Gibbs Manifolds

Dmitrii Pavlov, Bernd Sturmfels and Simon Telen

Abstract

Gibbs manifolds are images of affine spaces of symmetric matrices under the exponential map. They arise in applications such as optimization, statistics and quantum physics, where they extend the ubiquitous role of toric geometry. The Gibbs variety is the zero locus of all polynomials that vanish on the Gibbs manifold. We compute these polynomials and show that the Gibbs variety is low-dimensional. Our theory is applied to a wide range of scenarios, including matrix pencils and quantum optimal transport.

1 Introduction

Toric varieties provide the geometric foundations for many successes in the mathematical sciences. In statistics they appear as discrete exponential families [23, p. 2], and their ideals reveal Markov bases for sampling from conditional distributions [5]. In optimization, they furnish nonnegativity certificates [7] and they govern the entropic regularization of linear programming [21]. Notable sightings in phylogenetics, stochastic analysis, Gaussian inference and chemical reaction networks gave us the slogan that *the world is toric* [16, Section 8.3].

In all of these applications, the key player is the positive part of the toric variety. That real manifold is identified with a convex polytope by the moment map [16, Theorem 8.24]. The fibers of the underlying linear map are polytopes of complementary dimension, and each fiber intersects the toric variety uniquely, in the *Birch point*. This is the unique maximizer of the entropy over the polytope [19, Theorem 1.10]. In statistical physics and computer science [25], the Birch point is known as the *Gibbs distribution*. The name Gibbs refers to the maximum entropy state in a quantum system, and this is also the reason behind our title.

This paper initiates a non-commutative extension of applied toric geometry. In that extension, points in $\mathbb{R}^n$ are replaced by real symmetric $n \times n$ matrices, and linear programing is replaced by semidefinite programming. There is a moment map which takes the cone of positive semidefinite matrices onto a spectrahedral shadow, and whose fibers are spectrahedra of complementary dimension. The Gibbs manifold plays the role of the positive toric variety. Each spectrahedron intersects the Gibbs manifold uniquely, in the *Gibbs point*. Just like in the toric case, we study these objects algebraically by passing to the Zariski closure of our positive manifold. The resulting analogues of toric varieties are called *Gibbs varieties*.

We illustrate these concepts for the following linear space of symmetric 3×3-matrices:

$$\mathcal{L} \;=\; \left\{ \begin{bmatrix} y_1 + y_2 + y_3 & y_1 & y_2 \\ y_1 & y_1 + y_2 + y_3 & y_3 \\ y_2 & y_3 & y_1 + y_2 + y_3 \end{bmatrix} \;:\; y_1, y_2, y_3 \in \mathbb{R} \right\}. \tag{1}$$

The Gibbs manifold $GM(\mathcal{L})$ is obtained by applying the exponential function to each matrix in $\mathcal{L}$. Since the matrix logarithm is the inverse to the matrix exponential, it is a 3-dimensional manifold, contained in the 6-dimensional cone $\mathrm{int}(\mathbb{S}^3_+)$ of positive definite 3×3 matrices.

Consider the quotient map from the matrix space $\mathbb{S}^3 \simeq \mathbb{R}^6$ onto $\mathbb{S}^3/\mathcal{L} \simeq \mathbb{R}^3$. This takes a positive semidefinite matrix $X = [x_{ij}]$ to its inner products with the matrices in a basis of $\mathcal{L}$:

$$\pi \,:\, \mathbb{S}^3_+ \rightarrow \mathbb{R}^3 \,:\, X \mapsto \big(\, \mathrm{trace}(X) + 2x_{12}, \, \mathrm{trace}(X) + 2x_{13}, \, \mathrm{trace}(X) + 2x_{23} \,\big).$$

Precisely this map appeared in the statistical study of Gaussian models in [22, Example 1.1]. The fibers $\pi^{-1}(b)$ are three-dimensional spectrahedra, and these serve as feasible regions in optimization, both for semidefinite programming and for maximum likelihood estimation.

We here consider yet another convex optimization problem over the spectrahedron $\pi^{-1}(b)$, namely maximizing the *von Neumann entropy* $h(X) = \mathrm{trace}(X - X \cdot \log(X))$. This problem has a unique local and global maximum, at the intersection $\pi^{-1}(b) \cap GM(\mathcal{L})$. See Theorem 5.1. This *Gibbs point* is the maximizer of the entropy over the spectrahedron. Therefore, the Gibbs manifold $GM(\mathcal{L})$ is the set of Gibbs points in all fibers $\pi^{-1}(b)$, as b ranges over $\mathbb{R}^3$.

To study these objects algebraically, we ask for the polynomials that vanish on $GM(\mathcal{L})$. The zeros of these polynomials form the *Gibbs variety* $GV(\mathcal{L})$. Thus, the Gibbs variety is the Zariski closure of the Gibbs manifold. In our example, the Gibbs manifold has dimension 3, whereas the Gibbs variety has dimension 5. The latter is the cubic hypersurface $GV(\mathcal{L}) =$

$$\big\{ X \in \mathbb{S}^3 : (x_{11} - x_{22})(x_{11} - x_{33})(x_{22} - x_{33}) = x_{33}(x_{13}^2 - x_{23}^2) + x_{22}(x_{23}^2 - x_{12}^2) + x_{11}(x_{12}^2 - x_{13}^2) \big\}.$$

As promised, the study of Gibbs manifolds and Gibbs varieties is a non-commutative extension of applied toric geometry. Indeed, every toric variety is a Gibbs variety arising from diagonal matrices. For instance, the toric surface $\{ x \in \mathbb{R}^3 : x_1 x_3 = x_2^2 \}$ is realized as

$$GV(\mathcal{L}') = \big\{ X \in \mathbb{S}^3 \,:\, x_{11}x_{33} - x_{22}^2 = x_{12} = x_{13} = x_{23} = 0 \big\}$$

for the diagonal matrix pencil

$$\mathcal{L}' = \left\{ \begin{bmatrix} 2y_1 & 0 & 0 \\ 0 & y_1 + y_2 & 0 \\ 0 & 0 & 2y_2 \end{bmatrix} \,:\, y_1, y_2 \in \mathbb{R} \right\}. \tag{2}$$

However, even for diagonal matrices, the dimension of the Gibbs variety can exceed that of the Gibbs manifold. To see this, replace the matrix entry $2y_1$ by $\sqrt{2}y_1$ in the definition of $\mathcal{L}'$. This explains why transcendental number theory will make an appearance in our study.

Our presentation in this paper is organized as follows. Section 2 gives a more thorough introduction to Gibbs manifolds and Gibbs varieties. Theorem 2.4 states that the dimension of the Gibbs variety is usually quite small. The proof of this result is presented in Section 3. In that section we present algorithms for computing the prime ideal of the Gibbs variety. This is an implicitization problem, where the parametrization uses transcendental functions. We compare exact symbolic methods for solving that problem with a numerical approach. A key ingredient is the Galois group for the eigenvalues of a linear space of symmetric matrices. We

implemented our algorithms in `Julia`, making use of the computer algebra package `Oscar.jl` [18]. Our code and data are available at https://mathrepo.mis.mpg.de/GibbsManifolds.

In Section 4 we study the Gibbs varieties given by two-dimensional spaces of symmetric matrices. This rests on the classical Segre-Kronecker classification of matrix pencils [6].

In Section 5 we turn to the application that led us to start this project, namely *entropic regularization* in convex optimization. That section develops the natural generalization of the geometric results in [21] from linear programming to semidefinite programming. We conclude in Section 6 with a study of *quantum optimal transport* [3]. This is the semidefinite programming analogue to the classical optimal transport problem [21, Section 3]. We show that its Gibbs manifold is the positive part of a Segre variety in matrix space.

2 From manifolds to varieties

We write $\mathbb{S}^n$ for the space of symmetric $n \times n$-matrices. This is a real vector space of dimension $\binom{n+1}{2}$. The subset of positive semidefinite matrices is denoted $\mathbb{S}^n_+$. This is a full-dimensional closed semialgebraic convex cone in $\mathbb{S}^n$, known as the *PSD cone*. The PSD cone is self-dual with respect to the trace inner product, given by $\langle X, Y \rangle := \mathrm{trace}(XY)$ for $X, Y \in \mathbb{S}^n$.

The matrix exponential function is defined by the usual power series, which converges for all real and complex $n \times n$ matrices. It maps symmetric matrices to positive definite symmetric matrices. The zero matrix 0_n is mapped to the identity matrix id_n. We write

$$\exp \; : \; \mathbb{S}^n \to \mathrm{int}(\mathbb{S}^n_+) \,, \;\; X \; \mapsto \; \sum_{i=0}^{\infty} \frac{1}{i!} X^i.$$

This map is invertible, with the inverse given by the familiar series for the logarithm:

$$\log \; : \; \mathrm{int}(\mathbb{S}^n_+) \to \mathbb{S}^n \,, \;\; Y \; \mapsto \; \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j} \, (Y - \mathrm{id}_n)^j.$$

We next introduce the geometric objects studied in this article. We fix any matrix $A_0 \in \mathbb{S}^n$ and d linearly independent matrices $A_1, A_2, \ldots, A_d$, also in $\mathbb{S}^n$. We write $\mathcal{L}$ for the affine subspace $A_0 + \mathrm{span}_{\mathbb{R}}(A_1, \ldots, A_d)$ of the vector space $\mathbb{S}^n \simeq \mathbb{R}^{\binom{n+1}{2}}$. Thus, $\mathcal{L}$ is an *affine space of symmetric matrices* (ASSM) of dimension d. If $A_0 = 0$, then $\mathcal{L}$ is a *linear space of symmetric matrices* (LSSM). We are interested in the image of $\mathcal{L}$ under the exponential map:

Definition 2.1. The *Gibbs manifold* $\mathrm{GM}(\mathcal{L})$ of $\mathcal{L}$ is the d-dimensional manifold $\exp(\mathcal{L}) \subset \mathbb{S}^n_+$.

This is indeed a d-dimensional manifold inside the convex cone $\mathbb{S}^n_+$. It is diffeomorphic to $\mathcal{L} \simeq \mathbb{R}^d$, with the identification given by the exponential map and the logarithm map.

In notable special cases (e.g. that in Section 6), the Gibbs manifold is semi-algebraic, namely it is the intersection of an algebraic variety with the PSD cone. However, this fails in general, as seen in the Introduction. It is still interesting to ask which polynomial relations hold between the entries of any matrix in $\mathrm{GM}(\mathcal{L})$. This motivates the following definition.

Definition 2.2. The *Gibbs variety* $\mathrm{GV}(\mathcal{L})$ of $\mathcal{L}$ is the Zariski closure of $\mathrm{GM}(\mathcal{L})$ in $\mathbb{C}^{\binom{n+1}{2}}$.

Example 2.3 ($n = 4, d = 2$). Consider the 2-dimensional linear space of symmetric matrices

$$\mathcal{L} \;=\; \left\{ \begin{bmatrix} 0 & 0 & 0 & y_1 \\ 0 & 0 & y_1 & y_2 \\ 0 & y_1 & y_2 & 0 \\ y_1 & y_2 & 0 & 0 \end{bmatrix} \;:\; y_1, y_2 \in \mathbb{R} \right\} \;\subset\; \mathbb{S}^4 .$$

Its Gibbs manifold $\mathrm{GM}(\mathcal{L})$ is a surface in $\mathbb{S}^4 \simeq \mathbb{R}^{10}$. The Gibbs variety $\mathrm{GV}(\mathcal{L})$ has dimension five and degree three. It consists of all symmetric matrices $X = (x_{ij})$ whose entries satisfy

$$x_{13} - x_{22} + x_{44} \;=\; x_{14} - x_{23} + x_{34} \;=\; x_{24} - x_{33} + x_{44} \;=\; 0,$$

$$\text{and} \quad \mathrm{rank} \begin{bmatrix} x_{11} - x_{44} & x_{12} - x_{34} & x_{22} - x_{33} \\ x_{12} & x_{22} - x_{44} & x_{23} - x_{34} \end{bmatrix} \;\leq\; 1. \tag{3}$$

This follows from the general result on matrix pencils in Theorem 4.4. $\diamond$

The following dimension bounds constitute our main result on Gibbs varieties.

Theorem 2.4. *Let $\mathcal{L} \subset \mathbb{S}^n$ be an ASSM of dimension d. The dimension of the Gibbs variety $\mathrm{GV}(\mathcal{L})$ is at most $n + d$. If $A_0 = 0$, i.e. $\mathcal{L}$ is an LSSM, then $\dim \mathrm{GV}(\mathcal{L})$ is at most $n + d - 1$.*

These bounds are attained in many cases, including Example 2.3. Our proof of Theorem 2.4 appears in Section 3, in the context of algorithms for computing the ideal of $\mathrm{GV}(\mathcal{L})$.

While it might be difficult to find all polynomials that vanish on the Gibbs manifold, finding *linear* relations is sometimes easier. Such relations are useful for semidefinite optimization, see Remark 5.2. This brings us to the final geometric object studied in this paper.

Definition 2.5. The *Gibbs plane* $\mathrm{GP}(\mathcal{L})$ is the smallest affine space containing $\mathrm{GV}(\mathcal{L})$.

Clearly, we have the chain of inclusions $\mathrm{GM}(\mathcal{L}) \subseteq \mathrm{GV}(\mathcal{L}) \subseteq \mathrm{GP}(\mathcal{L}) \subseteq \mathbb{C}^{\binom{n+1}{2}}$.

Example 2.6. The Gibbs plane of the LSSM $\mathcal{L}$ from Example 2.3 is the 7-dimensional linear space in $\mathbb{C}^{10}$ that is defined by the three linear relations listed in the first row of (3). $\diamond$

We claimed in the Introduction that this article offers a generalization of toric varieties. In what follows, we make that claim precise, by discussing the case when $\mathcal{L}$ is a commuting family. This means that the symmetric matrices $A_0, A_1, \ldots, A_d$ commute pairwise, i.e. $A_i A_j = A_j A_i$ for all i, j. We now assume that this holds. Then the ASSM $\mathcal{L}$ can be diagonalized, i.e. there is an orthogonal matrix V such that $\Lambda_i = V^\top A_i V$ is a diagonal matrix, for all i. The vector $\lambda_i \in \mathbb{R}^n$ of diagonal entries in $\Lambda_i = \mathrm{diag}(\lambda_i)$ contains the eigenvalues of A_i.

The matrix exponential of any element in $\mathcal{L}$ can be computed as follows:

$$\exp(A_0 + y_1 A_1 + \cdots + y_d A_d) \;=\; V \cdot \exp(\Lambda_0 + y_1 \Lambda_1 + \cdots + y_d \Lambda_d) \cdot V^\top . \tag{4}$$

Let $\mathcal{D}$ denote this ASSM of diagonal matrices, i.e. $\mathcal{D} = \{\Lambda_0 + y_1 \Lambda_1 + \cdots + y_d \Lambda_d \;:\; y \in \mathbb{R}^d\}$. Then the linear change of coordinates given by V identifies the respective Gibbs manifolds:

$$\mathrm{GM}(\mathcal{L}) \;=\; V \cdot \mathrm{GM}(\mathcal{D}) \cdot V^\top . \tag{5}$$

4

The same statement holds for the Gibbs varieties and the Gibbs planes:

$$\mathrm{GV}(\mathcal{L}) = V \cdot \mathrm{GV}(\mathcal{D}) \cdot V^\top \qquad \text{and} \qquad \mathrm{GP}(\mathcal{L}) = V \cdot \mathrm{GP}(\mathcal{D}) \cdot V^\top. \qquad (6)$$

The dimensions of these objects are determined by arithmetic properties of the eigenvalues.

Recall that $\Lambda_i = \mathrm{diag}(\lambda_i)$ where λ_i is a vector in $\mathbb{R}^n$. Let Λ denote the linear subspace of $\mathbb{R}^n$ that is spanned by the d vectors $\lambda_1, \ldots, \lambda_d$. We have $\mathcal{D} = \lambda_0 + \Lambda$, and therefore

$$\mathrm{GM}(\mathcal{D}) = \exp(\lambda_0) \star \exp(\Lambda) = \{(e^{\lambda_{01}} w_1, \ldots, e^{\lambda_{0n}} w_n) : w \in \exp(\Lambda)\} \subset \mathbb{R}^n.$$

Here $\star$ denotes coordinate-wise multiplication in $\mathbb{R}^n$. Let $\Lambda_\mathbb{Q}$ be the smallest vector subspace of $\mathbb{R}^n$ spanned by elements of $\mathbb{Q}^n$ which contains Λ. Its dimension $d_\mathbb{Q} = \dim \Lambda_\mathbb{Q}$ satisfies $d \leq d_\mathbb{Q} \leq n$. Fix lattice vectors $a_1, a_2, \ldots, a_{d_\mathbb{Q}}$ in $\mathbb{Z}^n$ that form a basis of $\Lambda_\mathbb{Q}$. Then, inside an n-dimensional linear space defined by the diagonality condition, we have

$$\mathrm{GV}(\mathcal{D}) = \overline{\left\{ \Big(e^{\lambda_{01}} \prod_{i=1}^{d_\mathbb{Q}} z_i^{a_{i1}}, \; e^{\lambda_{02}} \prod_{i=1}^{d_\mathbb{Q}} z_i^{a_{i2}}, \; \ldots, \; e^{\lambda_{0n}} \prod_{i=1}^{d_\mathbb{Q}} z_i^{a_{in}} \Big) \; : \; z \in (\mathbb{C}^*)^{d_\mathbb{Q}} \right\}}. \qquad (7)$$

This is a toric variety of dimension $d_\mathbb{Q}$. Just like in [21, Section 2], the closure is taken in $\mathbb{C}^n$. The Gibbs manifold $\mathrm{GM}(\mathcal{D})$ is a d-dimensional subset of the real points in $\mathrm{GV}(\mathcal{D})$ for which z has strictly positive coordinates. We summarize our discussion in the following theorem.

Theorem 2.7. *Let $\mathcal{L}$ be an affine space of pairwise commuting symmetric matrices. Then the Gibbs variety $\mathrm{GV}(\mathcal{L})$ is a toric variety of dimension $d_\mathbb{Q}$, given explicitly by (6) and (7).*

For an illustration, consider the seemingly simple case $d = 1$ and $A_0 = 0$. Here, $\mathrm{GM}(\mathcal{L})$ is the curve formed by all powers of $\exp(A_1)$, and $\mathrm{GV}(\mathcal{L})$ is a toric variety of generally higher dimension. This scenario is reminiscent of that studied by Galuppi and Stanojkovski in [8].

Example 2.8. Let $n = 3$ and consider the LSSM $\mathcal{L}$ spanned by $A_1 = \begin{bmatrix} 4 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$. We have

$$A_1 = V \cdot \mathrm{diag}(\lambda) \cdot V^\top, \text{ where } \lambda = \big(2, 4 + \sqrt{2}, 4 - \sqrt{2}\big) \text{ and } V = \frac{1}{2} \begin{bmatrix} 0 & \sqrt{2} & -\sqrt{2} \\ -\sqrt{2} & 1 & 1 \\ \sqrt{2} & 1 & 1 \end{bmatrix}.$$

Here, $\mathcal{D} = \Lambda = \mathbb{R}\lambda$, $d_\mathbb{Q} = 2$, and $\Lambda_\mathbb{Q} = \mathbb{R}\{(1, 2, 2), (0, 1, -1)\} = \{p \in \mathbb{R}^3 : 4p_1 = p_2 + p_3\}$. Hence $\mathrm{GV}(\mathcal{D})$ is the toric surface $\{q_{11}^4 = q_{22}q_{33}\}$ in $\mathrm{GP}(\mathcal{D}) = \{Q \in \mathbb{S}^3 : q_{12} = q_{13} = q_{23} = 0\}$. We transform that surface into the original coordinates via (6). The computation reveals

$$\mathrm{GV}(\mathcal{L}) = \{X \in \mathrm{GP}(\mathcal{L}) : x_{23}^4 - 4x_{23}^3 x_{33} + 6x_{23}^2 x_{33}^2 - 4x_{23}x_{33}^3 + x_{33}^4 + 2x_{13}^2 - x_{23}^2 - 2x_{23}x_{33} - x_{33}^2 = 0\}.$$

The ambient 3-space is $\mathrm{GP}(\mathcal{L}) = \{X \in \mathbb{S}^3 : x_{11} - x_{23} - x_{33} = x_{12} - x_{13} = x_{22} - x_{33} = 0\}$. $\diamond$

This concludes our discussion of the toric Gibbs varieties that arise from pairwise commuting matrices. In the next section we turn to the general case, which requires new ideas.

3 Implicitization of Gibbs varieties

Implicitization is the computational problem of finding implicit equations for an object that comes in the form of a parametrization. When the parametrizing functions are rational functions, these equations are polynomials and can be found using resultants or Gröbner bases [16, Section 4.2]. A different approach rests on polynomial interpolation and numerical nonlinear algebra. This section studies the implicitization problem for Gibbs varieties. The difficulty arises from the fact that Gibbs manifolds are transcendental, since their parametrizations involve the exponential function. We start out by presenting our proof of Theorem 2.4.

As in Section 2, $\mathcal{L} = A_0 + \operatorname{span}_{\mathbb{R}}(A_1, \ldots, A_d)$ is a d-dimensional affine space of symmetric $n \times n$-matrices. Its elements are $A_0 + y_1 A_1 + \cdots + y_d A_d$. We shall parametrize the Gibbs manifold $\operatorname{GM}(\mathcal{L})$ in terms of the coordinates $y_1, \ldots, y_d$ on $\mathcal{L}$. This uses the following formula.

Theorem 3.1 (Sylvester [24]). *Let $f : D \to \mathbb{R}$ be an analytic function on an open set $D \subset \mathbb{R}$ and $M \in \mathbb{R}^{n \times n}$ a matrix that has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ in D. Then*

$$f(M) = \sum_{i=1}^{n} f(\lambda_i) M_i, \quad with \quad M_i = \prod_{j \neq i} \frac{1}{\lambda_i - \lambda_j} (M - \lambda_j \cdot \operatorname{id}_n).$$

We note that the product on the right hand side takes place in the commutative ring $\mathbb{R}[M]$.

Proof of Theorem 2.4. The characteristic polynomial of $A(y) = A_0 + y_1 A_1 + \cdots + y_d A_d$ equals

$$P_{\mathcal{L}}(\lambda; y) = \det(A(y) - \lambda \cdot \operatorname{id}_n) = c_0(y) + c_1(y) \lambda + \cdots + c_{n-1}(y) \lambda^{n-1} + (-1)^n \lambda^n.$$

Its zeros λ are algebraic functions of the coordinates $y = (y_1, \ldots, y_d)$ on $\mathcal{L}$.

We first assume that $\mathcal{L}$ has distinct eigenvalues, i.e. there is a Zariski open subset $U \subset \mathbb{R}^d$ such that $P_{\mathcal{L}}(\lambda; y^*)$ has n distinct real roots λ for all $y^* \in U$. Sylvester's formula writes the entries of $\exp(A(y))$ as rational functions of $y, \lambda_i(y)$ and $e^{\lambda_i(y)}$ for $y \in U$. These functions are multisymmetric in the pairs $(\lambda_i, e^{\lambda_i})$. They evaluate to convergent power series on $\mathbb{R}^d$.

Let V be the subvariety of $U \times \mathbb{R}^n$ that is defined by the equations

$$c_i(y) = (-1)^i \sigma_{n-i}(\lambda) \quad \text{for} \quad i = 0, \ldots, n-1, \tag{8}$$

where $\sigma_t(\lambda)$ is the tth elementary symmetric polynomial evaluated at $(\lambda_1, \ldots, \lambda_n)$. We have $\dim V = d$. Define a map $\phi : V \times \mathbb{R}^n \to \mathbb{S}^n$, using coordinates $z_1, \ldots, z_n$ on $\mathbb{R}^n$, as follows:

$$(y_1, \ldots, y_d, \lambda_1, \ldots, \lambda_n, z_1, \ldots, z_n) \longmapsto \sum_{i=1}^{n} z_i \prod_{j \neq i} \frac{1}{\lambda_i - \lambda_j} (A(y) - \lambda_j \cdot \operatorname{id}_n). \tag{9}$$

The closure $\overline{\phi(V \times \mathbb{R}^n)}$ of the image of this map is a variety. It contains the Gibbs variety: setting $z_i = e^{\lambda_i}$ parametrizes a dense subset of the Gibbs manifold, by Theorem 3.1.

The Gibbs variety of the LSSM $\mathbb{R}\mathcal{L}$ spanned by the ASSM $\mathcal{L}$ also lies in $\overline{\phi(V \times \mathbb{R}^n)}$, because $\exp(y_0 A(y)) = \phi(y_0 \cdot y, y_0 \cdot \lambda, e^{y_0 \cdot \lambda})$ for any $y \in U$ and $y_0 \in \mathbb{R} \backslash \{0\}$. We thus have

$$\dim \operatorname{GV}(\mathcal{L}) \leq \dim \overline{\phi(V \times \mathbb{R}^n)} \leq d + n \quad \text{and} \quad \dim \operatorname{GV}(\mathbb{R}\mathcal{L}) \leq d + n.$$

Finally, suppose that $\mathcal{L}$ is an LSSM, i.e. $A_0 = 0$. Then $\mathcal{L}$ is the linear span of an ASSM of dimension $d-1$ in $\mathbb{S}^n$. The second inequality therefore gives $\dim \mathrm{GV}(\mathcal{L}) \leq d + n - 1$.

We finally consider the case when $\mathcal{L}$ has $m < n$ distinct eigenvalues. Since symmetric matrices are diagonalizable, Sylvester's formula can easily be adapted to this case: it suffices to sum over the distinct eigenvalues of M, and to adjust the parametrization (9) accordingly. That is, we replace n by m. See [12, Chapter 6.1, Problem 14] for details. $\square$

Remark 3.2. If the points $\exp(\lambda(y)) = (e^{\lambda_1(y)}, \ldots, e^{\lambda_n(y)})$, $y \in U$, lie in a lower-dimensional subvariety $W \subset \mathbb{R}^n$, then the proof of Theorem 2.4 gives the better bound $\dim \mathrm{GV}(\mathcal{L}) \leq d + \dim W$. We saw this in Example 2.8. In general, no such subvariety W exists, i.e. one expects $W = \mathbb{R}^n$. This is an issue of Galois theory, to be discussed at the end of this section.

For ease of exposition, we work only with LSSMs in the rest of this section. That is, we set $A_0 = 0$. We comment on the generalization to ASSMs in Remark 3.7. Our discussion and the proof of Theorem 2.4 suggest Algorithm 1, for computing the ideal of the Gibbs variety of an LSSM $\mathcal{L}$. That ideal lives in a polynomial ring $\mathbb{R}[X]$ whose variables are the entries

---

Algorithm 1 Implicitization of the Gibbs variety of an LSSM $\mathcal{L}$, defined over $\mathbb{Q}$

  Input: Linearly independent matrices $A_1, \ldots, A_d \in \mathbb{S}^n$ with rational entries
  Output: Polynomials that define $\mathrm{GV}(\mathcal{L})$, where $\mathcal{L} = \mathrm{span}_{\mathbb{R}}(A_1, \ldots, A_d)$
1: Compute the characteristic polynomial $P_{\mathcal{L}}(\lambda; y) = c_0(y) + c_1(y)\lambda + \cdots + c_n(y)\lambda^n$
Require: $P_{\mathcal{L}}(\lambda; y)$ has n distinct roots in $\overline{\mathbb{R}(y)}$
2: $E_1' \leftarrow \{$the n polynomials $(-1)^i \sigma_{n-i}(\lambda) - c_i(y)$ in (8)$\}$
3: $E_1 \leftarrow \{$generators of any associated prime over $\mathbb{Q}$ of $\langle E_1' \rangle\}$
4: $E_2 \leftarrow \{$the entries of $\phi(y, \lambda, z) - X\}$, with $X = (x_{ij})$ a symmetric matrix of variables
5: $E_2, D \leftarrow$ clear denominators in E_2 and record the least common denominator D
6: **if** the roots $\lambda_1, \ldots, \lambda_n$ of $P_{\mathcal{L}}(\lambda; y)$ are $\mathbb{Q}$-linearly dependent **then**
7: $E_3 \leftarrow \{z^\alpha - z^\beta : \sum \alpha_i \lambda_i = \sum \beta_j \lambda_j, \ \alpha, \beta \in \mathbb{Z}_{\geq 0}^n\}$
8: **else**
9: $E_3 \leftarrow \emptyset$
10: $I \leftarrow$ the ideal generated by E_1, E_2, E_3 in the polynomial ring $\mathbb{R}[y, \lambda, z, X]$
11: $I \leftarrow I : D^\infty$
12: $J \leftarrow$ elimination ideal obtained by eliminating y, λ, z from I
13: **return** a set of generators of J

---

of a symmetric $n \times n$ matrix. The algorithm builds three subsets E_1, E_2, E_3 of the larger polynomial ring $\mathbb{R}[y, \lambda, z, X]$. After the saturation (step 11), the auxiliary variables y, λ, z are eliminated. The equations E_1' come from (8). They constrain (y, λ) to lie in V. The set E_1 generates an associated prime of $\langle E_1' \rangle$ (step 3), see the discussion preceding Theorem 3.6. The equations E_2 come from the parametrization (9). Note that, if $\mathcal{L}$ has $m < n$ distinct eigenvalues, this formula can be adjusted as in the end of the proof of Theorem 2.4, and the requirement after step 1 can be dropped. Later in the algorithm, one replaces n with m. It is necessary to clear denominators in order to obtain polynomials (step 5). The saturation by the LCD D avoids spurious components arising from this step. Finally, E_3 accounts for toric

relations between the z_i arising from $\mathbb{Q}$-linear relations among the λ_i. If no such relations exist, then Theorem 3.5 ensures that the assignment $E_3 \leftarrow \emptyset$ in step 9 is correct.

Steps 6 and 7 in Algorithm 1 require a detailed discussion. Further below we shall explain the $\mathbb{Q}$-linear independence of eigenvalues, how to check this, and how to compute E_3. Ignoring this for now, one can also run Algorithm 1 with $E_3 = \emptyset$. Then step 13 still returns polynomials that vanish on the Gibbs variety $\mathrm{GV}(\mathcal{L})$ but these may cut out a larger variety.

We implemented Algorithm 1 in Julia (v1.8.3), using Oscar.jl [18], and tested it on many examples. The code is available at https://mathrepo.mis.mpg.de/GibbsManifolds.

Example 3.3. The Gibbs variety $\mathrm{GV}(\mathcal{L})$ for the LSSM $\mathcal{L}$ in (1) has the parametrization

$$\phi = \sum_{i=1}^{3} \frac{z_i}{q(\lambda_i, y_1, y_2, y_3)} \begin{bmatrix} p_{11}(\lambda_i, y_1, y_2, y_3) & p_{12}(\lambda_i, y_1, y_2, y_3) & p_{13}(\lambda_i, y_1, y_2, y_3) \\ p_{12}(\lambda_i, y_1, y_2, y_3) & p_{22}(\lambda_i, y_1, y_2, y_3) & p_{23}(\lambda_i, y_1, y_2, y_3) \\ p_{13}(\lambda_i, y_1, y_2, y_3) & p_{23}(\lambda_i, y_1, y_2, y_3) & p_{33}(\lambda_i, y_1, y_2, y_3) \end{bmatrix}, \quad \text{where}$$

$$
\begin{aligned}
q &= 2y_1^2 + 6y_1y_2 + 2y_2^2 + 6y_1y_3 + 6y_2y_3 + 2y_3^2 - 6y_1\lambda - 6y_2\lambda - 6y_3\lambda + 3\lambda^2, \\
p_{11} &= y_1^2 + 2y_1y_2 + y_2^2 + 2y_1y_3 + 2y_2y_3 - 2y_1\lambda - 2y_2\lambda - 2y_3\lambda + \lambda^2, \\
p_{12} &= -y_1^2 - y_1y_2 - y_1y_3 + y_2y_3 + y_1\lambda, \\
p_{13} &= -y_1y_2 - y_2^2 + y_1y_3 - y_2y_3 + y_2\lambda, \\
p_{22} &= y_1^2 + 2y_1y_2 + 2y_1y_3 + 2y_2y_3 + y_3^2 - 2y_1\lambda - 2y_2\lambda - 2y_3\lambda + \lambda^2, \\
p_{23} &= y_1y_2 - y_1y_3 - y_2y_3 - y_3^2 + y_3\lambda, \\
p_{33} &= 2y_1y_2 + y_2^2 + 2y_1y_3 + 2y_2y_3 + y_3^2 - 2y_1\lambda - 2y_2\lambda - 2y_3\lambda + \lambda^2.
\end{aligned}
$$

Our Julia code for Algorithm 1 easily finds the cubic polynomial defining $\mathrm{GV}(\mathcal{L})$. ◇

In spite of such successes, symbolic implicitization is limited to small n and d. Numerical computations can help, in some cases, to find equations for more challenging Gibbs varieties.

Example 3.4. We consider the LSSM of 4×4 Hankel matrices with upper left entry zero:

$$\mathcal{L} = \left\{ \begin{bmatrix} 0 & y_2 & y_3 & y_4 \\ y_2 & y_3 & y_4 & y_5 \\ y_3 & y_4 & y_5 & y_6 \\ y_4 & y_5 & y_6 & y_7 \end{bmatrix} : (y_2, \dots, y_7) \in \mathbb{R}^6 \right\}.$$

Algorithm 1 failed to compute its Gibbs variety. We proceed using numerics as follows. Fix a degree $D > 0$ and let $N = \binom{9+D}{D}$ be the number of monomials in the 10 coordinates $x_{11}, \dots, x_{44}$ on $\mathbb{S}^4$. We create $M \geq N$ samples on $\mathrm{GM}(\mathcal{L})$ by plugging in random values for the six parameters y_i and applying the matrix exponential. Finding all vanishing equations of degree D on these samples amounts to computing the kernel of an $M \times N$ Vandermonde matrix. If this matrix has full rank, then there are no relations of degree D. We implemented this procedure in Julia. In our example, Theorem 2.4 suggests that $\mathrm{GV}(\mathcal{L})$ is a hypersurface, and this is indeed the case. Its defining equation has degree $D = 6$. We found it using $M = 5205 \geq N = 5005$ samples. Our denary sextic has 853 terms with integer coefficients:

$$x_{11}^3 x_{22} x_{24} x_{34} - x_{11}^3 x_{23}^2 x_{34} - x_{11}^3 x_{23} x_{24}^2 + x_{11}^3 x_{23} x_{24} x_{33} + \dots + 3x_{23} x_{24}^2 x_{33} x_{34}^2 + x_{24}^4 x_{33} x_{34} - x_{24}^3 x_{33}^2 x_{34}.$$

Its Newton polytope has the f-vector $(456, 5538, 21560, 41172, 44707, 29088, 11236, 2370, 211)$. Note that the package `Oscar.jl` conveniently allows to perform symbolic and numerical implicitization and polyhedral computations in the same programming environment.

We emphasize that our numerical `Julia` code is set up to find *exact* integer coefficients. For this, we first normalize the numerical approximation of the coefficient vector by setting its first (numerically) nonzero entry to one. Then we rationalize the coefficients using the built in command `rationalize` in `Julia`, with error tolerance `tol = 1e-7`. Correctness of the result is proved by checking that the resulting polynomial vanishes on the parametrization. ⋄

We now turn to $\mathbb{Q}$-linear relations among eigenvalues of $\mathcal{L}$. Our arithmetic discussion begins with a version of [1, (SP)], which is well-known in transcendental number theory:

Theorem 3.5 (Ax-Schanuel). *If the eigenvalues $\lambda_1, \dots, \lambda_n$ of the LSSM $\mathcal{L}$ are $\mathbb{Q}$-linearly independent, then $e^{\lambda_1}, \dots, e^{\lambda_n}$ are algebraically independent over the field $\mathbb{C}(y_1, \dots, y_d)$.*

On the other hand, suppose that the eigenvalues of $\mathcal{L}$ satisfy some non-trivial linear relation over $\mathbb{Q}$. We can then find nonnegative integers α_i and β_j, not all zero, such that

$$\sum_{i=1}^{n} \alpha_i \lambda_i \;=\; \sum_{j=1}^{n} \beta_j \lambda_j. \tag{10}$$

This implies that the exponentials of the eigenvalues satisfy the toric relations

$$\prod_{i=1}^{n} z_i^{\alpha_i} \;=\; \prod_{j=1}^{n} z_j^{\beta_j}. \tag{11}$$

The linear relations (10) can be found from the ideal $\langle E_1' \rangle$ in step 2 which specifies that the λ_i are the eigenvalues of $A(y)$. This ideal is radical if we assume that $\mathcal{L}$ has distinct eigenvalues. We compute the prime decomposition of the ideal over $\mathbb{Q}$. All prime components are equivalent under permuting the λ_i, so we replace $\langle E_1' \rangle$ by any of these prime ideals in step 3. We compute (10) as the linear forms in that prime ideal. Using (11), we compute the toric ideal $\langle E_3 \rangle$ in step 7, which is also prime. This ideal defines a toric variety W', whose S_n-orbit is the variety W in Remark 3.2. We arrive at the following result.

Theorem 3.6. *Let $\mathcal{L} \subset \mathbb{S}^n$ be an LSSM with distinct eigenvalues. The Gibbs variety $\mathrm{GV}(\mathcal{L})$ is irreducible and unirational, and the ideal J found in Algorithm 1 is its prime ideal.*

Proof. Sylvester's formula yields a rational parametrization ψ of $\mathrm{GV}(\mathcal{L})$ with parameters $y_1, \dots, y_d, z_1, \dots, z_n$. The parameters λ_i in (9) can be omitted: the entries in the image are multisymmetric in (λ_i, z_i), so that they can be expressed in terms of elementary symmetric polynomials of the λ_i [2, Theorem 1]. The point $(z_1, \dots, z_n)$ lies on the toric variety W' defined above. The domain $\mathbb{C}^d \times W'$ of ψ is an irreducible variety, and it is also rational. The image of ψ is the Gibbs variety $\mathrm{GV}(\mathcal{L})$, which is therefore unirational and irreducible. The ideals given by E_1 and E_2 in Algorithm 1 are prime, after saturation, and elimination in step 12 preserves primality. Hence the output in J in step 13 is the desired prime ideal. □

We define the *Galois group* $G_{\mathcal{L}}$ of an LSSM $\mathcal{L}$ to be the Galois group of the characteristic polynomial $P_{\mathcal{L}}(\lambda, y)$ over the field $\mathbb{Q}(y_1, \ldots, y_d)$. Note that $G_{\mathcal{L}}$ is the subgroup of the symmetric group S_n whose elements are permutations that fix each associated prime of $\langle E_1' \rangle$. Hence the index of the Galois group $G_{\mathcal{L}}$ in S_n is the number of associated primes. In particular, the Galois group equals S_n if and only if the ideal $\langle E_1' \rangle$ formed in step 2 is prime.

The existence of linear relations (10) depends on the Galois group $G_{\mathcal{L}}$. If the Galois group is small then the primes of $\langle E_1 \rangle$ are large, and more likely to contain linear forms. There is a substantial literature in number theory on this topic. See [9, 10] and the references therein. For instance, by Kitaoka [14, Proposition 2], there are no linear relations if n is prime, or if $n \geq 6$ and the Galois group is S_n or A_n. If this holds, $E_3 = \emptyset$ in step 9 of Algorithm 1.

The computation of Galois groups is a well-studied topic in symbolic computation and number theory. Especially promising are methods based on numerical algebraic geometry (e.g. in [11]). These fit well with the approach to implicitization in Example 3.4. For a future theoretical project, it would be very interesting to classify LSSMs by their Galois groups.

Remark 3.7. We briefly comment on how to adjust Algorithm 1 to compute the Gibbs variety of an ASSM $\mathcal{L}$ with $A_0 \neq 0$. In this case, algebraic relations between $e^{\lambda_1}, \ldots, e^{\lambda_n}$ come from $\mathbb{Q}$-linear relations between the eigenvalues of $\mathcal{L}$, but this time modulo $\mathbb{C}$: an affine relation $\sum \alpha_i \lambda_i = \sum \beta_j \lambda_j + \gamma$ gives $z^\alpha - e^\gamma \cdot z^\beta = 0$, where $z_i = e^{\lambda_i}$, $\alpha_i, \beta_j \in \mathbb{Z}_{\geq 0}$, $\gamma \in \mathbb{C}$. Here γ is a $\mathbb{Q}$-linear combination of eigenvalues of A_0. Theorem 3.6 holds for ASSMs as well, provided that these $\mathbb{Q}$-linear relations modulo $\mathbb{C}$ can be computed in practice. This can usually not be done over $\mathbb{Q}$. We leave this algorithmic challenge for future research.

4 Pencils of quadrics

In this section we study the Gibbs variety $\mathrm{GV}(\mathcal{L})$ where $\mathcal{L} \subset \mathbb{S}^n$ is a pencil of quadrics, i.e. an LSSM of dimension $d = 2$. We follow the exposition in [6], where pencils $\mathcal{L}$ are classified by Segre symbols. The *Segre symbol* $\sigma = \sigma(\mathcal{L})$ is a multiset of partitions that sum up to n. It is computed as follows: Pick a basis $\{A_1, A_2\}$ of $\mathcal{L}$, where A_2 is invertible, and find the Jordan canonical form of $A_1 A_2^{-1}$. Each eigenvalue determines a partition, according to the sizes of the corresponding Jordan blocks. The multiset of these partitions is the Segre symbol σ.

We use the canonical form in [6, Section 2]. Suppose the Segre symbol is $\sigma = [\sigma_1, \ldots, \sigma_r]$, where the ith partition σ_i equals $(\sigma_{i,1} \geq \sigma_{i,2} \geq \cdots \geq \sigma_{i,n} \geq 0)$. There are r groups of blocks, one for each eigenvalue α_i of $A_1 A_2^{-1}$. The jth matrix in the ith group is the $\sigma_{i,j} \times \sigma_{i,j}$ matrix

$$
y_1 \cdot \begin{bmatrix} 0 & 0 & \ldots & 0 & \alpha_i \\ 0 & 0 & \ldots & \alpha_i & 1 \\ \vdots & \vdots & \reflectbox{$\ddots$} & \reflectbox{$\ddots$} & \vdots \\ 0 & \alpha_i & 1 & \reflectbox{$\ddots$} & 0 \\ \alpha_i & 1 & \ldots & 0 & 0 \end{bmatrix} + y_2 \cdot \begin{bmatrix} 0 & \ldots & 0 & 0 & 1 \\ 0 & \ldots & 0 & 1 & 0 \\ 0 & \ldots & 1 & 0 & 0 \\ \vdots & \reflectbox{$\ddots$} & \vdots & \vdots & \vdots \\ 1 & \ldots & 0 & 0 & 0 \end{bmatrix}.
$$

There are 13 Segre symbols for $n = 4$; see [6, Example 3.1]. It is instructive to compute their Gibbs varieties. All possible dimensions, $2, 3, 4$ and 5, are attained. Dimension 2 arises for the diagonal pencil $\mathcal{L}_\sigma = \mathrm{diag}(\alpha_1 y_1 + y_2, \alpha_2 y_1 + y_2, \alpha_3 y_1 + y_2, \alpha_4 y_1 + y_2)$, with Segre symbol

$\sigma = [1, 1, 1, 1]$. When the α_i are distinct integers, $\mathrm{GV}(\mathcal{L}_\sigma) = \mathrm{GM}(\mathcal{L}_\sigma)$ is a toric surface. This is similar to (2). Dimension 5 arises for $\sigma = [4]$, which was presented in Example 2.3.

The following examples, also computed with Algorithm 1, exhibit the dimensions $5, 4, 3$.

Example 4.1. Consider the Segre symbol $\sigma = [3, 1]$. The canonical pencil $\mathcal{L}_{[3,1]}$ is spanned by

$$
\begin{bmatrix} 0 & 0 & \alpha_1 & 0 \\ 0 & \alpha_1 & 1 & 0 \\ \alpha_1 & 1 & 0 & 0 \\ 0 & 0 & 0 & \alpha_2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \text{for} \ \ \alpha_1, \alpha_2 \in \mathbb{R} \ \text{distinct}.
$$

Here, $\dim \mathrm{GV}(\mathcal{L}_{[3,1]}) = 5$, the upper bound in Theorem 2.4. Algorithm 1 produces the ideal

$$
J \ = \ \big\langle\, x_{14}, x_{24}, x_{34},\, x_{13} - x_{22} + x_{33},\, x_{12}^2 - x_{11}x_{22} - x_{12}x_{23} + x_{11}x_{33} + x_{22}x_{33} - x_{33}^2 \,\big\rangle.
$$

If $\alpha_1 = \alpha_2$, then the Segre symbol changes to $\sigma = [(3, 1)]$. We now find the additional cubic

$$
x_{11}x_{22}x_{33} + 2x_{12}x_{13}x_{23} - x_{13}^2 x_{22} - x_{11}x_{23}^2 - x_{12}^2 x_{33} \ - \ x_{44} \quad \in \ J. \tag{12}
$$

This cuts down the dimension by one, and we now have $\dim \mathrm{GV}(\mathcal{L}_{[(3,1)]}) = 4$. $\qquad\qquad\diamond$

Example 4.2. Consider the Segre symbol $\sigma = [(2, 2)]$. The pencil $\mathcal{L}_{[(2,2)]}$ is spanned by

$$
\begin{bmatrix} 0 & \alpha & 0 & 0 \\ \alpha & 1 & 0 & 0 \\ 0 & 0 & 0 & \alpha \\ 0 & 0 & \alpha & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad \text{for some} \ \ \alpha \in \mathbb{R}.
$$

A version of Algorithm 1 for LSSMs with multiple eigenvalues produces the ideal

$$
J \ = \ \big\langle\, x_{11} - x_{33},\, x_{12} - x_{34},\, x_{22} - x_{44},\, x_{13},\, x_{14},\, x_{23},\, x_{24} \,\big\rangle.
$$

The Gibbs variety $\mathrm{GV}(\mathcal{L}_{[(2,2)]})$ is 3-dimensional and equals the Gibbs plane $\mathrm{GP}(\mathcal{L}_{[(2,2)]})$. $\quad\diamond$

The cubic (12) which distinguishes the Segre symbols $[3, 1]$ and $[(3, 1)]$ is explained by the following result. This applies not just to pencils but to all ASSMs with block structure.

Proposition 4.3. *Let $\mathcal{L}$ be a block-diagonal ASSM with r blocks $X_i(y)$ of size τ_i, where $\tau_1 + \cdots + \tau_r = n$. The Gibbs plane $\mathrm{GP}(\mathcal{L})$ is contained in $\mathbb{S}^{\tau_1} \times \cdots \times \mathbb{S}^{\tau_r} \subset \mathbb{S}^n$. Moreover, with the notation $\mathcal{J} = \{\{i, j\} \in \binom{[r]}{2} : \mathrm{trace}(X_i(y)) = \mathrm{trace}(X_j(y))\}$, we have*

$$
\mathrm{GV}(\mathcal{L}) \ \subseteq \ \{(X_1, \ldots, X_r) \in \mathrm{GP}(\mathcal{L}) \, : \, \det(X_i) = \det(X_j) \text{ for all } \{i, j\} \in \mathcal{J}\}.
$$

Proof. Block-diagonal matrices are exponentiated block-wise. The entries outside the diagonal blocks are zero. The statement follows from $\det(\exp(X_i(y))) = \exp(\mathrm{trace}(X_i(y)))$. $\quad\square$

Proposition 4.3 holds for the canonical pencil $\mathcal{L}_\sigma$ of any Segre symbol σ. First of all, for all indices (i,j) outside the diagonal blocks, we have $x_{ij} = 0$ on the Gibbs plane $\mathrm{GP}(\mathcal{L}_\sigma)$. Next, one has equations for the exponential of a single block, like those in Theorem 4.4 below. Finally, there are equations that link the blocks corresponding to entries σ_{ij} of the same partition σ_i. Some of these come from trace equalities between blocks of $\mathcal{L}_\sigma$, and this is the scope of Proposition 4.3. In particular, blocks ij and ik for which $\sigma_{ij} = \sigma_{ik} \bmod 2$ exponentiate to $X_{ij} \in \mathbb{S}_+^{\sigma_{ij}}$ and $X_{ik} \in \mathbb{S}_+^{\sigma_{ik}}$ with equal determinant. We saw this in (12). In all examples we computed, the three classes of equations above determine the Gibbs variety.

We now derive the equations that hold for the exponential of a single block. To this end, we fix $\sigma = [n]$ with $\alpha_1 = 0$. The canonical LSSM $\mathcal{L}_{[n]}$ consists of the symmetric matrices

$$
Y = \begin{bmatrix}
0 & 0 & \ldots & 0 & y_1 \\
0 & 0 & \ldots & y_1 & y_2 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & y_1 & y_2 & \vdots & 0 \\
y_1 & y_2 & \vdots & 0 & 0
\end{bmatrix}.
$$

The case $n = 4$ was featured in Example 2.3. In what follows we generalize that example.

Theorem 4.4. *The following linear equations hold on the Gibbs plane* $\mathrm{GP}(\mathcal{L}_{[n]})$:

$$
x_{i-1,j} + x_{i+1,j} \;=\; x_{i,j-1} + x_{i,j+1} \quad \text{for } 2 \leqslant i < j \leqslant n. \tag{13}
$$

The 2×2-minors of the following $2 \times (n-1)$ matrix vanish on the Gibbs variety $\mathrm{GV}(\mathcal{L}_{[n]})$:

$$
D(X) \;=\; \begin{bmatrix} x_{11} & x_{12} & x_{22} & \cdots \\ x_{12} & x_{22} & x_{23} & \cdots \end{bmatrix} - \begin{bmatrix} x_{n,n} & x_{n-1,n} & x_{n-1,n-1} & \cdots \\ 0 & x_{n,n} & x_{n-1,n} & \cdots \end{bmatrix}. \tag{14}
$$

If the Galois group $G_{\mathcal{L}_{[n]}}$ is the symmetric group S_n, then the prime ideal of $\mathrm{GV}(\mathcal{L}_{[n]})$ is generated by (13) and (14), and we have $\dim \mathrm{GP}(\mathcal{L}_n) = 2n - 1$ *and* $\dim \mathrm{GV}(\mathcal{L}_{[n]}) = n + 1$.

Remark 4.5. We conjecture that $G_{\mathcal{L}_{[n]}} = S_n$. This was verified computationally for many values of n, but we currently do not have a proof that works for all n. This gap underscores the need, pointed out at the end of Section 3, for a study of the Galois groups of LSSMs.

Proof. We claim that the linear equations (13) hold for every non-negative integer power of Y. This implies that they hold for $\exp(Y)$. We will show this by induction. The equations clearly hold for $Y^0 = \mathrm{id}_n$. Suppose they hold for $(m_{ij}) = M = Y^k$. Write $(b_{ij}) = B := Y^{k+1} = MY$.

The two-banded structure of Y implies $b_{i,j} = y_1 \cdot m_{i,n-j+1} + y_2 \cdot m_{i,n-j+2}$ for $1 \leqslant i < j$. The following identity holds for $2 \leqslant i < j$, and it shows that $\exp(Y)$ satisfies the equations (13):

$$
\begin{aligned}
b_{i-1,j} - b_{i,j-1} - b_{i,j+1} + b_{i+1,j} \;=\;\; & y_1 \cdot m_{i-1,n-j+1} + y_2 \cdot m_{i-1,n-j+2} - y_1 \cdot m_{i,n-j+2} - \\
& y_2 \cdot m_{i,n-j+3} - y_1 \cdot m_{i,n-j} - y_2 \cdot m_{i,n-j+1} + y_1 \cdot m_{i+1,n-j+1} + y_2 \cdot m_{i+1,n-j+2} \\
=\;\; & y_1 \cdot (m_{i-1,n-j+1} - m_{i,n-j+2} - m_{i,n-j} + m_{i+1,n-j+1}) + \\
& y_2 \cdot (m_{i-1,n-j+2} - m_{i,n-j+3} - m_{i,n-j+1} + m_{i+1,n-j+2}) \;=\; 0.
\end{aligned}
$$

12

We next consider the matrix $D(X)$ in (14). We must show that $D(X)$ has rank ≤ 1 for $X \in \mathrm{GV}(\mathcal{L}_{[n]})$. We claim that the rows of $D(Y^k)$ are proportional with the same coefficient for all $k \in \mathbb{Z}_{\geq 0}$. This will imply that the rows of $D(\exp{(Y)})$ are proportional. For the proof, let $\vec{v}_1$ and $\vec{v}_2$ be the rows of $D(B)$, where $B = Y^k$. We will show that $y_1 \vec{v}_1 + y_2 \vec{v}_2 = 0$.

First note that $D(\mathrm{id}_n) = 0$. Also note that each column of $D(B)$ has the form

$$\begin{bmatrix} b_{i,i} - b_{n+1-i,n+1-i} \\ b_{i,i+1} - b_{n+1-i,n+2-i} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} b_{i,i+1} - b_{n-i,n+1-i} \\ b_{i+1,i+1} - b_{n-i+1,n-i+1} \end{bmatrix}.$$

We start with the left case. We must show $y_1(b_{i,i} - b_{n+1-i,n+1-i}) + y_2(b_{i,i+1} - b_{n+1-i,n+2-i}) = 0$.

Recall from above that $b_{i,j} = y_1 \cdot m_{i,n-j+1} + y_2 \cdot m_{i,n-j+2}$, where $(m_{i,j}) = M = Y^{k-1}$ for $i < j$. Using this and the fact that the powers of Y are symmetric, we write

$$y_1(b_{i,i} - b_{n+1-i,n+1-i}) + y_2(b_{i,i+1} - b_{n+1-i,n+2-i}) =$$
$$= y_1((y_1 \cdot m_{i,n-i+1} + y_2 \cdot m_{i,n-i+2}) - (y_1 \cdot m_{n+1-i,i} + y_2 \cdot m_{n+1-i,i+1})) +$$
$$y_2((y_1 \cdot m_{i,n-i} + y_2 \cdot m_{i,n-i+1}) - (y_1 \cdot m_{n+1-i,i-1} + y_2 \cdot m_{n+1-i,i})) $$
$$= y_1 y_2(m_{i,n-i+2} - m_{i+1,n+1-i} + m_{i,n-i} - m_{n+1-i,i-1}) = 0,$$

where the last equality follows from (13). Now, for the second case we have

$$y_1(b_{i,i+1} - b_{n-i,n+1-i}) + y_2(b_{i+1,i+1} - b_{n-i+1,n-i+1}) =$$
$$y_1(y_1 \cdot m_{i,n-i} + y_2 \cdot m_{i,n-i+1} - y_1 \cdot m_{n-i,i} - y_2 \cdot m_{n-i,i+1}) +$$
$$y_2(y_1 \cdot m_{i+1,n-i} + y_2 \cdot m_{i+1,n-i+1} - y_1 \cdot m_{n-i+1,i} - y_2 \cdot m_{n-i+1,i+1}) = 0.$$

This proves that the 2×2 minors of $D(X)$ vanish on the Gibbs variety $\mathrm{GV}(\mathcal{L}_{[n]})$.

Suppose now that the eigenvalues of Y are $\mathbb{Q}$-linearly independent. We can check this directly for $n \leq 5$. For $n \geq 6$ it follows from our hypothesis $G_{\mathcal{L}_{[n]}} = S_n$, by [14, Proposition 2]. That hypothesis implies $\dim \mathrm{GV}(\mathcal{L}_{[n]}) = n + 1$, by Theorems 2.4 and 3.5. The 2×2-minors of $D(X)$ generate a prime ideal of codimension $n - 2$ in the coordinates of the $(2n-1)$-dimensional space given by (13). The equality of dimensions yields $\mathrm{GP}(\mathcal{L}_{[n]}) = 2n - 1$, and we conclude that our linear and quadratic constraints generate the prime ideal of $\mathrm{GV}(\mathcal{L}_{[n]})$. $\quad \square$

Theorem 2.4 and its refinement in Remark 3.2 furnish an upper bound on the dimension of any Gibbs variety. This raises the question when this bound is attained. In what follows, we offer a complete answer for $d = 2$. Let $\mathcal{L}$ be a pencil with eigenvalues $\lambda_i(y)$, and let W denote the Zariski closure in $\mathbb{R}^n$ of the set of points $\exp(\lambda(y)) = (e^{\lambda_1(y)}, \ldots, e^{\lambda_n(y)})$, $y \in \mathbb{R}^2$.

Theorem 4.6. *Let* $\mathcal{L} = \mathrm{span}_{\mathbb{R}}(A_1, A_2)$*, where* $A_1 A_2 \neq A_2 A_1$*. Then* $\dim \mathrm{GV}(\mathcal{L}) = \dim(W) + 1$*. In particular, if the Galois group* $G_{\mathcal{L}}$ *is the symmetric group* S_n *then* $\dim \mathrm{GV}(\mathcal{L}) = n + 1$*.*

Proof. We claim that the fibers of the map $\phi : V \times W \to \mathbb{S}^n$ defined by (9) are one-dimensional. Let $B \in \phi(V \times W)$ and consider any point $p = (y_1, y_2, \lambda_1, \ldots, \lambda_n, z_1, \ldots, z_n) \in \phi^{-1}(B)$. The condition that p lies in the fiber $\phi^{-1}(B)$ is equivalent to

(1) $z_1, \ldots, z_n$ are the eigenvalues of B, and

(2) $X = y_1 A_1 + y_2 A_2$ and B have the same eigenvectors, and

(3) $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of X.

Condition (1) follows from Theorem 3.1 for $f = \exp$. It implies that there are only finitely many possibilities for the z-coordinates of the point p in the fiber: up to permutations, they are the eigenvalues of B. Condition (3) follows from $(y_1, y_2, \lambda_1, \ldots, \lambda_n) \in V$. It says that the λ-coordinates are determined, up to permutation, by y_1, y_2. Therefore, it suffices to show that the matrices in $\mathcal{L}$ whose eigenvectors are those of B form a one-dimensional subvariety.

Symmetric matrices have common eigenvectors if and only if they commute. Define $S = \{X = y_1 A_1 + y_2 A_2 \in \mathcal{L} : X \cdot B = B \cdot X\} \subset \mathcal{L}$. This is a pairwise commuting linear subspace. Note that S contains a nonzero matrix X, since there is a point in $\phi^{-1}(B)$ whose y-coordinates define a nonzero matrix in $\mathcal{L}$. Therefore $\dim S \geq 1$. Since $A_1 A_2 \neq A_2 A_1$, we also have $\dim S \leq 1$. Hence $\dim S = \dim \phi^{-1}(B) = 1$ and the upper bound $\dim W + 1$ for the dimension of $\mathrm{GV}(\mathcal{L})$, which is given by Remark 3.2, is attained in our situation. $\square$

5 Convex optimization

In this section we show how Gibbs manifolds arise from entropic regularization in optimization (cf. [21]). We fix an arbitrary linear map $\pi : \mathbb{S}^n \to \mathbb{R}^d$. This can be written in the form

$$\pi(X) = \big(\langle A_1, X \rangle, \langle A_2, X \rangle, \ldots, \langle A_d, X \rangle \big).$$

Here the $A_i \in \mathbb{S}^n$, and $\langle A_i, X \rangle := \mathrm{trace}(A_i X)$. The image $\pi(\mathbb{S}^n_+)$ of the PSD cone $\mathbb{S}^n_+$ under our linear map π is a *spectrahedral shadow*. Here it is a full-dimensional semialgebraic convex cone in $\mathbb{R}^d$. Interestingly, $\pi(\mathbb{S}^n_+)$ can fail to be closed, as explained in [13].

Semidefinite programming (SDP) is the following convex optimization problem:

$$\text{Minimize} \quad \langle C, X \rangle \quad \text{subject to} \quad X \in \mathbb{S}^n_+ \text{ and } \pi(X) = b. \tag{15}$$

See e.g. [16, Chapter 12]. The instance (15) is specified by the cost matrix $C \in \mathbb{S}^n$ and the right hand side vector $b \in \mathbb{R}^d$. The feasible region $\mathbb{S}^n_+ \cap \pi^{-1}(b)$ is a *spectrahedron*. The SDP problem (15) is feasible, i.e. the spectrahedron is non-empty, if and only if b is in $\pi(\mathbb{S}^n_+)$.

Consider the LSSM $\mathcal{L} = \mathrm{span}_{\mathbb{R}}(A_1, \ldots, A_d)$. We usually assume that $\mathcal{L}$ contains a positive definite matrix. This hypothesis ensures that each spectrahedron $\pi^{-1}(b)$ is compact.

As a natural extension of [21, eqn (2)], we now define the entropic regularization of SDP:

$$\text{Minimize} \quad \langle C, X \rangle - \epsilon \cdot h(X) \quad \text{subject to} \quad X \in \mathbb{S}^n_+ \text{ and } \pi(X) = b. \tag{16}$$

Here $\epsilon > 0$ is a small parameter, and h denotes the *von Neumann entropy*, here defined as

$$h : \mathbb{S}^n_+ \to \mathbb{R}, \ X \mapsto \mathrm{trace}\big(X - X \cdot \log(X) \big).$$

We note that h is invariant under the action of the orthogonal group on $\mathbb{S}^n_+$. This implies $h(X) = \sum_{i=1}^n (\lambda_i - \lambda_i \log(\lambda_i))$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of X. Hence the von Neumann entropy h is the matrix version of the entropy function on $\mathbb{R}^n_+$ used in [21].

Our next result makes the role of Gibbs manifolds in semidefinite programming explicit. The following ASSM is obtained by incorporating ϵ and the cost matrix C into the LSSM:

$$\mathcal{L}_\epsilon \; := \; \mathcal{L} - \frac{1}{\epsilon} C \quad \text{for any } \epsilon > 0.$$

Here we allow the case $\epsilon = \infty$, where the dependency on C disappears and the ASSM is simply the LSSM, i.e. $\mathcal{L}_\infty = \mathcal{L}$. The following theorem is the main result in this section.

Theorem 5.1. *For $b \in \pi(\mathbb{S}_+^n)$, the intersection of $\pi^{-1}(b)$ with the Gibbs manifold $\mathrm{GM}(\mathcal{L}_\epsilon)$ consists of a single point X_ϵ^*. This point is the optimal solution to the regularized SDP (16). For $\epsilon = \infty$, it is the unique maximizer of von Neumann entropy on the spectrahedron $\pi^{-1}(b)$.*

The importance of this result for semidefinite programming lies in taking the limit as ϵ tends to zero. This limit $\lim_{\epsilon \to 0} X_\epsilon^*$ exists and it is an optimal solution to (15). The optimal solution is unique for generic C. Entropic regularization is about approximating that limit.

Remark 5.2. Theorem 5.1 implies that adding the condition $X \in \mathrm{GV}(\mathcal{L}_\epsilon)$ to (16) leaves the optimizer unchanged. Hence, if we know equations for the Gibbs variety, we may shrink the feasible region by adding polynomial constraints. Most practical are the affine-linear equations: imposing $X \in \mathrm{GP}(\mathcal{L}_\epsilon)$ allows to solve (16) on a spectrahedron of lower dimension.

To prove Theorem 5.1, we derive two key properties of the von Neumann entropy:

Proposition 5.3. *The function h satisfies:*

 (a) h is strictly concave on the PSD cone $\mathbb{S}_+^n$, and
 (b) the gradient of h is the negative matrix logarithm: $\nabla(h)(X) = -\log(X)$.

Proof. For (a), we use a classical result by Davis [4]. The function h is invariant in the sense that its value $h(X)$ depends on the eigenvalues of X. In fact, it is a symmetric function of the n eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. This function equals $h(\lambda_1, \lambda_2, \ldots, \lambda_n) = \sum_{i=1}^n (\lambda_i - \lambda_i \log(\lambda_i))$, and this is a concave function $\mathbb{R}_+^n \to \mathbb{R}$. The assertion hence follows from the theorem in [4].

For (b) we prove a more general result. For convenience, we change variables $Y = X - \mathrm{id}_n$ so that $f(Y) = h(Y + \mathrm{id}_n)$ is analytic at $Y = 0$. Fix any function $f : \mathbb{R} \to \mathbb{R}$ that is analytic in a neighborhood of the origin. Then $Y \mapsto \mathrm{trace}(f(Y))$ is a well-defined real-valued analytic function of $n \times n$ matrices $Y = (y_{ij})$ that are close to zero. The gradient of this function is the $n \times n$ matrix whose entries are the partial derivatives $\partial \mathrm{trace}(f(Y))/\partial y_{ij}$. We claim that

$$\nabla \mathrm{trace}(f(Y)) \; = \; f'(Y^\top). \tag{17}$$

Both sides are linear in f, and f is analytic, so it suffices to prove this for monomials, i.e.

$$\nabla \mathrm{trace}(Y^k) \; = \; k \cdot (Y^\top)^{k-1} \qquad \text{for all integers } k \geq 1. \tag{18}$$

Note that $\mathrm{trace}(Y^k)$ is a homogeneous polynomial of degree k in the matrix entries y_{ij}, namely it is the sum over all products $y_{i_1 i_2} y_{i_2 i_3} \cdots y_{i_{k-2} i_{k-1}} y_{i_{k-1} i_1}$ that represent closed walks in the complete graph on k nodes. When taking the derivative $\partial/\partial y_{ij}$ of that sum, we obtain

k times the sum over all walks that start at node j and end at node i. Here each walk occurs with the factor k because y_{ij} can be inserted in k different ways to create one of the closed walks above. This polynomial of degree $k-1$ is the entry of the matrix power Y^{k-1} in row j and column i, so it is the entry of its transpose $(Y^\top)^{k-1}$ is row i and column j. To prove the proposition, we now apply (17) to the function $f(y) = (y+1) - (y+1) \cdot \log(y+1)$. $\quad\square$

If $\mathcal{L} = \mathcal{D}$ consists of diagonal matrices then the Gibbs manifold $\mathrm{GM}(\mathcal{D})$ is a discrete exponential family [23, §6.2], and $\pi(\mathrm{GM}(\mathcal{D}))$ is the associated convex polytope. This uses the moment map from toric geometry [16, Theorem 8.24]. In particular, if the linear space $\mathcal{D}$ is defined over $\mathbb{Q}$ then the polytope is rational and the Zariski closure of $\mathrm{GM}(\mathcal{D})$ is the toric variety of that polytope. If the space $\mathcal{D}$ is not defined over $\mathbb{Q}$ then $\mathrm{GM}(\mathcal{D})$ is an analytic toric manifold, whose Zariski closure is the larger toric variety $\mathrm{GV}(\mathcal{D}) = \mathrm{GM}(\mathcal{D}_{\mathbb{Q}})$ seen in (7).

The key step to proving Theorem 5.1 is a non-abelian version of the toric moment map.

Theorem 5.4. *The restriction of the linear map $\pi : \mathbb{S}_+^n \to \mathbb{R}^d$ to the Gibbs manifold $\mathrm{GM}(\mathcal{L})$ defines a bijection between $\mathrm{GM}(\mathcal{L})$ and the open spectrahedral shadow $\mathrm{int}(\pi(\mathbb{S}_+^n))$ in $\mathbb{R}^d$.*

Proof. Fix an arbitrary positive definite matrix $X \in \mathrm{int}(\mathbb{S}_+^n)$ and set $b = \pi(X)$. We must show that the spectrahedron $\pi^{-1}(b)$ contains precisely one point that lies in $\mathrm{GM}(\mathcal{L})$.

Consider the restriction of the von Neumann entropy h to the spectrahedron $\pi^{-1}(b)$. This restriction is strictly concave on the convex body $\pi^{-1}(b)$ by Proposition 5.3. Therefore h attains a unique maximum X^* in the relative interior of $\pi^{-1}(b)$. The first order condition at this maximum tells us that $\nabla(h)(X^*) = -\log(X^*)$ lies in $\mathcal{L}$, which is the span of the gradients of the constraints $\langle A_i, X \rangle = b_i$. Hence, the optimal matrix X^* lies in the Gibbs manifold

$$\mathrm{GM}(\mathcal{L}) \;=\; \left\{ X \in \mathbb{S}_+^n : \log(X) \in \mathcal{L} \right\}.$$

The assignment $b \mapsto X^* = X^*(b)$ is well defined and continuous on the interior of the cone $\pi(\mathbb{S}_+^n)$. We have shown that it is a section of the linear map π, which means $\pi(X^*(b)) = b$. We conclude that π defines a homeomorphism between $\mathrm{GM}(\mathcal{L})$ and $\mathrm{int}(\pi(\mathbb{S}_+^n))$. $\quad\square$

Proof of Theorem 5.1. For any fixed $\epsilon > 0$, any minimizer $X^* = X_\epsilon^*$ of the regularized problem (16) in the spectrahedron $\pi^{-1}(b)$ must satisfy $C + \epsilon \cdot \log(X^*) \in \mathcal{L}$. This is equivalent to $X^* \in \mathrm{GM}(\mathcal{L}_\epsilon)$, and it follows from the first order optimality conditions. By the same convexity argument as in the proof of Theorem 5.4, the objective function in (16) has only one critical point in the spectrahedron $\pi^{-1}(b)$. This implies $\pi^{-1}(b) \cap \mathrm{GM}(\mathcal{L}_\epsilon) = \{X_\epsilon^*\}$. $\quad\square$

We can now turn the discussion around and offer a definition of Gibbs manifolds and Gibbs varieties purely in terms of convex optimization. Fix any LSSM $\mathcal{L}$ of dimension d in $\mathbb{S}^n$. This defines a canonical linear map $\pi : \mathbb{S}_+^n \to \mathbb{S}^n/\mathcal{L}^\perp \simeq \mathbb{R}^d$. Each fiber $\pi^{-1}(b)$ is a spectrahedron. If this is non-empty then the entropy $h(X)$ has a unique maximizer $X^*(b)$ in $\pi^{-1}(b)$. The Gibbs manifold $\mathrm{GM}(\mathcal{L})$ is the set of these entropy maximizers $X^*(b)$ for $b \in \mathbb{R}^d$. The Gibbs variety $\mathrm{GV}(\mathcal{L})$ is defined by all polynomial constraints satisfied by these $X^*(b)$.

This extends naturally to any ASSM $A_0 + \mathcal{L}$. We now maximize the concave function $h(X) + \langle A_0, X \rangle$ over the spectrahedra $\pi^{-1}(b)$. The Gibbs manifold $\mathrm{GM}(A_0 + \mathcal{L})$ collects all maximizers, and the Gibbs variety $\mathrm{GV}(A_0 + \mathcal{L})$ is defined by their polynomial constraints.

Example 5.5. Let $\mathcal{L}$ denote the space of all Hankel matrices $[y_{i+j-1}]_{1 \le i,j \le n}$ in $\mathbb{S}^n$. This LSSM has dimension $d = 2n - 1$. The linear map $\pi : \mathbb{S}^n_+ \to \mathbb{R}^d$ takes any positive definite matrix X to a nonnegative polynomial $b = b(t)$ in one variable t of degree $2n - 2$. We have $b(t) = (1, t, \ldots, t^{n-1}) X (1, t, \ldots, t^{n-1})^\top$, so the matrix X gives a sum-of-squares (SOS) representation of $b(t)$. The fiber $\pi^{-1}(b)$ is the *Gram spectrahedron* [20] of the polynomial b. The entropy maximizer $X^*(b)$ in the Gram spectrahedron is a favorite SOS representation of b. The Gibbs manifold $\mathrm{GM}(\mathcal{L})$ gathers the favorite SOS representations for all nonnegative polynomials b. The Gibbs variety $\mathrm{GV}(\mathcal{L})$, which has dimension $\le 3n - 2$, is the tightest outer approximation of $\mathrm{GM}(\mathcal{L})$ that is definable by polynomials in the matrix entries.

In Example 3.4 we saw a variant of $\mathcal{L}$, namely the sub-LSSM where the upper left entry of the Hankel matrix was fixed to be zero. If $C = -E_{11}$ is the corresponding negated matrix unit, then (15) is the problem of minimizing $b(t)$ over $t \in \mathbb{R}$. See [16, Section 12.3] for a first introduction to polynomial optimization via SOS representations. It would be interesting to explore the potential of the entropic regularization (16) for polynomial optimization. $\qquad \diamond$

One of the topics of [21] was a scaling algorithm for solving the optimization problem (16) for linear programming (LP), i.e. the case when $A_1, \ldots, A_d$ are diagonal matrices. This algorithm extends the Darroch-Ratcliff algorithm for Iterative Proportional Fitting in statistics. Combining this with a method for driving ϵ to zero leads to a numerical algorithm for large-scale LP problems, such as the optimal transport problems in [21, Section 3].

We are hopeful that the scaling algorithm can be extended to the problem (16) in full generality. By combining this with a method for driving ϵ to zero, one obtains a numerical framework for solving SDP problems such as quantum optimal transport in Section 6.

One important geometric object for SDP is the limiting Gibbs manifold, $\lim_{\epsilon \to 0} \mathrm{GM}(\mathcal{L}_\epsilon)$. This is the set of optimal solutions, as b ranges over $\mathbb{R}^d$. In the case of LP, with C generic, it is the simplicial complex which forms the regular triangulation given by C. This reveals the combinatorial essence of entropic regularization of LP, as explained in [21, Theorem 7]. From the perspective of *positive geometry*, it would be worthwhile to study $\lim_{\epsilon \to 0} \mathrm{GM}(\mathcal{L}_\epsilon)$ for SDP. This set is semialgebraic, and it defines a nonlinear subdivision of the spectrahedral shadow $\pi(\mathbb{S}^n_+)$. If we vary the cost matrix C, the theory of *fiber bodies* in [15] becomes relevant.

6 Quantum optimal transport

In this section we examine a semidefinite programming analogue of the classical optimal transport problem, known as *quantum optimal transport* (QOT). We follow the presentation by Cole, Eckstein, Friedland, and Zyczkowski in [3]. Our notation for the dimensions is as in [21, Section 3.1]. We consider the space $\mathbb{S}^{d_1 d_2}$ of real symmetric matrices X of size $d_1 d_2 \times d_1 d_2$. Rows and columns are indexed by $[d_1] \times [d_2]$. Thus, we write $X = (x_{ijkl})$, where (i, j) and (k, l) are in $[d_1] \times [d_2]$. The matrix being symmetric means that $x_{ijkl} = x_{klij}$ for all indices. Each such matrix is mapped to a pair of two *partial traces* by the following linear map:

$$\mathbb{S}^{d_1 d_2} \to \mathbb{S}^{d_1} \times \mathbb{S}^{d_2}, \quad X \mapsto (Y, Z),$$

where the $d_1 \times d_1$ matrix $Y = (y_{ik})$ satisfies $y_{ik} = \sum_{j=1}^{d_2} x_{ijkj}$, and the $d_2 \times d_2$ matrix $Z = (z_{jl})$ satisfies $z_{jl} = \sum_{i=1}^{d_1} x_{ijil}$. If X is positive semidefinite then so are its partial traces Y and Z.

Hence our *marginalization map* restricts to a linear projection of closed convex cones, denoted

$$\mu : \mathbb{S}_+^{d_1 d_2} \rightarrow \mathbb{S}_+^{d_1} \times \mathbb{S}_+^{d_2} , \quad X \mapsto (Y, Z). \tag{19}$$

Diagonal matrices in $\mathbb{S}_+^{d_1 d_2}$ can be identified with rectangular matrices of format $d_1 \times d_2$ whose entries are nonnegative. The map μ takes such a rectangular matrix to its row sums and column sums. Hence the restriction of μ to diagonal matrices in $\mathbb{S}_+^{d_1 d_2}$ is precisely the linear map that defines classical optimal transport in the discrete setting of [21, Section 3.1].

The quantum optimal transportation problem (QOT) is the task of minimizing a linear function $X \mapsto \langle C, X \rangle$ over any transportation spectrahedron $\mu^{-1}(Y, Z)$. This is an SDP. Our main theorem in this section states that its Gibbs manifold is semialgebraic.

Theorem 6.1. *The Gibbs manifold* $\mathrm{GM}(\mathcal{L})$ *for QOT is a semialgebraic subset of* $\mathbb{S}_+^{d_1 d_2}$. *It consists of all symmetric matrices* $Y \otimes Z$, *where* $Y \in \mathbb{S}_+^{d_1}$ *and* $Z \in \mathbb{S}_+^{d_2}$. *The Gibbs variety* $\mathrm{GV}(\mathcal{L}) \subset \mathbb{S}^{d_1 d_2}$ *is linearly isomorphic to the cone over the Segre variety* $\mathbb{P}^{\binom{d_1+1}{2}-1} \times \mathbb{P}^{\binom{d_2+1}{2}-1}$.

The image of the marginalization map μ generalizes the polytope $\Delta_{d_1-1} \times \Delta_{d_2-1}$, and the fibers of μ are quantum versions of transportation polytopes. These shapes are now nonlinear.

Lemma 6.2. *The image of the map* μ *is a convex cone of dimension* $\binom{d_1+1}{2} + \binom{d_2+1}{2} - 1$:

$$\mathrm{image}(\mu) = \left\{ (Y, Z) \in \mathbb{S}_+^{d_1} \times \mathbb{S}_+^{d_2} : \mathrm{trace}(Y) = \mathrm{trace}(Z) \right\}. \tag{20}$$

For any point (Y, Z) *in the relative interior of this cone, the* transportation spectrahedron $\mu^{-1}(Y, Z)$ *is a compact convex body of dimension* $\frac{1}{2}(d_1 - 1)(d_2 - 1)(d_1 d_2 + d_1 + d_2 + 2)$.

Proof of Lemma 6.2. The partial trace map μ in (19) restricts to tensor products as follows:

$$\mu(Y \otimes Z) = \left(\mathrm{trace}(Z) \cdot Y , \mathrm{trace}(Y) \cdot Z \right). \tag{21}$$

Hence, if $Y \in \mathbb{S}_+^{d_1}$ and $Z \in \mathbb{S}_+^{d_2}$ satisfy $t = \mathrm{trace}(Y) = \mathrm{trace}(Z)$ then $\frac{1}{t} Y \otimes Z$ is a positive semidefinite matrix in the fiber $\mu^{-1}(Y, Z)$. This shows that the image is as claimed on the right hand side of (20). The image is a spectrahedral cone of dimension $\binom{d_1+1}{2} + \binom{d_2+1}{2} - 1$. Subtracting this from $\dim \mathbb{S}_+^{d_1 d_2} = \binom{d_1 d_2 + 1}{2}$ yields the dimension of the interior fibers. $\square$

Example 6.3 ($d_1 = d_2 = 2$)**.** The map μ projects positive semidefinite 4×4 symmetric matrices

$$X = \begin{bmatrix} x_{1111} & x_{1112} & x_{1121} & x_{1122} \\ x_{1112} & x_{1212} & x_{1221} & x_{1222} \\ x_{1121} & x_{1221} & x_{2121} & x_{2122} \\ x_{1122} & x_{1222} & x_{2122} & x_{2222} \end{bmatrix}$$

onto a 5-dimensional convex cone, given by the direct product of two disks. The formula is

$$Y = \begin{bmatrix} x_{1111} + x_{1212} & x_{1121} + x_{1222} \\ x_{1121} + x_{1222} & x_{2121} + x_{2222} \end{bmatrix} \quad \text{and} \quad Z = \begin{bmatrix} x_{1111} + x_{2121} & x_{1112} + x_{2122} \\ x_{1112} + x_{2122} & x_{1212} + x_{2222} \end{bmatrix}.$$

The fibers of this map μ are the 5-dimensional transportation spectrahedra $\mu^{-1}(Y, Z)$.

To illustrate the QOT problem, we fix the margins and the cost matrix as follows:

$$Y = \begin{bmatrix} 5 & 1 \\ 1 & 6 \end{bmatrix} \quad \text{and} \quad Z = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 2 & 3 & 5 & 7 \\ 3 & 11 & 13 & 17 \\ 5 & 13 & 23 & 29 \\ 7 & 17 & 29 & 31 \end{bmatrix}. \tag{22}$$

We wish to minimize $\langle C, X \rangle$ subject to $\mu(X) = (Y, Z)$. The optimal solution X^* is equal to

$$\begin{bmatrix} 3.5791289951969672555885181314 & 2.1481033873373327210111731020 & 2.6712549910317892812299265149 & -2.0756620454202478990696017 \\ 2.1481033873373327210111731020 & 1.4208710048030274444114818686 & 1.16978382139276763200240537 & -1.6712549910317892812299265149 \\ 2.6712549910317892812299265149 & 1.169783821392767632002405371 & 3.4208710048030274444114818686 & -0.14810338733733272101173102 \\ -2.0756620454202478990696017 & -1.6712549910317892812299265149 & -0.14810338733733272101173102 & 2.5791289951969672555885181314 \end{bmatrix}.$$

This matrix has rank 2. The optimal value equals $v = 156.964485798827271035367539305....$ This is an algebraic number of degree 12. Its exact representation is the minimal polynomial

$$125v^{12} - 465480v^{11} + 770321646v^{10} - 744236670798v^9 + 463560077206539v^8 - 193865445786866004v^7$$
$$+ 54901023652716544539v^6 - 10330064181552258647604v^5 + 1219620644420527588643307v^4$$
$$- 77994100149206862070472310v^3 + 1395374211380010273312826701v^2$$
$$+ 83502957914204004050312708316v - 2047417613706778627978564647804 \;=\; 0.$$

This was derived from the KKT equations in [17, Theorem 3]. We conclude that the algebraic degree of QOT for $d_1 = d_2 = 2$ is equal to 12. This is smaller than the algebraic degree of semidefinite programming, which is 42. That is the entry for $m=5$ and $n=4$ in [17, Table 2].
    This drop arises because QOT is a very special SDP. The LSSM for our QOT problem is

$$\mathcal{L} = \left\{ \begin{bmatrix} y_1 + y_3 & y_5 & y_4 & 0 \\ y_5 & y_1 & 0 & y_4 \\ y_4 & 0 & y_2 + y_3 & y_5 \\ 0 & y_4 & y_5 & y_2 \end{bmatrix} \;:\; y_1, y_2, y_3, y_4, y_5 \in \mathbb{R} \right\}. \tag{23}$$

This defines our 5-dimensional Gibbs manifold $\mathrm{GM}(\mathcal{L})$ in the 10-dimensional cone $\mathbb{S}_+^4$. Theorem 6.1 states that it equals the positive part of the Gibbs variety, i.e. $\mathrm{GM}(\mathcal{L}) = \mathrm{GV}(\mathcal{L}) \cap \mathbb{S}_+^4$.
    We compute the entropy maximizer inside the 5-dimensional transportation spectrahedron $\mu^{-1}(Y, Z)$ for the marginal matrices Y and Z in (22). Notably, its entries are rational:

$$\mu^{-1}(Y,Z) \cap \mathrm{GV}(\mathcal{L}) \;=\; \mu^{-1}(Y,Z) \cap \mathrm{GM}(\mathcal{L}) \;=\; \left\{ \frac{1}{11} \begin{bmatrix} 35 & 10 & 7 & 2 \\ 10 & 20 & 2 & 4 \\ 7 & 2 & 42 & 12 \\ 2 & 4 & 12 & 24 \end{bmatrix} \right\}. \qquad \diamond$$

Proof of Theorem 6.1. By linear extension, the equation (21) serves as a definition of the marginalization map μ on $\mathbb{S}^{d_1 d_2}$. We observe the following for the trace inner product on $\mathbb{S}^{d_1 d_2}$:

$$\mathrm{trace}\big((A \otimes \mathrm{id}_{d_2})(Y \otimes Z)\big) = \mathrm{trace}(Z) \cdot \mathrm{trace}(AY) \quad \text{for all } A \in \mathbb{S}^{d_1}$$
$$\text{and} \quad \mathrm{trace}\big((\mathrm{id}_{d_1} \otimes B)(Y \otimes Z)\big) = \mathrm{trace}(Y) \cdot \mathrm{trace}(BZ) \quad \text{for all } B \in \mathbb{S}^{d_2}.$$

Therefore, the (i, j) entry of $\mathrm{trace}(Z) \cdot Y$ is obtained as $\frac{1}{2}\langle (E_{ij} + E_{ji}) \otimes \mathrm{id}_{d_2}, Y \otimes Z \rangle$, where E_{ij} is the (i, j)-th matrix unit. A similar observation holds for the entries of $\mathrm{trace}(Y) \cdot Z$. This

means that $\mu(X)$ is computed by evaluating $\mathrm{trace}\big((A \otimes \mathrm{id}_{d_2})X\big)$ and $\mathrm{trace}\big((\mathrm{id}_{d_1} \otimes B)X\big)$, where A ranges over a basis of $\mathbb{S}^{d_1}$ and B ranges over a basis of $\mathbb{S}^{d_2}$. Therefore, we have

$$\mathcal{L} \;=\; \big\{\, A \otimes \mathrm{id}_{d_2} \,+\, \mathrm{id}_{d_1} \otimes B \;:\; A \in \mathbb{S}^{d_1} \text{ and } B \in \mathbb{S}^{d_2} \,\big\}. \tag{24}$$

Now, the key step in the proof consists of the following formula for the matrix logarithm

$$\log(Y \otimes Z) \;=\; \log(Y) \otimes \mathrm{id}_{d_2} \,+\, \mathrm{id}_{d_1} \otimes \log(Z).$$

This holds for positive semidefinite matrices Y and Z, and it is verified by diagonalizing these matrices. By setting $Y = \exp(A)$ and $Z = \exp(B)$, we now conclude that the Gibbs manifold $\mathrm{GM}(\mathcal{L})$ consists of all tensor products $Y \otimes Z$ where $Y \in \mathbb{S}_+^{d_1}$ and $Z \in \mathbb{S}_+^{d_2}$.

We have shown that $\mathrm{GM}(\mathcal{L})$ is the intersection of a variety with $\mathbb{S}_+^{d_1 d_2}$. This variety must be the Gibbs variety $\mathrm{GV}(\mathcal{L})$. More precisely, $\mathrm{GV}(\mathcal{L})$ consists of all tensor products $Y \otimes Z$ where Y, Z are complex symmetric. This is the cone over the Segre variety, which is the projective variety in $\mathbb{P}^{\binom{d_1 d_2 + 1}{2} - 1}$ whose points are the tensor products $X = Y \otimes Z$. $\qquad\square$

We have the following immediate consequence of the proof of Theorem 6.1. The entropy maximizers have rational entries. This explains the matrix at the end of Example 6.3

Corollary 6.4. *The Gibbs point for QOT is given by $\frac{Y \otimes Z}{\mathrm{trace}(Y)}$, with Y, Z the given margins.*

At this point, it pays off to revisit Section 3 and to study its thread for the LSSM in (24).

Example 6.5. We apply Algorithm 1 to the LSSM $\mathcal{L}$ in (23). The eigenvalues of $\mathcal{L}$ are distinct, and the ideal $\langle E_1' \rangle$ in step 2 is the intersection of six prime ideals. One of them is

$$\langle\, \lambda_1 + \lambda_2 - y_1 - y_2 - y_3 \,,\; \lambda_3 + \lambda_4 - y_1 - y_2 - y_3 \,,$$
$$2\lambda_2\lambda_4 - \lambda_2 y_1 - \lambda_4 y_1 - \lambda_2 y_2 - \lambda_4 y_2 + 2y_1 y_2 - \lambda_2 y_3 - \lambda_4 y_3 + y_1 y_3 + y_2 y_3 + y_3^2 - 2y_4^2 + 2y_5^2,$$
$$\lambda_2^2 + \lambda_4^2 - \lambda_2 y_1 - \lambda_4 y_1 - \lambda_2 y_2 - \lambda_4 y_2 + 2y_1 y_2 - \lambda_2 y_3 - \lambda_4 y_3 + y_1 y_3 + y_2 y_3 - 2y_4^2 - 2y_5^2 \,\rangle.$$

The other five associated primes are found by permuting indices of $\lambda_1, \lambda_2, \lambda_3, \lambda_4$. Hence, the Galois group $G_{\mathcal{L}}$ is the Klein four-group $S_2 \times S_2$ in S_4, and we infer the linear relation $\lambda_1 + \lambda_2 - \lambda_3 - \lambda_4$. The set E_3 in step 7 is the singleton $\{z_1 z_2 - z_3 z_4\}$. The elimination in step 12 reveals the prime ideal in $\mathbb{R}[X]$ that is shown for arbitrary d_1, d_2 in Corollary 6.6. $\quad\diamond$

Our final result is derived from Theorem 6.1 using tools of toric algebra [16, Chapter 8].

Corollary 6.6. *The Gibbs variety for QOT is parametrized by monomials $x_{ijkl} = y_{ik} z_{jl}$ that are not all distinct. Its prime ideal in $\mathbb{R}[X]$ is minimally generated by the 2×2 minors of a matrix of format $\binom{d_1+1}{2} \times \binom{d_2+1}{2}$, together with $\binom{d_1}{2}\binom{d_2}{2}$ linear forms in the entries of X.*

We propose to extend QOT to quantum graphical models [26]. In statistics, every undirected graph G on s vertices defines such a model [23, Section 13.2]. The graphical model lives in the probability simplex $\Delta_{d_1 d_2 \cdots d_s - 1}$. Its points are nonnegative tensors of format $d_1 \times d_2 \times \cdots \times d_s$ whose entries sum to 1. The quantum graphical model lives in the high-dimensional PSD cone $\mathbb{S}_+^{d_1 d_2 \cdots d_s}$, where the marginalization records the partial trace for every clique in G. It would be interesting to study the Gibbs manifold and the Gibbs varieties for these models. One may ask whether they agree for all graphs G that are decomposable. By Theorem 6.1, this holds for QOT, where G is the graph with two nodes and no edges.

Acknowledgements

References

[1] J. Ax: *On Schanuel's conjectures*, Annals of Mathematics **93** (1971) 252–268.

[2] E. Briand: *When is the algebra of multisymmetric polynomials generated by the elementary multisymmetric polynomials?*, Beiträge zur Algebra und Geometrie **45** (2004) 353–368.

[3] S. Cole, M. Eckstein, S. Friedland and K. Zyczkowski: *Quantum Monge-Kantorovich problem and transport distance between density matrices*, Phys. Rev. Lett. **129** (2022) 110402.

[4] C. Davis: *All convex invariant functions of hermitian matrices*, Arch. Math. **8** (1957) 276–278.

[5] P. Diaconis and B. Sturmfels: *Algebraic algorithms for sampling from conditional distributions*, Annals of Statistics **26** (1998) 363–397.

[6] C. Fevola, Y. Mandelshtam and B. Sturmfels: *Pencils of quadrics: old and new*, Le Matematiche **76** (2021) 319–335.

[7] J. Forsgård and T. de Wolff: *The algebraic boundary of the sonc-cone*, SIAM J. Appl. Algebra Geom. **6** (2022) 468–502.

[8] F. Galuppi and M. Stanojkovski: *Toric varieties from cyclic matrix semigroups*, Rend. Istit. Mat. Univ. Trieste **53** (2021), Art. No. 17.

[9] K. Girstmair: *Linear dependence of zeros of polynomials and construction of primitive elements*, Manuscripta Math. **39** (1982) 81–97.

[10] K. Girstmair: *Linear relations between roots of polynomials*, Acta Arith. **89** (1999) 53–96.

[11] J. Hauenstein, J. Rodriguez and F. Sottile: *Numerical computation of Galois groups*, Found. Comput. Math. **18** (2018) 867–890.

[12] R. Horn and C. Johnson: *Topics in Matrix Analysis*, Cambridge University Press, 1991.

[13] Y. Jiang and B. Sturmfels: *Bad projections of the PSD cone*, Collectanea Mathematica **72** (2021) 261–280.

[14] Y. Kitaoka: *Notes on the distribution of roots modulo a prime of a polynomial*, Uniform Distribution Theory **12** (2017) 91-117.

[15] L. Mathis and C. Meroni: *Fiber convex bodies*, Discrete Comput. Geom. (2022), online first.

[16] M. Michałek and B. Sturmfels: *Invitation to Nonlinear Algebra*, Graduate Studies in Mathematics, vol 211, American Mathematical Society, Providence, 2021.

[17] J. Nie, K. Ranestad and B. Sturmfels: *The algebraic degree of semidefinite programming*, Mathematical Programming **122** (2010) 379–405.

[18] The OSCAR Team: *OSCAR – Open Source Computer Algebra Research system, Version 0.11.0*, https://oscar.computeralgebra.de, 2022.

[19] L. Pachter and B. Sturmfels: *Algebraic Statistics for Computational Biology*, Cambridge University Press, 2005.

[20] C. Scheiderer: *Extreme points of Gram spectrahedra of binary forms*, Discrete Comput. Geom. **67** (2022) 1174–1190.

[21] B. Sturmfels, S. Telen, F.-X. Vialard and M. von Renesse: *Toric geometry of entropic regularization*, presented at MEGA 2022, Kraków, `arXiv:2202.01571`.

[22] B. Sturmfels and C. Uhler: *Multivariate Gaussians, semidefinite matrix completion, and convex algebraic geometry*, Annals of the Institute of Statistical Mathematics **62** (2010) 603–638.

[23] S. Sullivant: *Algebraic Statistics*, Graduate Studies in Mathematics, 194, American Mathematical Society, Providence, RI, 2018.

[24] J. J. Sylvester: *On the equation to the secular inequalities in the planetary theory*, Philosophical Magazine Series 5, 16:100 (1883) 267-269.

[25] E. Vigoda: *Sampling from Gibbs distributions*, PhD Dissertation, Computer Science Dept., UC Berkeley, 1999.

[26] S. Weis and J. Gouveia: *Quantum marginals, faces, and coatoms*, `arXiv:2103.08360`.

Authors' addresses:

Dmitrii Pavlov, MPI-MiS Leipzig   dmitrii.pavlov@mis.mpg.de

Bernd Sturmfels, MPI-MiS Leipzig and UC Berkeley   bernd@mis.mpg.de

Simon Telen, CWI Amsterdam and MPI-MiS Leipzig   Simon.Telen@cwi.nl