WHAT IS A MARKOV BASIS AND WHAT IS IT GOOD FOR?

THOMAS KAHLE

Let $\Bbbk$ be a field and $R = \Bbbk[x_1, \dots, x_n]$ the polynomial ring in n indeterminates.

Definition 1. Let $A \subseteq \mathbb{Z}^{d\times n}$ be an integer matrix. The *toric ideal for A* is

$$I_A := \langle x^u - x^v : u - v \in \ker_{\mathbb{Z}} A\rangle \subseteq R.$$

Find a minimal (or just finite) generating set of I_A. As you have seen in other lectures, a lattice basis of $\ker_{\mathbb{Z}}(A)$ is usually not sufficient. Generating sets have interesting applications to *random walks on discrete objects* and because Markov chains are used in these applications, there came about the name *Markov basis* (which is typically used for the exponents of generators).

Hypothesis testing in statistics. Think of two random variables that take only finitely many values. These could for example be traits of the individuals in a population, such as their gender, color of their hair, or the number of hours they watch sports on TV every week. For maximum simplicity, let $X = (X_1, X_2)$ be a vector of only two random variables, taking values in $[r] \times [s]$. In a population, such as the population of Romania, or Europe, there is a true distribution of X and we want to learn something about this true distribution from a small *sample* (this is statistics!).

Now, the simplest question about this data is to ask: Is there evidence against the hypothesis that X_1 is independent of X_2? In statistics we never prove things, we just *statistically disprove things* using an argument like: If the the hypothesis was true, what would typical data look like? If, under the hypothesis, it is very unlikely to get the data that looks like the data we got, then this presents evidence against the hypothesis (although it could be the case that we just got unlucky.)

We can't argue the other way around, because if we get data that confirms the hypothesis, then maybe the hypothesis was just a very weak hypothesis so that no data would actually refute it. There is no systematic way to rule this out, so we don't argue in this way.

Let's assume independence of X_1 and X_2. This defines a statistical model, a subset of all joint distributions for X. It turns out that this subset is defined by 2-minors and equals the non-negative real part of the Segre embedding, but this is a different story.

Date: August 2016.

We then determine the best explanation of the data within the model of independence using the *maximum likelihood method*. The best explanation is the disitribution under which our given data has the highest probability. When you analyze this, it turns out that the determination of the maximum likelihood estimate does not actually depend on the entries of the table, but only on the marginals defined as follows. The sample data is given in form of a *contingency table*

$$U = \begin{pmatrix} u_{11} & \dots & u_{1s} \\ \vdots & & \vdots \\ u_{r1} & \dots & u_{rs} \end{pmatrix}$$

containing counts u_{ij} which give the number of observations $X_1 = i, X_2 = j$. The *marginals* of U are the row sums and column sums:

$$\begin{array}{ccc|c} u_{11} & \dots & u_{1s} & u_{1+} \\ \vdots & & \vdots & \vdots \\ u_{r1} & \dots & u_{rs} & u_{r+} \\ \hline u_{+1} & \dots & u_{+s} & u_{++} \end{array}$$

That is, there is a linear map $Au = (u_{+1}, u_{+2}, u_{1+}, u_{2+})$. If X_1 is independent of X_2, then the count in position ij should be proportional to the product $u_{i+}u_{+j}$ just as the distribution factorizes. Now the big question is: How good is the maximum likelihood estimate? It explains the data best among the distributions in the model, but how good is that really? We can clearly measure the distance of the actual data to the estimate using an appropriate norm (it's called the X^2-statistics, a variant of L_2). Now assume you get a distance of 5. How big is 5? As a reference measure, Fisher proposed to compare this 5 with the values gotten for other fake data that would lead to the same estimate. To carry out this test, we therefore need to generate fake data tables v that have the same marginals, that is, such that $Au = Av$. We need points from

Definition 2. Let $\mathbf{b} \in \mathbb{Z}^d$. The *fiber of* $\mathbf{b}$ is

$$A^{-1}[\mathbf{b}] := \{v \in \mathbb{N}^n : Av = \mathbf{b}\}$$

Note that the fiber is the set of interger points in a polytope. Fisher did this test for very small Au, such that he could enumerate all tables v with $Av = Au$. The story is told in the popular book " The Lady Tasting Tea" by David Salsburg.

To conclude the method, if we could enumerate the fiber of $\mathbf{b} = Au$, and then compare the value of $X^2(u)$ to that of $X^2(v)$ for all other tables v. We reject the hypothesis if the probability of observing an $X^2(u)$ as high as ours is very low. This probability is the p-value that is given in EVERY SCIENTIFIC STUDY. (You can google "p-value hacking" at this point).

In reality, this is hard, and this is why mainstream statistics studies the asymptotics of $X^2(u)$ as the sample size grows $u_{++} \to \infty$. Then the distribution of $X^2(u)$ converges to a χ^2-distribution (a certain mixture of Gaussians). The assessment is then done by replacing X^2 by its asymptotic distribution.

But algebra wants to be exact! Diaconis and Sturmels idea: Use a Markov chain to sample from the fiber. Starting from u, if we have a sufficiently large set of "moves" to connect the fiber, then we can run a random walk to sample from it.

Definition 3. Let $\mathcal{M} \subseteq \ker_{\mathbb{Z}} A$ be a finite set, and $A^{-1}[\mathbf{b}]_{\mathcal{M}}$ the graph that has the fiber as its set of vertices and $u \sim v \Leftrightarrow (u-v) \in \pm\mathcal{M}$ as its edges.

$\mathcal{M}$ is a *Markov subbasis for* $\mathbf{b}$ if $A^{-1}[\mathbf{b}]_{\mathcal{M}}$ is connected and *a Markov basis of* A if it is a subbasis for all $\mathbf{b}$.

Why should a finite Markov basis exist? The answer is the Noetherianity of polynomial rings, as we will demonstrate now. Let us abstract the situation a bit. Considering any finite set $\mathcal{M} \subseteq \mathbb{Z}^n$ that spans a saturated lattice (that is, one that is equal to $\ker_{\mathbb{Z}}(A)$ for some matrix $A \in \mathbb{Z}^{d\times n}$). Given two integer points $u, v \in \mathbb{N}^n$, we can ask if there exists a walk

$$u = u_0, u_1, \ldots, u_s = v, \qquad u_i \in \mathbb{N}^n, u_{i-1} - u_i \in \pm\mathcal{M}. \tag{1}$$

Clearly, a necessary condition of the existence of such a walk is that $u - v \in \ker_{\mathbb{Z}}(A) = \mathbb{Z}\mathcal{M}$. Proposition 4 (probably originally due to [MM82]) gives an algebraic sufficient condition. To see it, we construct an ideal out of $\mathcal{M}$ as follows. Decompose each $m \in \mathcal{M}$ into its positive and negative parts $(m^{\pm})_i = \max\{\pm m_i, 0\}$ such that $m = m^+ - m^-$. Then we can define a binomial $x^{m^+} - x^{m^-}$ in a polynoimal ring with n indeterminates. To the set $\mathcal{M}$ corresponds an ideal

$$I_{\mathcal{M}} := \langle x^{m^+} - x^{m^-} : m \in \mathcal{M} \rangle.$$

Proposition 4. *There exists a nonnegative walk* (1) *between* $u, v \in \mathbb{N}^n$ *if and only if* $x^u - x^v \in I_{\mathcal{M}}$.

Proof. If there exists a nonnegative walk as in (1), then

$$x^u - x^v = x^u - x^{u_1} + x^{u_1} - x^{u_2} + \ldots x^{u_{s-1}} + x^v.$$

Now each step $x^{u_{l-1}} - x^{u_l} = x^w(x^{(u_{l-1}-u_l)^+} - x^{(u_{l-1}-u_l)^-})$ is contained in I, so the conclusion follows. In the other direction, if $x^u - x^v \in I$, then it can be written in terms of the generators as

$$x^u - x^v = \sum_i x^{w_i}(x^{m_i^+} - x^{m_i^+}) \qquad \text{Ex.: Why are monomial coefficients enough?}$$

Comparing coefficients, there must be an index i_0 such that $x^u = x^{w_{i_0}+m^+_{i_0}}$. We use the corresponding binomial as the first step in the walk: $u = w_{i_0} + m^+_{i_0} \to w_{i_0} + m^-_{i_0}$. Now $-x^{w_{i_0}+m^-_{i_0}}$ either equals $-x^v$, or it is cancelled in the sum by a term $x^{w_{i_1}+m^+_{i_1}}$. So either we are done, or we use $w_{i_0} + m^-_{i_0} \to w_{i_0} + m^-_{i_0} = w_{i_1} + m^+_{i_1}$ as the second step in the walk. The result now follows after finitely many applications of this step. □

The whole situation can be interpreted in terms of multigradings. Let $S = \Bbbk[x_1, \ldots, x_n]$ be a shorthand for our polynomial ring. The matrix A defines a (multi-)grading on S where $\deg(x^u) = Au \in \mathbb{Z}^d$. This means in particular that $\deg(x_i)$ equals the i-th column of A. The A-graded Hilbert function of the quotient $S/\langle x^u - x^v : Au = Av\rangle$ takes values only zero and one—the quotient is *finely graded*. We have

Theorem 5 (Fundamental Theorem of Markov bases)**.** *The following are equivalent for finite a set $\mathcal{M} \subseteq \ker_{\mathbb{Z}}(A)$:*

(1) *$\mathcal{M}$ is a Markov basis for A.*

(2) $I_A = I_{\mathcal{M}} = \langle x^{m^+} - x^{m^-} : m \in \mathcal{M}\rangle$.

(3) $I_{\mathcal{M}} = I_{\mathcal{M}} : (\prod_{i=1}^n x_i)^\infty = \{f \in R : x^w f \in I_{\mathcal{M}}$ *for some monomial* $x^w\}$.

Proof. According to Proposition 4, if $\mathcal{M} \subseteq \mathbb{Z}^n$ is a Markov basis, then $x^u - x^v \in I_A$ whenever $Au = Av$. This means that $I_A \subseteq I_{\mathcal{M}}$. The other direction is clear since $\mathcal{M} \subseteq \ker_{\mathbb{Z}}(A)$. The equivalence of 3 follows from the more general fact that $I_A = I_{\mathcal{M}} : (\prod_i x_i)^\infty$ whenever $\mathcal{M}$ spans $\ker_{\mathbb{Z}}(A)$. □

From here you can to into a few different directions. For example:

- Efficient computation of Markov bases → [4ti207].
- Theoretical computation of Markov bases, e.g. toric fiber products [Sul07, RS16, EKS14].
- Convergence of random walks using Markov bases ([Win16] and Tobias' thesis).
- Non-Markov bases, decompositions, connectivity analsys.

Exercise 6. Compute the Markov basis for independent matrices.

References

[4ti207] 4ti2 team, *4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces*, available at www.4ti2.de, 2007.

[EKS14] Alexander Engström, Thomas Kahle, and Seth Sullivant, *Multigraded commutative algebra of graph decompositions*, Journal of Algebraic Combinatorics **39** (2014), no. 2, 335–372.

[MM82] Ernst W. Mayr and Albert A. Meyer, *The complexity of the word problems for commutative semigroups and polynomial ideals*, Advances in Mathematics **46** (1982), no. 3, 305–329.

[RS16] Johannes Rauh and Seth Sullivant, *Lifting Markov bases and higher codimension toric fiber products*, Journal of Symbolic Computation **74** (2016), 276–307.

[Sul07] Seth Sullivant, *Toric fiber products*, J. Algebra **316** (2007), no. 2, 560–577.

[Win16] Tobias Windisch, *Rapid mixing and Markov bases*, to appear in SIAM Discrete Mathematics, arXiv:1505.03018 (2016).

Otto-von-Guericke Universität, Magdeburg, Germany
URL: http://www.thomas-kahle.de