THE THREE-STATE TORIC HOMOGENEOUS MARKOV CHAIN MODEL HAS MARKOV DEGREE TWO

PATRIK NORÉN

Abstract. We prove that the three-state toric homogenous Markov chain model has Markov degree two. In algebraic terminology, that a certain class of toric ideals are generated by quadratic binomials. This was conjectured by Haws, Martin del Campo, Takemura and Yoshida, who proved that the Markov degree was at most six.

1. Introduction

In this note we prove that the Markov degree of three state toric homogeneous Markov chain model is two, as conjectured. Let S and T be positive integers,

$$R_{S,T} = \mathbb{K}[x_w \mid w \text{ is a } T\text{-letter word } i_1 \dots i_T \text{ on the alphabet } [S] \text{ with } i_j \neq i_{j+1}],$$

and define the *S-state toric homogeneous Markov T–chain ideal*, $I_{S,T}$, as the kernel of the ring homomorphism

$$\Phi_{S,T} : R_{S,T} \to R_{S,2}$$

given by $\Phi(x_{i_1 \dots i_T}) = x_{i_1 i_2} x_{i_2 i_3} \cdots x_{i_{T-1} i_T}$. Most results are independent of T when it is sufficiently large. The *Markov degree* is the smallest degree of a generating set, and the *Gröbner degree* is the smallest degree of a Gröbner basis. A brief overview of previous and related results:

Theorem (Hara and Takamura, [4] and [7])**.** *The Markov degree of the two-state model is 2.*

Theorem (D. Haws, A. Martin del Campo, A. Takemura and R. Yoshida, [5], [6])**.** *The Markov degree of the three-state model is at most 6.*

This is the 'big conjecture' according to Ruriko Yoshida [8].

Conjecture. *For $S > 2$ the S-state model has Markov degree $S - 1$ and Gröbner degree S.*

We prove the Markov part of the conjecture for three state model by combinatorial arguments.

Theorem 2.8. *The Markov degree of the three-state model is 2.*

The S-state toric homogeneous Markov T–chain ideal is almost an ideal of graph homomorphisms from the path of length T into the complete graph on S vertices. Thanks to the following result it is believed that their structure should be possible to understand.

Theorem (Engström and Norén [3])**.** *The ideal of graph homomorphisms from any forest into any graph has a square-free quadratic Gröbner basis.*

That results was proved using the toric fiber product, and adaptations of that has been useful before, as in [1] and [2], when the ideals under consideration are not toric fiber products right off.

2. Proof of the main result

Let P_T be the directed path on vertex set $[T]$ and edges $12, 23, \ldots, (T-1)T$. Let K_3 be the directed complete graph on vertex set [3]. A T-letter word $i_1 \ldots i_T$ on the alphabet [3] with $i_j \neq i_{j+1}$ encode a graph homomorphism $P_T \to K_3$, the word $i_1 \cdots i_T$ corresponds to the homomorphism sending vertex j to i_j.

A *state graph* is a directed graph on vertex set [3] with multiple edges allowed but no loops.

Now the variables in $R_{3,T}$ are indexed by graph homomorphisms $P_T \to K_3$. A graph homomorphism $P_T \to K_3$ induce a state graph and if x, y are two variables with the same state graph then $x - y \in I_{3,T}$. It is enough to have one variable for each state graph. A state graph G of a variable can be uniquely decomposed into a set of 2-cycles, triangles with the same orientation and potentially a leftover path directed in the same way as the triangles. Note that any collection of 2-cycles, triangles and a possibly empty path whose number of edges add up to $T-1$ is a decomposition of a state graph of a variable. The graph can be reconstructed from the decomposition and so alternatively there is one variable for each decomposition.

The notation for a path is ij or ijk depending on the length. The notation for 2-cycles is (ij). The notation for triangles is (ijk), this cycle contains the path ijk, note that the orientation or direction matters. The notation for a the step $xy - x'y'$ is $A, B \to A', B'$ where A is a collection of paths and cycles in x and B is a collection of cycles and paths in y, the variable x' is obtained from x by removing the cycles and paths A and adding A' and y' is obtained from y be removing B and adding B'. Note that when this has been done it might be necessary to decompose the the graphs in a new way, for example $(123), (321) \to (321), (123)$ might actually be $(123), (321)(321) \to (321), (12)(23)(13)$.

Example 2.1. The word 123231323123 has the decomposition (13)(23)(23)(123)23.

To prove that the ideals $I_{3,T}$ are generated in degree 2 some lemmas are needed.

Lemma 2.2. *Given a monomial m' it is possible to use degree 2-moves to reach a monomial m so that all triangles of m have the same direction and if x, y divide m then the number of triangles in x and y differ by at most two.*

Furthermore if it is impossible to get all the paths and triangles of the variables dividing m directed in the same way then there is at most one triangle in the variables of m.

Proof. Suppose the monomial m' have variables with triangles directed differently and one variable with more than one triangle then the move $(123), (321) \to (321), (123)$ decrease the number of triangles. In the end either all triangles have the same direction or the variables have at most one triangle. If there still is cycles of different directions and paths on some of the variables with cycles it is possible to reduce the number of cycles in the same way, otherwise there is a move $(123), (132) \to (12)31, (23)13$ reducing the number of cycles of different directions. If all paths and cycles have the same directions the move

$$(ijk)(ijk), (i_1 i_2)(i_3 i_4)(i_4 i_5) \to (i_1 i_2)(i_3 i_4)(i_4 i_5), (ijk)(ijk)$$

reduces the difference of triangles between variables. If there is two or more triangles in some variables and a path with the other direction on some, then the moves $(ijk), kj \to (jk)ij, ki$ or $(ijk), kji \to (ij)ki, (jk)$ might be needed together with swapping the extra paths to reduce the number of triangles and paths directed the wrong way. If the variable with triangles have a path with one edge then the moves above are enough since the variable with the wrong directed path have at least six 2-cycles in this case. □

Call the monomials m in Lemma 2.2 *normal*.

Lemma 2.3. *If m, n are normal monomials so that $m - n \in I_{3,T}$ and m has all paths and cycles directed in the same way then so do n.*

Proof. Assume m have the direction (123) and that n have something in the opposite direction.

First case is odd T and no triangles in n. In m everything have the same direction and so at least one edge with this direction outside a 2-cycle is needed for every variable in n, however in n there is only one edge outside of 2-cycles in each variable and so all of them have the right direction.

Second case is even T and no triangles in n. By assumption n contains some path directed in the wrong way, it is possible to chose one of these paths to be 321. Without loss of generality n do not contain the path 123 as this would allow the move $321, 123 \to (12), (23)$ that reduces the number of wrong directed paths. Now in this situation it is no loss of generality to assume that the only type of path directed the wrong way in n is 321, since other paths in the right direction is needed to cancel the edges with wrong direction. This give a surplus of edges 13 that can not be accounted for in m, in m there is at most one extra edge 13 for any other edge in the right direction.

Now there are two situations, that the triangles in n have the same direction as in m and that they have the opposite direction. The next two cases deal with triangles in the same direction.

Third case is odd T and triangles in n. Without loss of generality assume that n contains the path 21. Now n will not contain the paths $12, 123$ and 312. Furthermore it will not contain any variables with triangles but no path. In this situation n can not contain paths 32 and 13, this gives more extra 23 and 31 edges than can be in m.

Fourth case is even T and triangles in n. Again assume that n contain 321. This implies that n can not contain the paths $123, 12$ and 23. Now n can not contain any other type of wrong directed path. The extra edges 31 make n impossible.

When the triangles in n are directed in the opposite way as in m the situation is easier, it guarantees the existence of paths needed to cancel all triangles and and so it reduces to one of the previous four cases. □

Lemma 2.4. *Let m', n' be normal monomials so that $m' - n' \in I_{3,T}$. It is possible using degree 2 steps to go to normal m from m' and normal n from n' so that if x divides m and y divides n then the number of triangles in x and y differ by at most three.*

Proof. There is only something to prove if both monomials have all triangles and paths directed the same way.

Suppose that the minimal number of triangles in a variable in m' and n' are the same, then there is nothing to prove. If the minimal number differ by two or more then the number of edges in the triangles and paths can not add up to the same in both monomials and this is required since the set of 2-cycles have to be the same for both monomials. □

Lemma 2.5. *Let m and n be normal monomials with paths and triangles in the same direction and containing variables with the same number of triangles. If $m - n \in I_{3,T}$ then it is possible to go from m to n using degree two steps.*

Proof. Let x, y be variables with the same number of triangles so that x divide m and y divide n.

Suppose the variables x, y have the same path then it is possible to do a sequence of moves $(ij),(jk) \to (jk),(ij)$ to m, n that share a variable. This is because the set of 2-cycles are the same for both monomials.

Suppose x have the path 12 and y have the path 23. Now m can not contain any paths $23, 231$ and n can not contain any paths $12, 312$. Any extra edge 23 in m are locked up in paths or triangles also containing 12 and so there are more extra edges 12, however the same argument for 12 and n give that there are more extra edges 23. This is impossible.

Suppose x have the path 123 and y have the path 231. Now at most one of m and n can contain the path 213, assume without loss of generality that n contain no 213. Now any edge 12 in n are in triangles, while in m there are more edges 12 than edges 31. This is impossible.

The case that remains is when x have no path and y have the path 123. If m contains any variable with a path on two edges, a variable with a triangle and a path with one edge or a variable that is not x containing a triangle with no path. Then there is a move giving a path to x and not changing the triangles in x or the orientation of the other variables. The case that remains is that all variables in m except x contain no triangle and no path of length 2. If T is odd then the other variables contain no path and if T is even the other variables have paths with one edge. Note that x contain fewer extra edges than y and the other variables in m contain the lowest possible number of extra edges, this is impossible. □

Lemma 2.6. *If m and n are normal monomials so that $m - n \in I_{3,T}$ and m and n have all paths and cycles directed in the same way, then it is possible to use degree 2-steps to go from m to n.*

Proof. Note that by Lemma 2.5 the only case that remains is that the monomials have no variables with the same number of triangles and by Lemm 2.4 all the variables dividing one of the monomials have an even number of triangles and the variables dividing the other have an odd number of triangles.

One of the monomials have paths with one edge on each variable, assume that this monomial is m. The step $(ijk)ij, (i_1i_2)(i_3i_4)jk \to (i_1i_2)(i_3i_4), (ijk)ijk$ from m if possible creates variables with the same number of triangles in both monomials. This step is impossible if all the variables have the same number of triangles and fewer than two 2-cycles or if all paths are the same. If all paths are the same assume it to be 12 then the other monomial have to have paths on all variables containing the edge 12, however two paths of length two can at most give one extra edge 12 and so this is impossible. Now all variables have exactly one 2-cycle, zero 2-cycles is impossible since the number of 2-cycles in the other monomial is nonzero. Now again all paths still have to be the same otherwise the move $(ijk)(i_1i_2)ij, (i_3i_4)jk \to (i_1i_2)(i_3i_4)ijk, (ijk)$ is possible. $\square$

Lemma 2.7. *If m and n are normal monomials with paths and triangles in different directions and if $m - n \in I_{3,T}$ then it is possible to go from m to n using degree two steps.*

Proof. Again divide into the cases that T is evan and odd.

Assume that T is even and that m contains triangles. Let the triangles have the orientation of (123) and assume that there is a path 21. Now every triangle have to have the path 231 or there is a move destroying a triangle. This proves that all the paths with orientation opposite of the cycles are 21 and this proves that the only paths with the orientation of (123) are $23, 31$ and 231. Now there is more than one extra edge for each variable even after canceling the edges 21 from paths and 12 from triangles and this forces n to have triangles even after minimizing them as was done for m. In fact counting the extra edges give that the set of triangles and paths have to be equal for both monomials and so it is possible to go between them with the steps $(ij), (jk) \to (jk), (ij)$. When m do not contain triangles then n will not contain triangles by the previous argument and then it is possible to go between them with the moves $(ij), (jk) \to (jk), (ij)$ and $ij, ji(jk) \to jk, kj(ij)$.

Assume that T is odd and that m contains triangles. Again the triangles have the orientation (123) and m contains the path 321. All paths on variables with triangles have to be 31 and so all paths of opposite direction of the triangles are 321. The paths 123 may not occur and counting the extra edges as in the even case give that the sets of paths and triangles have to be the same for both monomials and it is possible to go between them. Now assume that m have no triangles then neither do n and it is clear that it is enough to use the moves $(ij), (jk) \to (jk), (ij)$ and $ijk, kji \to (ij), (jk)$ $\square$

Theorem 2.8. *Three-state toric homogenous Markov chain model has Markov degree two.*

Proof. Let m', n' be two monomials and m, n be the corresponding normal monomials obtained from Lemma 2.2, then $m' - n' \in I_{3,T}$ if and only if $m - n \in I_{3,T}$. Either both monomials have everything directed in the same way or both have paths and or triangles oriented differently according to Lemma 2.3. Now it is possible to go between the monomials by Lemma 2.6 or Lemma 2.7. $\square$

References

[1] A. Engström. Cut ideals of K4-minor free graphs are generated by quadrics. *Michigan Math. J.* **60** (2011), no. 3, 705–714.

[2] A. Engström, Thomas Kahle and Seth Sullivant. Multigraded Commutative Algebra of Graph Decompositions, `arXiv:1102.2601`, preprint 2011, 38 pp.

[3] A. Engström and P. Norén. Ideals of Graph Homomorphisms, *Ann. Comb.*, to appear, 34 pp.

[4] H. Hara and A. Takemura. A Markov basis for two-state toric homogeneous Markov chain model without initial paramaters. *J. Japan Statist. Soc.* **41** (2011) , 33–49.

[5] D. Haws, A. Martin Del Campo, R. Yoshida. Degree bounds for a minimal Markov basis for the three-state toric homogeneous Markov chain model. Proceedings of the Second CREST–SBM International Conference “Harmony of Gröbner Bases and the Modern Industrial Society.” 2012. 63–98.

[6] D. Haws, A. Martn del Campo, A. Takemura and R. Yoshida. Markov degree of the three-state toric homogeneous Markov chain model. `arXiv:1204.3070`, preprint 2012, 20 pp.

[7] A. Takemura and H. Hara, H. Markov chain Monte Carlo test of toric homogeneous Markov chains. *Stat. Methodol.* **9** (2012), no. 3, 392–406.
[8] R. Yoshida. Plenary talk at the Algebraic Statistics 2012 conference. June 2012 at Penn State.

Department of Mathematics and Systems Analysis, Aalto University
E-mail address: `patrik.noren@aalto.fi`