Séminaire Lotharingien de Combinatoire **93B** (2025)
Article #160, 9 pp.

Proceedings of the 37th Conference on Formal Power Series and Algebraic Combinatorics (Sapporo)

Computing all Markov bases

Oliver Clarke*[1] and Alexander Milner†[2]

[1] *Department of Mathematical Sciences & Department of Computer Science, Durham University, Stockton Rd, Durham DH1 3LE, United Kingdom*
[2] *The University of Edinburgh, James Clerk Maxwell Building, Peter Guthrie Tait Rd, Edinburgh EH9 3FD, United Kingdom*

Abstract. We introduce the package *AllMarkovBases* for *Macaulay2*, which computes all minimal Markov bases of a given toric ideal. The package builds on functionality of *4ti2* by producing the *fiber graph* of the toric ideal. The package uses this graph to compute properties of the toric ideal such as its indispensable set of binomials as well as its universal Markov basis.

Keywords: Markov bases, Toric ideals, Macaulay2

1 Introduction

A central theme of Algebraic Statistics is the study of Markov bases [19, 14, 1]. From an algebraic perspective, a Markov basis is simply a generating set of a toric ideal [18]. The celebrated result of Diaconis and Sturmfels [9], often called *the fundamental theorem of Markov bases*, provides an interpretation for these generating sets as moves in a Markov chain, which allows for sampling from conditional distributions. There are many distinct algorithms for producing one Markov basis of a toric ideal [13, 2, 12]. We refer the reader to the thesis of Malkin [16] for a comprehensive overview of these methods, including the state of the art *project-and-lift* algorithm. An implementation of this algorithm is included in *4ti2* [11], which is available in many computer-algebra systems including *Macaulay2* [10] via the package `FourTiTwo`.

There are a number of important properties of Markov bases, such as the *strongly robust property* [20, 15] and *distance reduction property* [21, 7], which can be read from the configuration matrix of the toric ideal. Thus, for testing ideas about Markov bases, it may be useful enumerate all minimal Markov bases or (uniformly) sample from them. So, in this extended abstract, we present the package `AllMarkovBases` [8] for *Macaulay2*, which is available at `https://github.com/ollieclarke8787/WorkshopDurham2024/tree/main/ToricIdeals`. Our package builds on `FourTiTwo` with functionality that allows the user to compute: all minimal Markov bases, the number of minimal Markov bases,

*oliver.clarke@durham.ac.uk
†a.j.c.milner@sms.ed.ac.uk

the universal Markov basis, and indispensable elements of a toric ideal. These families of elements have alternative combinatorial descriptions and prominently feature in the study of Markov bases. See, for example, [4, 15, 7, 6]. Our package also allows the user to sample uniformly from the set of minimal Markov bases, which may be particularly useful if the number of minimal Markov bases is very large.

Setup. Throughout the abstract, we fix the following setup. Let k be a field and $A \in \mathbb{Z}^{d \times n}$ a matrix. We assume that $\ker(A) \cap \mathbb{N}^n = \{0\}$. The toric ideal $I_A \subseteq k[x_1, \dots, x_n]$ is the ideal generated by the binomials $x^u - x^v$ where $u, v \in \mathbb{N}^n$ such that $u - v \in \ker(A)$. A *(minimal) Markov basis* of A is a (inclusion-minimal) set of generators M of I_A and we identify M with its set of binomial exponents $\{u - v : x^u - x^v \in M\}$. We often write the elements of M as the row-vectors of a matrix. The *indispensable set* $S(A)$ is the intersection of all minimal Markov bases of A. And the *universal Markov basis* $U(A)$ is the union of all minimal Markov bases of A.

Example 1.1. The matrix $A = \begin{pmatrix} 7 & 8 & 9 & 10 \end{pmatrix}$ defines a 1-dimensional toric ideal (monomial curve). The ideal has four minimal Markov bases, which are computed as follows.

```
i1 : needsPackage "AllMarkovBases";
i2 : A = matrix "7,8,9,10";

i3 : countMarkov A
o3 = 4

i4 : markovBases A
o4 = {| -1 2  -1 0  |, | -1 2  -1 0  |, | -1 2  -1 0  |, | -1 2  -1 0  |}
      | -1 1  1  -1 |  | -1 1  1  -1 |  | -1 1  1  -1 |  | -1 1  1  -1 |
      | 0  -1 2  -1 |  | 0  -1 2  -1 |  | 0  -1 2  -1 |  | 0  -1 2  -1 |
      | 4  0  -2 -1 |  | 4  0  -2 -1 |  | 4  -1 0  -2 |  | 4  -1 0  -2 |
      | 3  1  -1 -2 |  | 3  1  -1 -2 |  | 3  1  -1 -2 |  | 3  1  -1 -2 |
      | 3  0  1  -3 |  | 2  2  0  -3 |  | 3  0  1  -3 |  | 2  2  0  -3 |
```

So, for instance, the toric ideal $I_A \subseteq k[a, b, c, d]$ is minimally generated by the six polynomials

$$ac - b^2,\ ad - bc,\ bd - c^2,\ a^4 - c^2 d,\ a^3 b - cd^2,\ a^3 c - d^3.$$

Example 1.2. The monomial curve given by the matrix $A' = \begin{pmatrix} 51 & 52 & 53 & 54 & 55 & 56 \end{pmatrix}$ has 24300 minimal Markov bases. It may not be feasible to work with all of them so we can produce random samples as follows.

```
i5 : A' = matrix "51,52,53,54,55,56";

i6 : countMarkov A'
o6 = 24300

i7 : randomMarkov A'
o7 = | 8  4  0  0  0  -11 |
     | -1 2  -1 0  0  0   |
```

```
     | -1 1  1  -1 0  0   |
     | -1 0  2  0  -1 0   |
     | -1 1  0  1  -1 0   |
     | 0  -1 1  1  -1 0   |
     | -1 1  0  0  1  -1  |
     | 0  0  -1 2  -1 0   |
     | 0  -1 0  2  0  -1  |
     | 0  0  -1 1  1  -1  |
     | 0  0  0  -1 2  -1  |
     | 12 0  -1 0  -1 -9  |
     | 11 1  0  -1 -1 -9  |
     | 11 0  1  -1 0  -10 |
     | 9  3  0  0  -1 -10 |
```

The *indispensable set* $S(A')$ and the *universal Markov basis* $U(A')$ are computed as follows. We have abridged the output of `toricUniversalMarkov A'` to save space. We note that the indispensable set and universal Markov bases are read from the fiber graph, so the methods do not require the computation of all 24300 minimal Markov bases.

```
i8 : toricIndispensableSet A'
o8 = | -1 2 -1 0  0 0  |
     | -1 1 1  -1 0 0  |
     | 0  0 -1 1  1 -1 |
     | 0  0 0  -1 2 -1 |

i9 : toricUniversalMarkov A'
o9 = | 10 0  2  0  0  -11 |
     | 8  4  0  0  0  -11 |
      ...
     | 9  3  0  0  -1 -10 |
     | 11 0  0  1  -1 -10 |
               33        6
o9 : Matrix ZZ   <-- ZZ
```

Example 1.3. In Algebraic Statistics, Segre embeddings play an important role as they correspond to independence models. Consider the Segre embedding $\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2 \to \mathbb{P}^{26}$ defined by $([x_1 : x_2 : x_3], [y_1 : y_2 : y_3], [z_1 : z_2 : z_3]) \mapsto [x_i y_j z_k]_{ijk}$. The image of this embedding is the toric variety whose toric ideal is given by the matrix

$$A'' = \begin{bmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0 \\ 1&1&1&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0 \\ 0&0&0&1&1&1&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0&1&1&1&0&0&0 \\ 1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0 \\ 0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0&0&1&0 \end{bmatrix}.$$

We compute the number of Markov bases, the indispensable set, and universal Markov basis as follows. The semicolon suppresses the output, so the following shows that there are 81 indispensable binomials and the universal Markov bases has size 243.

```
i10 : A'' = matrix {
    {1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1},
    {1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0},
    {0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0},
    {1,1,1,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0},
    {0,0,0,1,1,1,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,1,1,1,0,0,0},
    {1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0},
    {0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0,0,1,0}};

i11 : countMarkov A''
o11 = 3245185536584267267831560205762 56

i12 : toricIndispensableSet A'';
                 81         27
o12 : Matrix ZZ    <-- ZZ

i13 : toricUniversalMarkov A'';
                 243         27
o13 : Matrix ZZ     <-- ZZ
```

We note that this example is particularly amenable to our method as the fiber graphs are all very small.

2 Implementation details

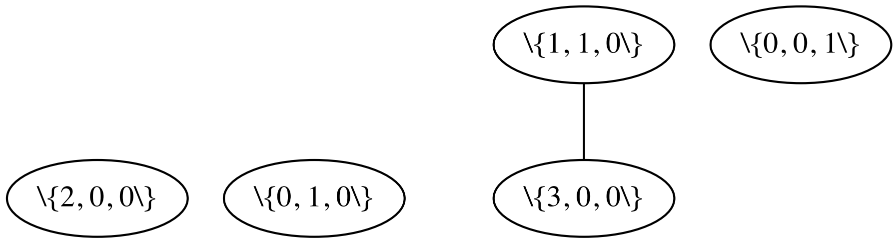
We translate the main results of Charalambous, Katsabekis, and Thoma [3, Section 2] into a procedure (Algorithm 1) that produces all minimal Markov bases. In this section, we explain the the procedure through a simple running example $A = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \in \mathbb{Z}^{1\times 3}$. Let us also fix a distinguished Markov basis

$$M = \begin{bmatrix} 2 & -1 & 0 \\ 3 & 0 & -1 \end{bmatrix}.$$

So the toric ideal $I_A \subseteq k[x, y, z]$ is minimally generated by the binomials $x^2 - y$ and $x^3 - z$. The *affine semigroup* of A is $\mathbb{N}A := \{Ax \in \mathbb{Z}^d : x \in \mathbb{N}^n\}$. For each $t \in \mathbb{N}A$, the *t-fiber* of A is the subset $\mathcal{F}_t := \{u \in \mathbb{N}^n : Au = t\}$. For each binomial $x^u - x^v \in I_A$, we have that $u, v \in \mathcal{F}_t$ belong to the same fiber of A, and we define its *A-degree* $\deg_A(x^u - x^v) = t$. In our running example we have

$$\deg_A(x^2 - y) = 2 \quad \text{and} \quad \deg_A(x^3 - z) = 3.$$

The following result follows directly from [3, Theorems 2.6 and 2.7].



(a) A-degree 2 (b) A-degree 3

Figure 1: Fiber graphs of A for the generating fibers

Theorem 2.1. *If M and M' are minimal Markov bases of A, then $\deg_A(M)$ and $\deg_A(M')$ are equal as multisets.*

We say that $t \in \mathbb{N}A$ is a *generating fiber* for A if there exists an element $x^u - x^v$ of a minimal Markov basis with A-degree $\deg_A(x^u - x^v) = t$. So, by Theorem 2.1, the set of all generating fibers is determined by a single Markov basis.

Definition 2.2. Let $t \in \mathbb{N}^d$. The *fiber graph* G_t has vertex set $\mathcal{F}_t$ and edge set

$$E(G_t) = \{uv : \text{there exists } i \in \{1, \dots, d\} \text{ such that } u_i > 0 \text{ and } v_i > 0\}.$$

To determine the minimal Markov bases, it suffices to consider only the fiber graphs of the generating fibers. With our package, these fiber graphs can be listed with the function `fiberGraph`. For our running example $A = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$ they are computed as follows.

```
i1 : needsPackage "AllMarkovBases";
i2 : A = matrix "1,2,3"
i3 : fiberGraph A
o3 = {Graph{{0, 1, 0} => {}}, Graph{{0, 0, 1} => {}          }}
            {2, 0, 0} => {}         {1, 1, 0} => {{3, 0, 0}}
                                    {3, 0, 0} => {{1, 1, 0}}
```

A visualisation of these fibers, produced by the package `Graphs`, is shown in Figure 1. Each element $z \in \ker(A)$ is thought of as a *move* that connects pairs of points $u, v \in \mathbb{N}^n$ whenever $u = v + z$. In our example, the move $(2, -1, 0)$ connects $(2,0,0)$ and $(0,1,0)$ and connects $(1,1,0)$ and $(3,0,0)$. A subset $M \subseteq \ker(A)$ is said to *connect* the t-fiber, for some $t \in \mathbb{N}A$, if the graph $G_{M,t}$ on $\mathcal{F}_t$ with edges $\{uv : u - v \in M\}$ is connected. The *fundamental theorem of Markov bases* [9, Theorem 3.1] tells us that Markov bases are in one-to-one correspondence with subsets of $\ker(A)$ that connect each fiber of A. Similarly, by [3, Theorems 2.6 and 2.7], the minimal Markov bases are in one-to-one correspondence with sets of moves that *minimally connect* G_t for each generating fiber t of A.

Remark 2.3. In the literature, the graphs $G_{M,t}$ are often called fiber graphs [12]. To avoid confusion, for us the term fiber graph will always refer to the graph G_t for some t.

In our running example, observe that the 2-fiber in Figure 1a is connected only by the element $(2, -1, 0)$, up to sign. It follows that the corresponding binomial $x^2 - y$ is *indispensable*, i.e. it appears in every Markov basis, as there is no other way to connect the 2-fiber. On the other hand, the 3-fiber in Figure 1b is minimally connected with either the move $(1, 1, -1)$ or $(3, 0, -1)$. These form the two minimal Markov bases of A. In our package, this process is carried out by the function `markovBases`:

```
i4 : markovBases A
o4 = {| 2 -1 0  |, | 2 -1 0  |}
      | 3 0  -1 |   | 1 1  -1 |
```

We now state the algorithm implemented by `markovBases`.

Algorithm 1: All minimal Markov bases of a matrix

Input : $A \in \mathbb{Z}^{d \times n}$ a configuration matrix
Output: $\mathcal{M}$ the set of minimal Markov bases of A

1: Initialise $\mathcal{M} = \emptyset$;
2: **for** $t \in \mathbb{N}A$ *a generating fiber of* A **do**
3: Compute $C_t :=$ the set of connected components of the fiber graph G_t ;
4: **end**
5: $\mathcal{T} := \prod_t \{T : \text{spanning tree on vertex set } C_t\}$, where the product is taken over all generating fibers $t \in \mathbb{N}A$ of A
6: **for** $\{T_t : \textit{spanning tree on } C_t\} \in \mathcal{T}$ **do**
7: **for** *each collection* $\{f_{T,e} : T \in \mathcal{T},\, e \in E(T)\}$ *where* $f_{T,e}$ *is a choice function on* e **do**
8: $M := \{f_{T,e}(u) - f_{T,e}(v) : T \in \mathcal{T},\, uv \in E(T)\}$ Markov basis of A;
9: $\mathcal{M} \leftarrow \mathcal{M} \cup \{M\}$;
10: **end**
11: **end**

Let us give some notes on how the algorithm works. Recall that each minimal Markov basis corresponds to a collection of moves that minimally connects the fiber graph G_t for each generating fiber t of A. We enumerate these collections of moves in two steps. First, we find the connected components of each fiber graph and connect them with spanning trees. This corresponds to lines 2 to 5 and the loop spanning lines 6 to 11. Next, for each edge of such a spanning tree we choose a pair of points in the fiber that lie in the connected components in the edge. This gives rise to a move, which lies the minimal Markov basis. Let us expand on this explanation with the following example.

Example 2.4. Suppose $\mathcal{F}_t = \{v_1, \dots, v_6\}$ is a generating fiber and G_t has three connected components $C_t = \{v_1v_2,\, v_3v_4,\, v_5v_6\}$. If we fix a spanning tree with edges $e_1 = \{v_1v_2, v_3v_4\}$ and $e_2 = \{v_1v_2, v_5v_6\}$, then there are four choices for turning the edge e_1

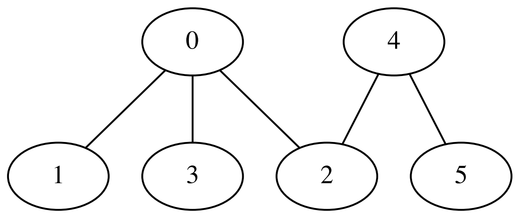


Figure 2: Spanning tree on $\{0,1,\ldots,5\}$ associated to the Prüfer sequence $\{0,0,2,4\}$

into a move:

$$v_1 - v_3,\ v_1 - v_4,\ v_2 - v_3,\ v_2 - v_4.$$

Similarly, there are four ways to turn e_2 into a move. Thus, this generating fiber contributes two elements of a minimal Markov basis in 16 ways.

The spanning trees are enumerated using the well-known bijection of Prüfer [17]. This bijection between labelled spanning trees on n vertices and *Prüfer sequences* (sequences in $\{0,\ldots,n-1\}$ of length $n-2$) is completely constructive. Our package implements the helpful function `pruferSequence` that takes a Prüfer sequence and returns the edge set of the corresponding spanning tree. For example, the spanning tree associated to the sequence $\{0,0,2,4\}$ is given computed as follows, see Figure 2.

```
i5 : pruferSequence {0,0,2,4}
o5 = {set {0, 1}, set {0, 3}, set {0, 2}, set {4, 2}, set {4, 5}}
```

It is straightforward to see from Algorithm 1 and Example 2.4 that the number of minimal Markov bases can be computed from the sizes of connected components of each fiber graph.

Theorem 2.5 ([3, Theorem 2.9])**.** *For each generating fiber $t \in \mathbb{N}A$, assume G_t has n_t connected components of size $m_{t,1}, m_{t,2}, \ldots, m_{t,n_t}$ for some positive integers $m_{t,i}$ with $i \in \{1,2,\ldots,n_t\}$. The number of minimal Markov bases of A is given by*

$$\prod_t m_{t,1} \cdot m_{t,2} \cdots m_{t,n_t} \cdot (m_{t,1} + m_{t,2} + \cdots + m_{t,n_t})^{n_t - 2}.$$

Our package implements this as the function `countMarkov`, which computes the number of minimal Markov bases without enumerating all of them.

2.1 Applications of fiber graphs

Once we have computed the fiber graphs of the generating fibers of a toric ideal, other properties become immediately available. Our package extracts these properties with variation on Algorithm 1, which allows us to sample from the set of Markov bases and compute the indispensable set and univeral Markov basis.

Random sampling. The function `randomMarkov` produces a random Markov basis. In lines 6 and 7 of Algorithm 1, we sample uniformly an element of $\mathcal{T}$ and a random collection of choice functions. This produces a single uniformly distributed minimal Markov basis of A.

Indispensable set and universal Markov basis. The universal Markov basis $U(A)$ of the matrix A is the union of its minimal Markov bases. Since every minimal Markov basis arises from Algorithm 1, we compute $U(A)$ by replacing the collection of spanning trees $\mathcal{T}$ with a single complete graph.

By [3, Corollary 2.10], the indispensable elements are detected directly from the fiber graphs. The function `toricIndispensableSet` determines the indispensable elements by finding each fiber graph G_t with exactly two vertices u, v such that $|u| = |v| = 1$, i.e. the two vertices of G_t are subsets of $\mathbb{N}^n$ of size one. This method is completely analogous to [5, Algorithm 1, Theorem 3.3].

References

[1] F. Almendra-Hernández, J. A. De Loera, and S. Petrović. "Markov bases: a 25 year update". *Journal of the American Statistical Association* **119**.546 (2024), pp. 1671–1686. DOI.

[2] A. M. Bigatti, R. La Scala, and L. Robbiano. "Computing toric ideals". *Journal of Symbolic Computation* **27**.4 (1999), pp. 351–365. DOI.

[3] H. Charalambous, A. Katsabekis, and A. Thoma. "Minimal systems of binomial generators and the indispensable complex of a toric ideal". *Proceedings of the American Mathematical Society* **135**.11 (2007), pp. 3443–3451. DOI.

[4] H. Charalambous, A. Thoma, and M. Vladoiu. "Markov complexity of monomial curves". *Journal of Algebra* **417** (2014), pp. 391–411. DOI.

[5] H. Charalambous, A. Thoma, and M. Vladoiu. "Binomial fibers and indispensable binomials". *Journal of Symbolic Computation* **74** (2016), pp. 578–591. DOI.

[6] H. Charalambous, A. Thoma, and M. Vladoiu. "Minimal generating sets of lattice ideals". *Collectanea Mathematica* **68**.3 (2017), pp. 377–400. DOI.

[7] O. Clarke and D. Kosta. "Distance Reducing Markov Bases". 2024. arXiv:2406.17730.

[8] O. Clarke and A. Milner. "Computing all minimal Markov bases in Macaulay2". 2025. arXiv:2502.19031.

[9] P. Diaconis and B. Sturmfels. "Algebraic algorithms for sampling from conditional distributions". *The Annals of Statistics* **26**.1 (1998), pp. 363–397. DOI.

[10] D. R. Grayson and M. E. Stillman. "Macaulay2, a software system for research in algebraic geometry". Available at `http://www2.macaulay2.com`.

[11] R. Hemmecke, R. Hemmecke, M. Köppe, P. Malkin, and M. Walter. "4ti2—A software package for algebraic, geometric and combinatorial problems on linear spaces". Link.

[12] R. Hemmecke and P. N. Malkin. "Computing generating sets of lattice ideals and Markov bases of lattices". *Journal of Symbolic Computation* **44**.10 (2009), pp. 1463–1476. DOI.

[13] S. Ho¸sten and B. Sturmfels. "GRIN: an implementation of Gröbner bases for integer programming". *Integer programming and combinatorial optimization (Copenhagen, 1995)*. Vol. 920. Lecture Notes in Comput. Sci. Springer, Berlin, 1995, pp. 267–276. DOI.

[14] D. Kosta. "Markov Bases of Toric Ideals: Connecting Commutative Algebra and Statistics". *London Mathematical Society Newsletter* **486** (2020), pp. 34–37.

[15] D. Kosta, A. Thoma, and M. Vladoiu. "On the strongly robust property of toric ideals". *Journal of Algebra* **616** (2023), pp. 1–25. DOI.

[16] P. Malkin. "Computing Markov bases, Gröbner bases, and extreme rays." PhD thesis. Catholic University of Louvain, Louvain-la-Neuve, Belgium, 2007.

[17] H. Prüfer. "Neuer Beweis eines Satzes über Permutationen". *Archiv der Mathematischen Physik* **27** (1918), pp. 742–744.

[18] B. Sturmfels. *Gröbner bases and convex polytopes*. Vol. 8. University Lecture Series. American Mathematical Society, Providence, RI, 1996, xii+162 pp. DOI.

[19] S. Sullivant. *Algebraic statistics*. Vol. 194. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2018, xiii+490 pp. DOI.

[20] S. Sullivant. "Strongly robust toric ideals in codimension 2". *Journal of Algebraic Statistics* **10**.1 (2019), pp. 128–136. DOI.

[21] A. Takemura and S. Aoki. "Distance-reducing Markov bases for sampling from a discrete sample space". *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability* **11**.5 (2005), pp. 793–813. DOI.