Algebra & Number Theory

Volume 5
2011
No. 5

mathematical sciences publishers

Algebra & Number Theory

www.jant.org

PRODUCTION

contact@msp.org

Silvio Levy, Scientific Editor

See inside back cover or www.jant.org for submission instructions.

The subscription price for 2011 is US $150/year for the electronic version, and $210/year (+$35 shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA.

Algebra & Number Theory (ISSN 1937-0652) at Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY
mathematical sciences publishers
http://msp.org/

A NON-PROFIT CORPORATION

Typeset in LaTeX

ALGEBRA AND NUMBER THEORY 5:5(2011)

Mutation classes of $\tilde{A}_n$-quivers and derived equivalence classification of cluster tilted algebras of type $\tilde{A}_n$

Janine Bastian

We give an explicit description of the mutation classes of quivers of type $\tilde{A}_n$. Furthermore, we provide a complete classification of cluster tilted algebras of type $\tilde{A}_n$ up to derived equivalence. We show that the bounded derived category of such an algebra depends on four combinatorial parameters of the corresponding quiver.

1. Introduction

A few years ago, Fomin and Zelevinsky [2002] introduced the concept of cluster algebras, which rapidly became a successful research area. Cluster algebras nowadays link various areas of mathematics, like combinatorics, Lie theory, algebraic geometry, representation theory, integrable systems, Teichmüller theory, Poisson geometry and also string theory in physics (via recent work on quivers with superpotentials; see [Derksen et al. 2008; Labardini-Fragoso 2009]).

In an attempt to categorify cluster algebras, which a priori are combinatorially defined, cluster categories have been introduced by Buan, Marsh, Reineke, Reiten and Todorov [Buan et al. 2006]. For a quiver Q without loops and oriented 2-cycles and the corresponding path algebra KQ (over an algebraically closed field K), the cluster category $\mathscr{C}_Q$ is the orbit category of the bounded derived category $D^b(KQ)$ by the functor $\tau^{-1}[1]$, where τ denotes the Auslander–Reiten translation and $[1]$ is the shift functor on the triangulated category $D^b(KQ)$.

Important objects in cluster categories are the cluster-tilting objects. The endomorphism algebras of such objects in the cluster category $\mathscr{C}_Q$ are called cluster tilted algebras of type Q [Buan et al. 2007]. Cluster tilted algebras have several interesting properties; for example, their representation theory can be completely understood in terms of the representation theory of the corresponding path algebra of a quiver (ibid.). These algebras have been studied by various authors; see for instance [Assem et al. 2008a, 2008b; Buan et al. 2008; Caldero et al. 2006].

MSC2000: primary 16G20; secondary 18E35, 18E30.

Keywords: cluster tilted algebra, quiver mutation, derived equivalence.

In recent years, a focal point in the representation theory of algebras has been the investigation of derived equivalences of algebras. Since a lot of properties and invariants of rings and algebras are preserved by derived equivalences, it is important for many purposes to classify classes of algebras up to derived equivalence, instead of Morita equivalence. For self-injective algebras, the representation type is preserved under derived equivalences [Krause 1997; Rickard 1989a]. It has also been proved in [Rickard 1991] that the class of symmetric algebras is closed under derived equivalences. Additionally, derived equivalent algebras have the same number of pairwise nonisomorphic simple modules and isomorphic centers.

In this work, we are concerned with the problem of derived equivalence classification of cluster tilted algebras of type $\tilde{A}_n$. Such a classification was done for cluster tilted algebras of type A_n by Buan and Vatne [2008]; see also [Murphy 2010] on the more general case of m-cluster tilted algebras of type A_n.

Since the quivers of cluster tilted algebras of type $\tilde{A}_n$ are exactly the quivers in the mutation classes of $\tilde{A}_n$, our first aim in this paper is to give a description of the mutation classes of $\tilde{A}_n$-quivers; these mutation classes are known to be finite (for example, see [Fomin et al. 2008]). The second purpose of this note is to describe, when two cluster tilted algebras of type Q have equivalent derived categories, where Q is a quiver whose underlying graph is $\tilde{A}_n$.

In Definition 3.3 we present a class $\mathcal{Q}_n$ of quivers with $n+1$ vertices that includes all nonoriented cycles of length $n+1$. To show that this class contains all quivers mutation-equivalent to some quiver of type $\tilde{A}_n$ we first prove that this class is closed under quiver mutation. Furthermore, we define parameters $r_1,\ r_2,\ s_1$ and s_2 for any quiver $Q \in \mathcal{Q}_n$ in Definition 3.7 and prove that every quiver in $\mathcal{Q}_n$ with parameters $r_1,\ r_2,\ s_1$ and s_2 can be mutated to a normal form, see Figure 1, without changing the parameters.

With the help of the result above we can show that every quiver $Q \in \mathcal{Q}_n$ with parameters $r_1,\ r_2,\ s_1$ and s_2 is mutation-equivalent to some nonoriented cycle with $r := r_1 + 2r_2$ arrows in one direction and $s := s_1 + 2s_2$ arrows in the other. Hence, if two quivers Q_1 and Q_2 of $\mathcal{Q}_n$ have the parameters $r_1,\ r_2,\ s_1,\ s_2$, respectively $\tilde{r}_1,\ \tilde{r}_2,\ \tilde{s}_1,\ \tilde{s}_2$ and $r_1 + 2r_2 = \tilde{r}_1 + 2\tilde{r}_2,\ s_1 + 2s_2 = \tilde{s}_1 + 2\tilde{s}_2$ (or vice versa), then Q_1 is mutation-equivalent to Q_2.

The converse of this result — an explicit description of the mutation classes of quivers of type $\tilde{A}_n$ — can be shown with the help of [Fomin et al. 2008, Lemma 6.8].

The main result of the derived equivalence classification of cluster tilted algebras of type $\tilde{A}_n$ is the following theorem:

Theorem 1.1. *Two cluster tilted algebras of type $\tilde{A}_n$ are derived equivalent if and only if their quivers have the same parameters $r_1,\ r_2,\ s_1$ and s_2 (up to changing the roles of r_i and s_i, $i \in \{1, 2\}$).*

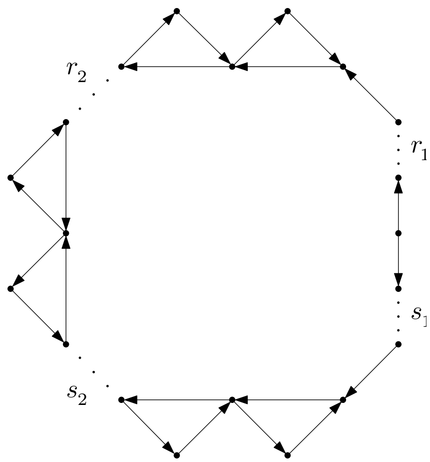


Figure 1. Normal form for quivers in $\mathcal{Q}_n$.

We prove that every cluster tilted algebra of type $\tilde{A}_n$ with parameters r_1, r_2, s_1 and s_2 is derived equivalent to a cluster tilted algebra corresponding to a quiver in normal form. Furthermore, we compute the parameters r_1, r_2, s_1 and s_2 as combinatorial derived invariants for a quiver $Q \in \mathcal{Q}_n$ with the help of an algorithm defined by Avella-Alaminos and Geiß [2008].

The paper is organized as follows. In Section 2 we collect some basic notions about quiver mutations. In Section 3 we present the set $\mathcal{Q}_n$ of quivers that can be obtained by iterated mutation from quivers whose underlying graph is of type $\tilde{A}_n$. Moreover, we describe, when two quivers of $\mathcal{Q}_n$ are in the same mutation class. In the fourth section we describe the cluster tilted algebras of type $\tilde{A}_n$ and their relations (as shown in [Assem et al. 2010]). In Section 5 we first briefly review the fundamental results on derived equivalences. Afterwards, we prove our main result, that is, we show, when two cluster tilted algebras of type $\tilde{A}_n$ are derived equivalent.

2. Quiver mutations

A *quiver* is a finite directed graph Q, consisting of a finite set of vertices Q_0 and a finite set of arrows Q_1 between them.

Let Q be a quiver and K be an algebraically closed field. We can form the *path algebra* KQ, where the basis of KQ is given by all paths in Q, including trivial paths e_i of length zero at each vertex i of Q. Multiplication in KQ is defined by concatenation of paths. Our convention is to read paths from right to left. For any path α in Q let $s(\alpha)$ denote its start vertex and $t(\alpha)$ its end vertex. Then the product of two paths α and β is defined to be the concatenated path $\alpha\beta$ if $s(\alpha) = t(\beta)$. The unit element of KQ is the sum of all trivial paths: $1_{KQ} = \sum_{i \in Q_0} e_i$.

We now recall the definition of quiver mutations.

Definition 2.1 [Fomin and Zelevinsky 2002]. Let Q be a quiver without loops and oriented 2-cycles. The *mutation* of Q at a vertex k to a new quiver Q^* can be described as follows:

(1) Add a new vertex k^*.

(2) If there are $r > 0$ arrows $i \to k$, $s > 0$ arrows $k \to j$ and $t \in \mathbb{Z}$ arrows $j \to i$ in Q, there are $t - rs$ arrows $j \to i$ in Q^*. (Here, a negative number of arrows means arrows in the opposite direction.)

(3) For any vertex i replace all arrows from i to k with arrows from k^* to i, and replace all arrows from k to i with arrows from i to k^*.

(4) Remove the vertex k.

Note that mutation at sinks or sources only means changing the direction of all incoming or outgoing arrows. Two quivers are called *mutation-equivalent* (or *sink/source equivalent*) if one can be obtained from the other by a finite sequence of mutations (at sinks or sources). The *mutation class* of a quiver Q is the class of all quivers mutation-equivalent to Q.

3. Mutation classes of $\tilde{A}_n$-quivers

Remark 3.1. Quivers of type $\tilde{A}_n$ are just cycles with $n+1$ vertices. If the cycle is oriented, we get the mutation class of D_{n+1} (see [Derksen and Owen 2008; Fomin et al. 2008, 2003] or Type IV in type D in [Vatne 2010]). If the cycle is nonoriented, we get what we call the mutation classes of $\tilde{A}_n$.

First, we have to fix one drawing (plane embedding) of this nonoriented cycle. Thus, we can speak of clockwise and anticlockwise oriented arrows. But we have to consider that this notation is only unique up to reflection of the cycle, i.e., up to changing the roles of clockwise and anticlockwise oriented arrows.

Lemma 3.2 [Fomin et al. 2008, Lemma 6.8]. *Let C_1 and C_2 be two nonoriented cycles, so that in C_1 there are s arrows oriented clockwise and r arrows oriented anticlockwise. Similarly, in C_2 there are $\tilde{s}$ arrows oriented clockwise and $\tilde{r}$ arrows oriented anticlockwise. Then C_1 and C_2 are mutation-equivalent if and only if the unordered pairs $\{r, s\}$ and $\{\tilde{r}, \tilde{s}\}$ coincide.*

Thus, two nonoriented cycles of length $n+1$ are mutation-equivalent if and only if they have the same parameters r and s (up to changing the roles of r and s).

Next we provide an explicit description of the mutation classes of $\tilde{A}_n$-quivers. For this we need a description of the mutation class of quivers of type A_k. We use one given in [Buan and Vatne 2008]:

- Each quiver has k vertices.
- All nontrivial cycles are oriented and of length 3.
- A vertex has at most four incident arrows.
- If a vertex has four incident arrows, then two of them belong to one oriented 3-cycle, and the other two belong to another oriented 3-cycle.
- If a vertex has three incident arrows, then two of them belong to an oriented 3-cycle, and the third arrow does not belong to any oriented 3-cycle.

(By a *cycle* in the second condition we mean a cycle in the underlying graph not passing through the same edge twice. In particular, this condition excludes multiple arrows.)

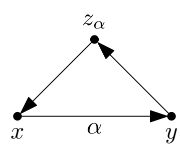
For another description of mutation classes of type A quivers, see [Seven 2007].

Now we can formulate the description of the mutation classes of $\tilde{A}_n$-quivers, similar to the description for Type IV in type D in [Vatne 2010].

Definition 3.3. Let $\mathcal{Q}_n$ be the class of connected quivers with $n+1$ vertices that satisfy the following conditions (see Figure 2 for an illustration):

(i) There exists precisely one full subquiver that is a nonoriented cycle of length ≥ 2. Thus, if the length is two, it is a double arrow.

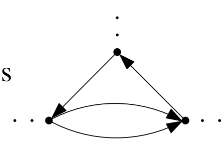
(ii) For each arrow $x \xrightarrow{\alpha} y$ in this nonoriented cycle, there may (or may not) be a vertex z_α not on the nonoriented cycle and such that there is an oriented 3-cycle of the form



Apart from the arrows of these oriented 3-cycles there are no other arrows incident to vertices on the nonoriented cycle.

(iii) If we remove all vertices in the nonoriented cycle and their incident arrows, the result is a disjoint union of quivers $Q_1, Q_2, \ldots$, one for each z_α (which we call Q_α). These are quivers of type A_{k_α} for $k_\alpha \geq 1$, and the vertices z_α have at most two incident arrows in these quivers. Furthermore, if a vertex z_α has two incident arrows in such a quiver, then z_α is a vertex in an oriented 3-cycle in Q_α.

Our convention is to choose only one of the double arrows to be part of the oriented 3-cycle in the case shown here:



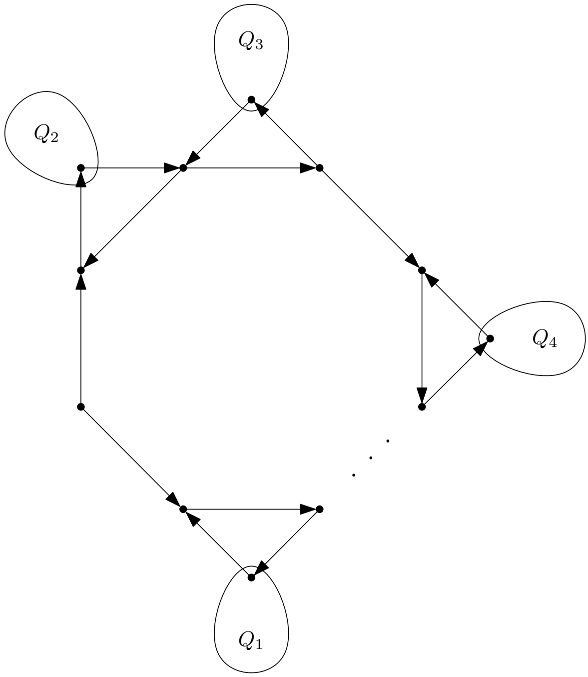
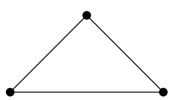


Figure 2. Quiver in $\mathcal{Q}_n$.

Notation 3.4. Whenever we draw an edge 

the direction of the arrow between j and k is not important for this situation; and whenever we draw a cycle



it is an oriented 3-cycle.

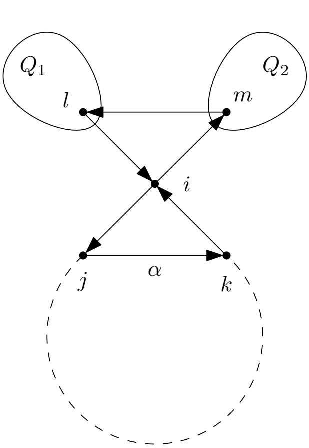
Lemma 3.5. *$\mathcal{Q}_n$ is closed under quiver mutation.*

Proof. Let Q be a quiver in $\mathcal{Q}_n$ and let i be some vertex of Q. The subquivers Q_1 and Q_2 highlighted in the pictures are quivers of type A.

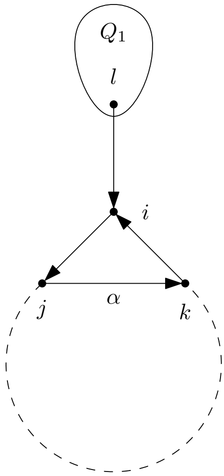
If i is a vertex in one of the quivers Q_α of type A, but not one of the vertices z_α connecting this quiver of type A to the rest of the quiver Q, then the mutation at i leads to a quiver $Q^* \in \mathcal{Q}_n$ since type A is closed under quiver mutation.

It therefore suffices to check what happens when we mutate at the other vertices, and we will consider four cases:

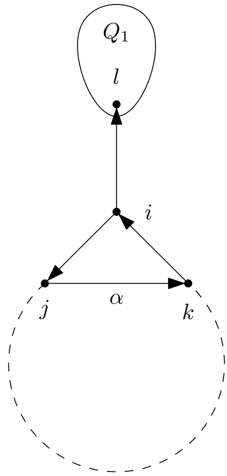
(1) Let i be one of the vertices z_α, hence not on the nonoriented cycle. For the situation where the quiver Q_α of type A attached to z_α consists only of one vertex, we can look at the first mutated quiver in case (2) below since quiver mutation is an involution. Thus, we have three cases:



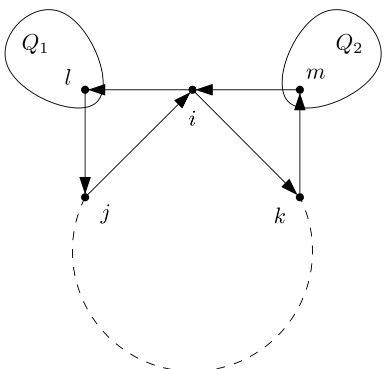
or



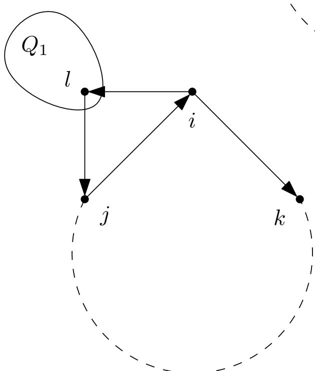
or



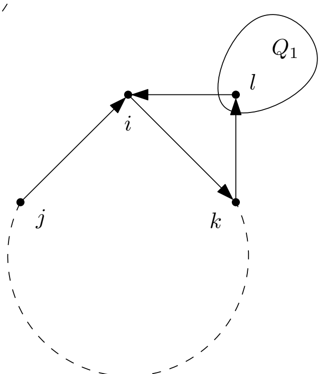
Then the mutation at i leads to the following three quivers, which have a nonoriented cycle one arrow longer than for Q, and this is indeed a nonoriented cycle since the arrows $j \to i \to k$ have the same orientation as α had before.



or

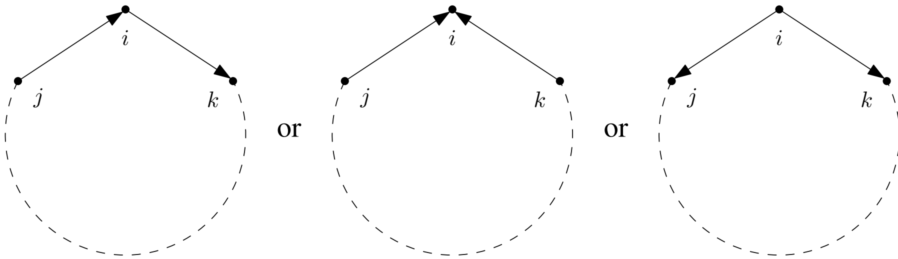


or

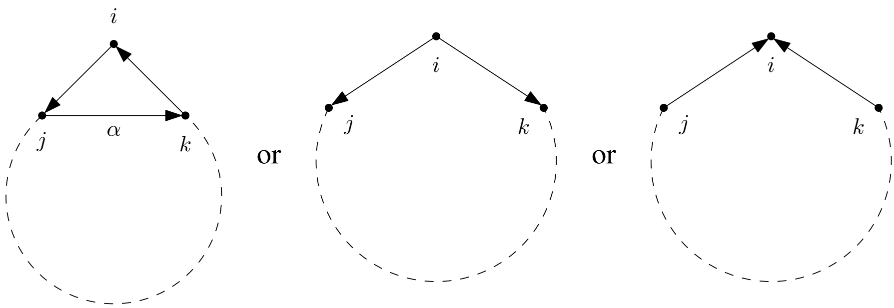


The vertices l and m have at most two incident arrows in the quivers Q_1 and Q_2 since they had at most four resp. three incident arrows in Q (see the description of quivers of type A). Furthermore, if l or m has two incident arrows in the quiver Q_1 or Q_2, then these two arrows form an oriented 3-cycle as in Q. Thus, the mutated quiver Q^* is also in $\mathcal{Q}_n$.

(2) Let i be a vertex on the nonoriented cycle, and not part of any oriented 3-cycle. Three cases can occur:



and mutation at i leads to



If i is a sink or a source in Q, the nonoriented cycle in Q^* is of the same length as before and Q^* is in $\mathcal{Q}_n$. If there is a path $j \to i \to k$ in Q, mutation at i leads to a quiver Q^*, which has a nonoriented cycle one arrow shorter than in Q.

Note that in this case the nonoriented cycle in Q consists of at least three arrows and thus, the nonoriented cycle in Q^* has at least two arrows. Thus, the mutated quiver Q^* is also in $\mathcal{Q}_n$.

(3) Let i be a vertex on the nonoriented cycle that is part of exactly one oriented 3-cycle. Then four cases can occur, but two of them have been dealt with by the second and third mutated quiver in case (1) since quiver mutation is an involution. Thus, we only have to consider the two situations shown in Figure 3 and their special cases where the nonoriented cycle is a double arrow. (The two-headed arrows indicate mutation at i.)

After mutating at vertex i, the nonoriented cycle has the same length as before. Moreover, l has the same number of incident arrows as before. Thus, Q^* is in $\mathcal{Q}_n$.

(4) Let i be a vertex on the nonoriented cycle that is part of two oriented 3-cycles. Then three cases can occur, but one of them has been dealt with by the first mutated quiver in case (1). Thus, we have to consider only the situations in Figure 4 and their special cases where the nonoriented cycle is a double arrow.

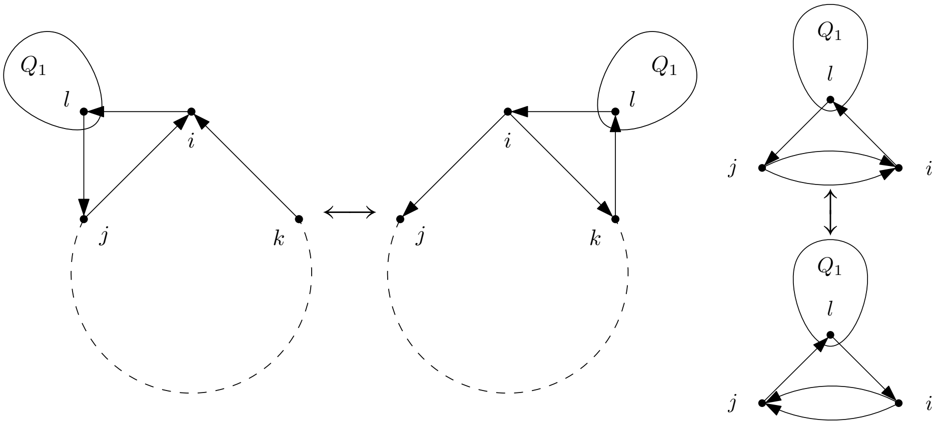


Figure 3. Possibilities in case (3).

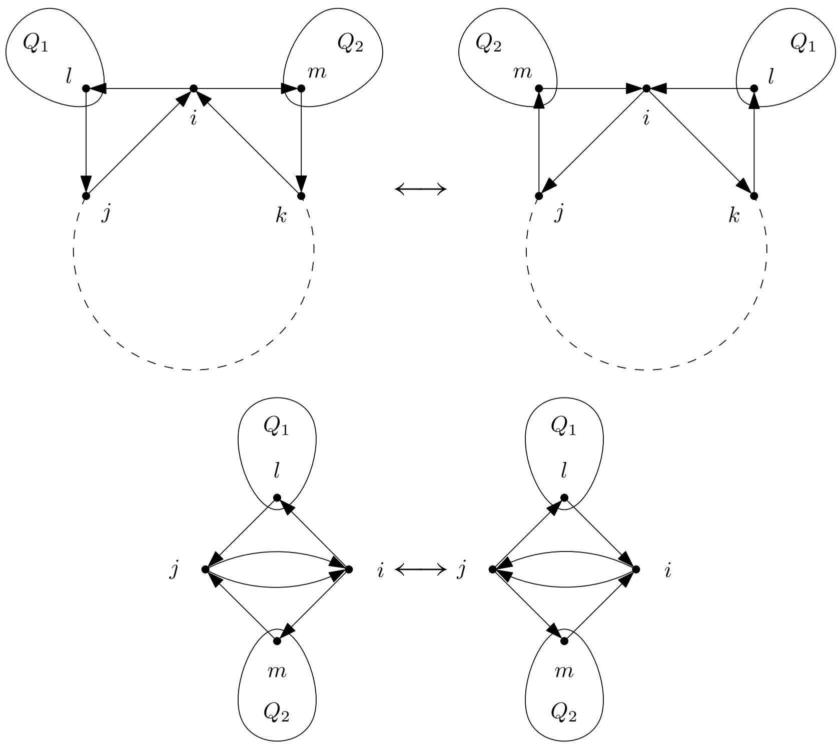


Figure 4. Possibilities in case (4).

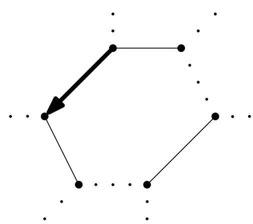
The nonoriented cycle has the same length as before. Moreover, l and m have the same number of incident arrows as before. Thus, again, the mutated quiver Q^* belongs to $\mathfrak{Q}_n$. $\square$

Remark 3.6. It is easy to see that all orientations of a circular quiver of type $\tilde{A}_n$ are in $\mathcal{Q}_n$ (except the oriented case; but this leads to the mutation class of D_{n+1}). Since $\mathcal{Q}_n$ is closed under quiver mutation every quiver mutation-equivalent to some quiver of type $\tilde{A}_n$ is in $\mathcal{Q}_n$, too.

Now we fix one drawing of a quiver $Q \in \mathcal{Q}_n$, without arrow crossing. Thus, we can again speak of clockwise and anticlockwise oriented arrows of the nonoriented cycle. But we have to consider that this notation is only unique up to reflection of the nonoriented cycle, that is, up to changing the roles of clockwise and anticlockwise oriented arrows. We define four parameters r_1, r_2, s_1 and s_2 for a quiver $Q \in \mathcal{Q}_n$ as follows:

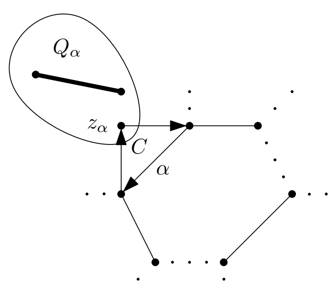
Definition 3.7. • Let r_1 be the number of arrows that are not part of any oriented 3-cycle and that fulfill one of two conditions:

(1) The arrow is part of the nonoriented cycle and is oriented anticlockwise:



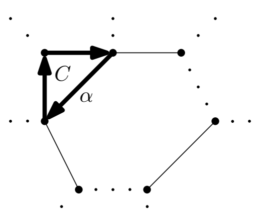
(2) The arrow is not part of the nonoriented cycle, but is attached to an oriented 3-cycle C sharing with the nonoriented cycle one arrow α that is oriented anticlockwise (see figure on the right).

In this sense, "attached" means that the arrow is part of the quiver Q_α of type A that shares the vertex z_α with the cycle C (see Definition 3.3).

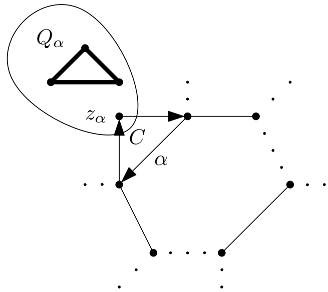


• Let r_2 be the number of oriented 3-cycles that fulfill one of two conditions:

(1) The cycle shares with the nonoriented cycle one arrow α that is oriented anticlockwise:



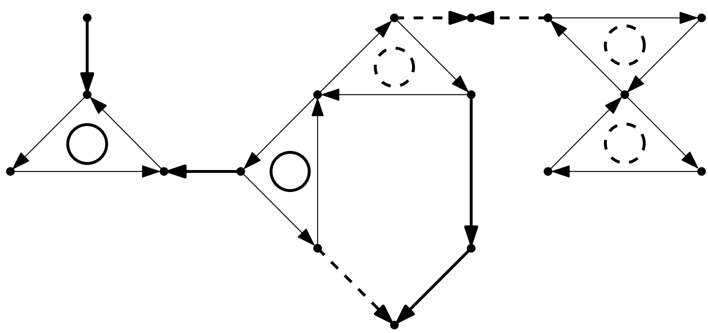
(2) The cycle is attached to an oriented 3-cycle C sharing one arrow α with the nonoriented cycle and α is oriented anticlockwise:



Here, "attached" is in the same sense as above.

- Similarly we define the parameters s_1 and s_2 with "clockwise" instead of "anticlockwise".

Example 3.8. We denote the arrows that count for the parameter r_1 by and the arrows that count for s_1 by . Furthermore, the oriented 3-cycles of r_2 are denoted by and the oriented 3-cycles of s_2 are denoted by . Let $Q \in \mathcal{Q}_{16}$ be a quiver of the form



Then $r_1 = 3,\ r_2 = 3,\ s_1 = 4$ and $s_2 = 2$.

Lemma 3.9. *If Q_1 and Q_2 are quivers in $\mathcal{Q}_n$, and Q_1 and Q_2 have the same parameters $r_1,\ r_2,\ s_1$ and s_2 (up to interchanging r_1 with s_1 and r_2 with s_2), then Q_2 can be obtained from Q_1 by iterated mutation, where all the intermediate quivers have the same parameters as well.*

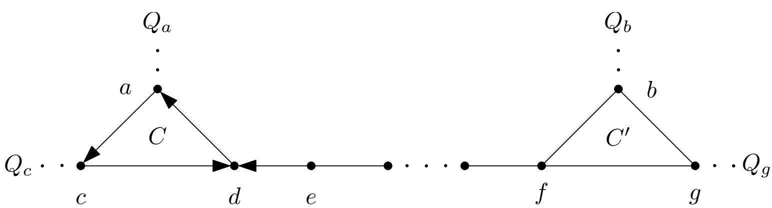
Proof. It is enough to show that all quivers in $\mathcal{Q}_n$ with parameters $r_1,\ r_2,\ s_1\ s_2$ can be mutated to a quiver in *normal form* (see Figure 1) without changing the parameters $r_1,\ r_2,\ s_1\ s_2$. In such a quiver, r_1 is the number of anticlockwise arrows in the nonoriented cycle that do not share any arrow with an oriented 3-cycle and

s_1 is the number of clockwise arrows in the nonoriented cycle that do not share any arrow with an oriented 3-cycle. Furthermore, r_2 is the number of oriented 3-cycles sharing one arrow α with the nonoriented cycle and α is oriented anticlockwise and s_2 is the number of oriented 3-cycles sharing one arrow β with the nonoriented cycle and β is oriented clockwise (see Definition 3.7).

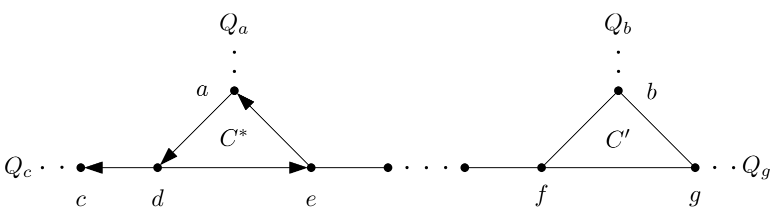
We divide this process into five steps.

Step 1: Let Q be a quiver in $\mathcal{Q}_n$. We move all oriented 3-cycles of Q sharing no arrow with the nonoriented cycle towards the oriented 3-cycle that is attached to them and that shares one arrow with the nonoriented cycle.

Method: Let C and C' be a pair of neighboring oriented 3-cycles in Q (i.e., no arrow in the path between them is part of an oriented 3-cycle) such that the length of the path between them is at least one. We want to move C and C' closer together by mutation.



In the picture the Q_i are subquivers of Q. Mutating at d will produce a quiver Q^* looking like this:

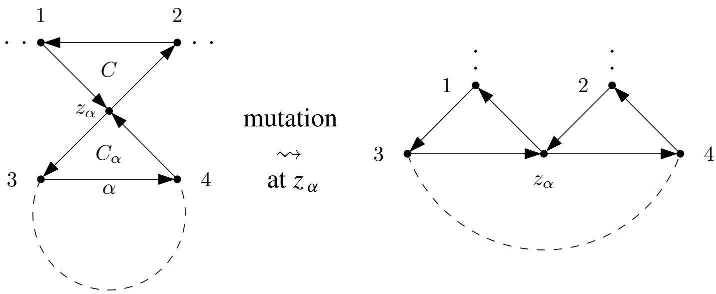


Thus, the length of the path between C^* and C' decreases by 1 and there is a path of length one between C^* and Q_c. The arguments for a quiver with arrow $d \to e$ are analogous and these mutations can also be used if the arrows between d and f are part of the nonoriented cycle (see Step 4).

In this procedure, the parameters r_1, r_2, s_1 and s_2 are left unchanged since we are not changing the number of arrows and the number of oriented 3-cycles which are attached to an oriented 3-cycle sharing one arrow with the nonoriented cycle.

Step 2: We move all oriented 3-cycles onto the nonoriented cycle.

Method: Let C be an oriented 3-cycle that shares one vertex z_α with an oriented 3-cycle C_α sharing an arrow α with the nonoriented cycle. Then we mutate at the vertex z_α:



Hence, both of the oriented 3-cycles share one arrow with the nonoriented cycle and these arrows are oriented as α was before. Thus, the parameters r_1, r_2, s_1 and s_2 are left unchanged. Furthermore, the length of the nonoriented cycle increases by 1. By iterated mutation of that kind, we produce a quiver Q^*, where all the oriented 3-cycles share an arrow with the nonoriented cycle.

Step 3: We move all arrows onto the nonoriented cycle.

Method: This is a similar process as in Step 2: Let C_α be an oriented 3-cycle that shares an arrow α with the nonoriented cycle. All arrows attached to C_α can be moved into the nonoriented cycle by iteratively mutating at vertex z_α. After mutating, all these arrows have the same orientation as α in the nonoriented cycle. Thus, the parameters r_1, r_2, s_1 and s_2 are left unchanged.

Step 4: Move oriented 3-cycles along the nonoriented cycle.

Method: First, we number all oriented 3-cycles by $C_1, \ldots, C_{r_2+s_2}$ in such a way that C_{i+1} follows C_i anticlockwise. As in Step 1, we can move an oriented 3-cycle C_i towards C_{i+1} without changing the orientation of the arrows, that is, without changing the parameters r_1, r_2, s_1 and s_2.

If the nonoriented cycle includes the vertex a in the pictures of Step 1, the arrows between the two cycles move to the top of C_{i+1}, that is, they are no longer part of the nonoriented cycle. However, we can reverse their directions by mutating at the new sinks or sources and insert these arrows into the nonoriented cycle between C_{i+1} and C_{i+2} by mutations like in Step 3 (if C_{i+2} exists).

Doing this iteratively, we produce a quiver Q^* as in Figure 5, with r_1+s_1 arrows that are not part of any oriented 3-cycle and r_2+s_2 oriented 3-cycles sharing one arrow with the nonoriented cycle.

Step 5: Change orientation on the nonoriented cycle to the orientation of Figure 1.

Method: The part of the nonoriented cycle without oriented 3-cycles can be moved to the desired orientation of Figure 1 via sink/source mutations, without mutating

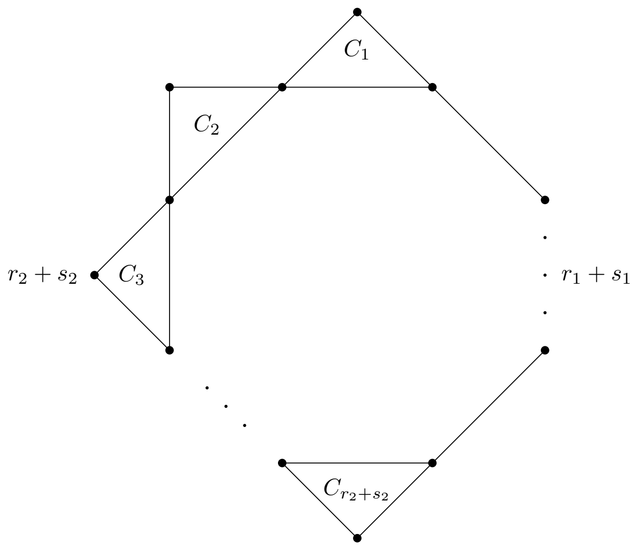
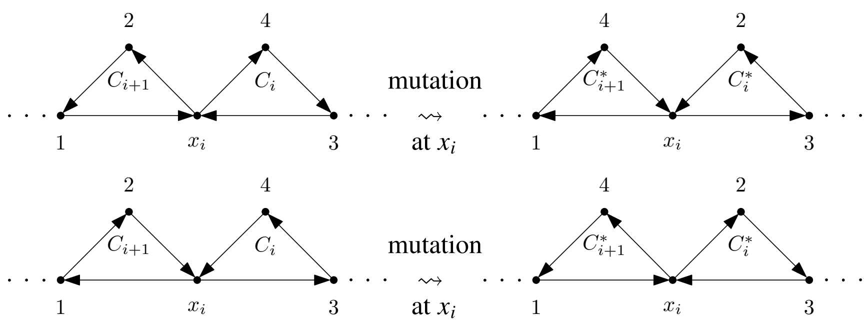


Figure 5. Normal form of Step 4.

at the end vertices that are attached to oriented 3-cycles. Thus, the parameters r_1 and s_1 are left unchanged.

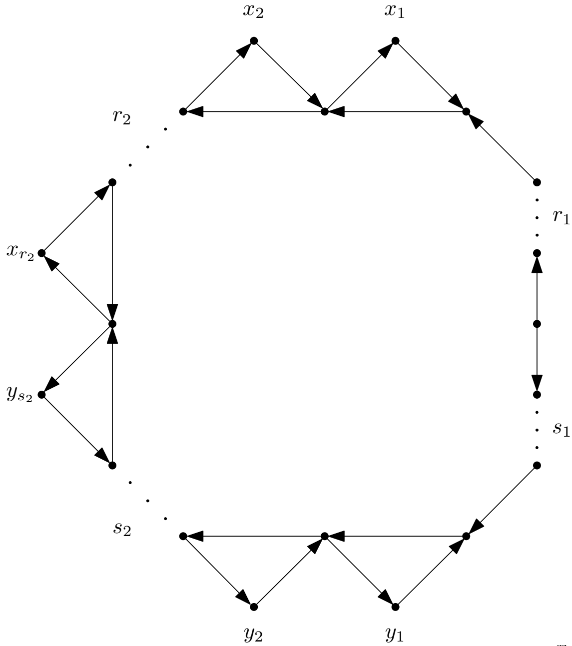
Each oriented 3-cycle shares one arrow with the nonoriented cycle. If all of these arrows are oriented in the same direction, the quiver is in the required form. Thus, we can assume that there are at least two arrows of two oriented 3-cycles C_i and C_{i+1} having opposite orientations. If we mutate at the connecting vertex of C_i and C_{i+1}, the directions of these arrows are changed:



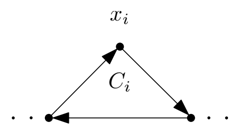
Hence, these mutations act like sink/source mutations at the nonoriented cycle and the parameters r_2 and s_2 are left unchanged. Thus, we can mutate at such connecting vertices as in the part without oriented 3-cycles to reach the desired orientation of Figure 1. $\square$

Theorem 3.10. *Let $Q \in \mathcal{Q}_n$ with parameters r_1, r_2, s_1 and s_2. Then Q is mutation-equivalent to a nonoriented cycle of length $n+1$ with parameters $r = r_1 + 2r_2$ and $s = s_1 + 2s_2$.*

Proof. We can assume that Q is in normal form (see Lemma 3.9) and we label the vertices z_α as follows:

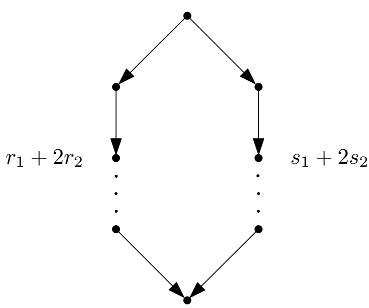


Mutation at the vertex x_i of an oriented 3-cycle



leads to two arrows of the form $\cdots \bullet \leftarrow \underset{x_i}{\bullet} \leftarrow \bullet \cdots$.

Thus, after mutating at all the x_i, the parameter r_2 is zero and we have a new parameter $r = r_1 + 2r_2$. Similarly, we get $s = s_1 + 2s_2$. Hence, mutating at all the x_i and y_i leads to a quiver with underlying graph $\tilde{A}_n$ as follows:



Since there is a nonoriented cycle in every $Q \in \mathcal{Q}_n$, both r and s are nonzero. Thus, the cycle above is also nonoriented. Hence, Q is mutation-equivalent to some quiver of type $\tilde{A}_n$ with parameters $r = r_1 + 2r_2$ and $s = s_1 + 2s_2$. $\square$

Corollary 3.11. *Let $Q_1, Q_2 \in \mathcal{Q}_n$ with parameters $r_1,\ r_2,\ s_1$ and s_2, respectively $\tilde{r}_1,\ \tilde{r}_2,\ \tilde{s}_1$ and $\tilde{s}_2$. If $r_1 + 2r_2 = \tilde{r}_1 + 2\tilde{r}_2$ and $s_1 + 2s_2 = \tilde{s}_1 + 2\tilde{s}_2$, or vice versa, then Q_1 is mutation-equivalent to Q_2.*

Theorem 3.12. *Let $Q_1, Q_2 \in \mathcal{Q}_n$ with parameters $r_1,\ r_2,\ s_1$ and s_2, respectively $\tilde{r}_1,\ \tilde{r}_2,\ \tilde{s}_1$ and $\tilde{s}_2$. Then Q_1 is mutation-equivalent to Q_2 if and only if*

$$r_1 + 2r_2 = \tilde{r}_1 + 2\tilde{r}_2 \quad \textit{and} \quad s_1 + 2s_2 = \tilde{s}_1 + 2\tilde{s}_2$$

or

$$r_1 + 2r_2 = \tilde{s}_1 + 2\tilde{s}_2 \quad \textit{and} \quad s_1 + 2s_2 = \tilde{r}_1 + 2\tilde{r}_2.$$

The "only if" part follows from Theorem 3.10 and Lemma 3.2.

4. Cluster tilted algebras of type $\tilde{A}_n$

In general, cluster tilted algebras arise as endomorphism algebras of cluster-tilting objects in a cluster category [Buan et al. 2007]. Since a cluster tilted algebra A of type $\tilde{A}_n$ is finite dimensional over an algebraically closed field K, there exists a quiver Q which is in the mutation classes of $\tilde{A}_n$ [Buan et al. 2008] and an admissible ideal I of the path algebra KQ of Q such that $A \cong KQ/I$. A nonzero linear combination $k_1\alpha_1 + \cdots + k_m\alpha_m,\ k_i \in K\backslash\{0\}$, of paths α_i of length at least two, with the same starting point and the same end point, is called a *relation* in Q. If $m = 1$, we call such a relation a *zero-relation*. Any admissible ideal of KQ is generated by a finite set of relations in Q.

From [Assem et al. 2010] and [Assem and Redondo 2009], we know that a cluster tilted algebra A of type $\tilde{A}_n$ is gentle, a notion whose definition we recall:

Definition 4.1. We call $A = KQ/I$ a *special biserial algebra* if these properties hold:

(1) Each vertex of Q is the starting point of at most two arrows and the end point of at most two arrows.

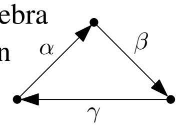
(2) For each arrow α in Q there is at most one arrow β such that $\alpha\beta \notin I$, and at most one arrow γ such that $\gamma\alpha \notin I$.

A is *gentle* if moreover:

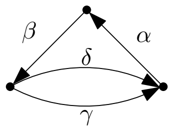
(3) The ideal I is generated by paths of length 2.

(4) For each arrow α in Q there is at most one arrow β' with $t(\alpha) = s(\beta')$ such that $\beta'\alpha \in I$, and there is at most one arrow γ' with $t(\gamma') = s(\alpha)$ such that $\alpha\gamma' \in I$.

Also from the same references, all relations in a cluster tilted algebra A of type $\tilde{A}_n$ occur in the oriented 3-cycles (cycles of the form on the right with (zero-)relations $\alpha\gamma,\ \beta\alpha$ and $\gamma\beta$).



Remark 4.2. According to our convention in Definition 3.3 there are only three (zero-)relations in the quiver



and here, these are $\alpha\delta$, $\beta\alpha$ and $\delta\beta$.

For the next section, we need the notion of Cartan matrices of an algebra A (for example, see [Holm 2005]). Let K be a field and $A = KQ/I$. Since $\sum_{i \in Q_0} e_i + I$ is the unit element in A we get $A = A \cdot 1 = \bigoplus_{i \in Q_0} Ae_i$; hence the (left) A-modules $P_i := Ae_i$ are the indecomposable projective A-modules. The *Cartan matrix* $C = (c_{ij})$ of A is a $|Q_0| \times |Q_0|$-matrix defined by setting $c_{ij} = \dim_K \operatorname{Hom}_A(P_j, P_i)$. Any homomorphism $\varphi : Ae_j \to Ae_i$ of left A-modules is uniquely determined by $\varphi(e_j) \in e_j A e_i$, the K-vector space generated by all paths in Q from vertex i to vertex j that are nonzero in A. In particular, $c_{ij} = \dim_K e_j A e_i$.

That means that computing entries of the Cartan matrix for A reduces to counting paths in Q that are nonzero in A.

5. Derived equivalence classification of cluster tilted algebras of type $\tilde{A}_n$

We briefly review the fundamental results on derived equivalences. For a K-algebra A the bounded derived category of A-modules is denoted by $D^b(A)$. Recall that two algebras A, B are called derived equivalent if $D^b(A)$ and $D^b(B)$ are equivalent as triangulated categories. By a celebrated theorem of Rickard (Theorem 5.2), derived equivalences can be found using the concept of tilting complexes.

Definition 5.1. A tilting complex T over A is a bounded complex of finitely generated projective A-modules satisfying the following conditions:

(i) $\operatorname{Hom}_{D^b(A)}(T, T[i]) = 0$ for all $i \neq 0$, where $[\,\cdot\,]$ denotes the shift functor in $D^b(A)$.

(ii) The category $\operatorname{add}(T)$ (i.e., the full subcategory consisting of direct summands of direct sums of T) generates the homotopy category $K^b(P_A)$ of projective A-modules as a triangulated category.

Theorem 5.2 [Rickard 1989b]. *Two algebras A and B are derived equivalent if and only if there exists a tilting complex T for A such that the endomorphism algebra* $\operatorname{End}_{D^b(A)}(T) \cong B$.

For calculating the endomorphism algebra $\operatorname{End}_{D^b(A)}(T)$ we can use the following alternating sum formula, which gives a general method for computing the Cartan matrix of an endomorphism algebra of a tilting complex from the Cartan matrix of the algebra A.

Proposition 5.3 [Happel 1988]. *For an algebra A let $Q = (Q^r)_{r\in\mathbb{Z}}$ and $R = (R^s)_{s\in\mathbb{Z}}$ be bounded complexes of projective A-modules. Then*

$$\sum_i (-1)^i \dim \mathrm{Hom}_{D^b(A)}(Q, R[i]) = \sum_{r,s} (-1)^{r-s} \dim \mathrm{Hom}_A(Q^r, R^s).$$

In particular, if Q and R are direct summands of the same tilting complex, then

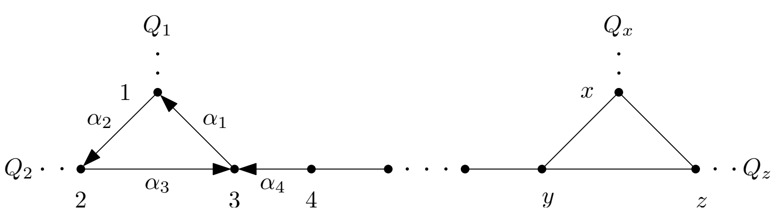
$$\dim \mathrm{Hom}_{D^b(A)}(Q, R) = \sum_{r,s} (-1)^{r-s} \dim \mathrm{Hom}_A(Q^r, R^s).$$

Lemma 5.4. *Let $A = KQ/I$ be a cluster tilted algebra of type $\tilde{A}_n$. Let r_1, r_2, s_1 and s_2 be the parameters of Q that are defined in Definition 3.7. Then A is derived equivalent to a cluster tilted algebra corresponding to a quiver in normal form as in Figure 1.*

Proof. First, the number of oriented 3-cycles with full relations is invariant under derived equivalence for gentle algebras [Holm 2005], so the number $r_2 + s_2$ is an invariant. From [Avella-Alaminos and Geiss 2008, Proposition B], we know that the number of arrows is also invariant under derived equivalence, so the number $r_1 + s_1$ is an invariant, too. Later, we show in the proof of Theorem 5.5 that the single parameters r_1, r_2, s_1 and s_2 are invariant under derived equivalence.

Our strategy in this proof is to go through the proof of Lemma 3.9 and define a tilting complex for each mutation in Steps 1 and 2. We can omit the other three steps since these are just the same situations as in the first two steps. We show that if we mutate at some vertex of the quiver Q and obtain a quiver Q^*, then the two corresponding cluster tilted algebras are derived equivalent.

Step 1: Let A be a cluster tilted algebra with corresponding quiver



We can compute the Cartan matrix to be $\begin{pmatrix} 1 & 1 & 0 & 0 & \dots \\ 0 & 1 & 1 & 0 & \dots \\ 1 & 0 & 1 & 0 & \dots \\ 1 & 0 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$.

Since we are dealing with left modules and read paths from right to left, a nonzero path from vertex i to j gives a homomorphism $P_j \to P_i$ by right multiplication. Thus, two arrows $\alpha : i \to j$ and $\beta : j \to k$ give a path $\beta\alpha$ from i to k and a homomorphism $\alpha\beta : P_k \to P_i$.

In the situation above, we have homomorphisms $P_3 \xrightarrow{\alpha_3} P_2$ and $P_3 \xrightarrow{\alpha_4} P_4$.

Let $T = \bigoplus_{i=1}^{n+1} T_i$ be the following bounded complex of projective A-modules, where $T_i : 0 \to P_i \to 0$, $i \in \{1, 2, 4, \dots, n+1\}$, are complexes concentrated in degree zero and

$$T_3 : 0 \to P_3 \xrightarrow{(\alpha_3,\alpha_4)} P_2 \oplus P_4 \to 0$$

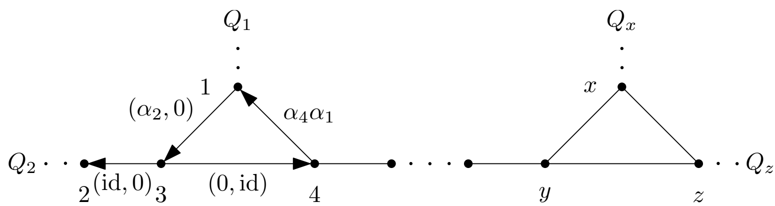
is a complex concentrated in degrees -1 and 0.

We leave it to the reader to verify that this is indeed a tilting complex.

By Rickard's Theorem 5.2, $E := \mathrm{End}_{D^b(A)}(T)$ is derived equivalent to A. Using the alternating sum formula of the Proposition 5.3 of Happel we can compute the Cartan matrix of E to be

$$\begin{pmatrix} 1 & 1 & 1 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ 0 & 1 & 1 & 1 & \dots \\ 1 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

We define homomorphisms in E as follows:



Now we have to check the relations, up to homotopy.

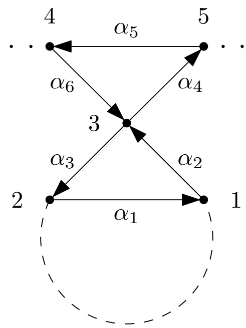
Clearly, the homomorphism $(\alpha_4\alpha_1\alpha_2, 0)$ in the oriented 3-cycle containing the vertices 1, 3 and 4 is zero since $\alpha_1\alpha_2$ was zero in A. Furthermore, the composition of $(\alpha_2, 0)$ and $(0, \mathrm{id})$ yields a zero-relation. The last zero-relation in this oriented 3-cycle is the concatenation of $(0, \mathrm{id})$ and $\alpha_4\alpha_1$ since this homomorphism is homotopic to zero:

$$\begin{tikzcd} & 0 \arrow[r] & P_1 \arrow[r] \arrow[dl, "\alpha_1"'] \arrow[d, "(0,\alpha_4\alpha_1)"] & 0 \\ 0 \arrow[r] & P_3 \arrow[r, "(\alpha_3,\alpha_4)"'] & P_2 \oplus P_4 \arrow[r] & 0 \end{tikzcd}$$

The relations in all other oriented 3-cycles of this quiver are the same as in the quiver of A.

Thus, we have defined homomorphisms between the summands of T corresponding to the arrows of the quiver that we obtain after mutating at vertex 3 in the quiver of A. We have shown that they satisfy the defining relations of this algebra and the Cartan matrices agree. Thus, A is derived equivalent to E and A^{op} is derived equivalent to E^{op}, where the quiver of E is the same as the quiver we obtain after mutating at vertex 3 in the quiver of A. Furthermore, the quivers of A^{op} and E^{op} are the quivers in the other case in Step 1.

Step 2: Let A be a cluster tilted algebra with corresponding quiver



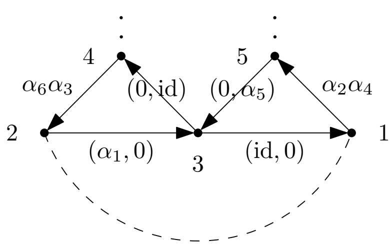
We define a tilting complex T as the bounded complex of projective A-modules $T = \bigoplus_{i=1}^{n+1} T_i$, where $T_i : 0 \to P_i \to 0$, for $i \in \{1, 2, 4, \ldots, n+1\}$, are complexes concentrated in degree zero and $T_3 : 0 \to P_3 \xrightarrow{(\alpha_2,\alpha_6)} P_1 \oplus P_4 \to 0$ is a complex concentrated in degrees -1 and 0.

By Rickard's theorem, $E := \mathrm{End}_{D^b(A)}(T)$ is derived equivalent to A. Using Happel's alternating sum formula (Proposition 5.3), we can compute the Cartan matrix of E to be

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & \ldots \\ 1 & 1 & 1 & 0 & 0 & \ldots \\ 1 & 0 & 1 & 1 & 0 & \ldots \\ 0 & 1 & 0 & 1 & 0 & \ldots \\ 0 & 0 & 1 & 1 & 1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

(This deals with the case where not all the arrows between 2 and 1 along the nonoriented cycle are oriented in the same direction. The case where they are can be handled similarly.)

We define homomorphisms in E as follows:



Thus, A is derived equivalent to E and A^{op} is derived equivalent to E^{op}, where the quiver of E is the same as the quiver we obtain after mutating at 3.

In Steps 3 and 4 of the proof of Lemma 3.9 we mutate at a vertex with three incident arrows as in Step 1. In Step 5 we mutate at sinks, sources and at vertices with four incident arrows as in Step 2.

Thus, we obtain a quiver of a derived equivalent cluster tilted algebra by all mutations in the proof of Lemma 3.9. Hence, every cluster tilted algebra $A = KQ/I$ of type $\tilde{A}_n$ is derived equivalent to a cluster tilted algebra with a quiver in normal form having the same parameters as Q. □

Our next aim is to prove the main result:

Theorem 5.5. *Two cluster tilted algebras of type $\tilde{A}_n$ are derived equivalent if and only if their quivers have the same parameters r_1, r_2, s_1 and s_2, up to changing the roles of r_i and s_i for $i \in \{1, 2\}$.*

But first, we recall some background from [Avella-Alaminos and Geiss 2008]. Let $A = KQ/I$ be a gentle algebra, where $Q = (Q_0, Q_1)$ is a connected quiver. A *permitted path* of A is a path $C = \alpha_l \dots \alpha_2\alpha_1$ that contains no zero-relations. A permitted path C is called a *nontrivial permitted thread* if for all $\beta \in Q_1$ neither $C\beta$ nor βC is a permitted path. Similarly a *forbidden path* of A is a sequence $\Pi = \alpha_l \dots \alpha_2\alpha_1$ formed by pairwise different arrows in Q with $\alpha_{i+1}\alpha_i \in I$ for all $i \in \{1, 2, \dots, l-1\}$. A forbidden path Π is called a *nontrivial forbidden thread* if for all $\beta \in Q_1$ neither $\Pi\beta$ nor $\beta\Pi$ is a forbidden path. Let $v \in Q_0$ such that $\#\{\alpha \in Q_1 : s(\alpha) = v\} \leq 1$, $\#\{\alpha \in Q_1 : t(\alpha) = v\} \leq 1$ and if $\beta, \gamma \in Q_1$ are such that $s(\gamma) = v = t(\beta)$, then $\gamma\beta \notin I$. Then we consider e_v a *trivial permitted thread* in v and denote it by h_v. Let $\mathcal{H}_A$ be the set of all permitted threads of A, trivial and nontrivial. Similarly, for $v \in Q_0$ such that $\#\{\alpha \in Q_1 : s(\alpha) = v\} \leq 1$, $\#\{\alpha \in Q_1 : t(\alpha) = v\} \leq 1$ and if $\beta, \gamma \in Q_1$ are such that $s(\gamma) = v = t(\beta)$, then $\gamma\beta \in I$, we consider e_v a *trivial forbidden thread* in v and denote it by p_v. Note that certain paths can be permitted and forbidden threads simultaneously.

Now, one can define functions $\sigma, \varepsilon : Q_1 \to \{1, -1\}$ that satisfy these conditions:

(1) If $\beta_1 \neq \beta_2$ are arrows with $s(\beta_1) = s(\beta_2)$, then $\sigma(\beta_1) = -\sigma(\beta_2)$.

(2) If $\gamma_1 \neq \gamma_2$ are arrows with $t(\gamma_1) = t(\gamma_2)$, then $\varepsilon(\gamma_1) = -\varepsilon(\gamma_2)$.

(3) If β and γ are arrows with $s(\gamma) = t(\beta)$ and $\gamma\beta \notin I$, then $\sigma(\gamma) = -\varepsilon(\beta)$.

We can extend these functions to threads of A as follows: for a nontrivial thread $H = \alpha_l \dots \alpha_2\alpha_1$ of A define $\sigma(H) := \sigma(\alpha_1)$ and $\varepsilon(H) := \varepsilon(\alpha_l)$. If there is a trivial permitted thread h_v for some $v \in Q_0$, the connectivity of Q assures the existence of some $\gamma \in Q_1$ with $s(\gamma) = v$ or some $\beta \in Q_1$ with $t(\beta) = v$. In the first case, we define $\sigma(h_v) = -\varepsilon(h_v) := -\sigma(\gamma)$, for the second case $\sigma(h_v) = -\varepsilon(h_v) := \varepsilon(\beta)$. If there is a trivial forbidden thread p_v for some $v \in Q_0$, we know that there exists $\gamma \in Q_1$ with $s(\gamma) = v$ or $\beta \in Q_1$ with $t(\beta) = v$. In the first case, we define $\sigma(p_v) = \varepsilon(h_v) := -\sigma(\gamma)$, for the second case $\sigma(p_v) = \varepsilon(h_v) := -\varepsilon(\beta)$.

We next use a combinatorial algorithm to produce certain pairs of natural numbers, using only the quiver with relations which defines a gentle algebra. In the algorithm we go forward through permitted threads and backwards through forbidden threads in such a way that each arrow and its inverse are used exactly once.

Algorithm 5.6 [Avella-Alaminos and Geiss 2008].

(1) Begin with a permitted thread H_0 of A.

- If H_i is defined, consider Π_i the forbidden thread that ends in $t(H_i)$ and such that $\varepsilon(H_i) = -\varepsilon(\Pi_i)$.
- Let H_{i+1} be the permitted thread that starts in $s(\Pi_i)$ and such that $\sigma(H_{i+1}) = -\sigma(\Pi_i)$.

The process stops when $H_k = H_0$ for some natural number k. Set

$$m = \sum_{1 \le i \le k} l(\Pi_{i-1}),$$

where $l(\cdot)$ is the length (number of arrows) of a path. We obtain the pair (k, m).

(2) Repeat the first step of the algorithm until all permitted threads of A have been considered.

(3) If there are oriented cycles in which each pair of consecutive arrows form a relation, we add a pair $(0, m)$ for each of those cycles, where m is the length of the cycle.

(4) Define $\phi_A : \mathbb{N}^2 \to \mathbb{N}$, where $\phi_A(k, m)$ is the number of times the pair (k, m) arises in the algorithm.

This function ϕ is invariant under derived equivalence:

Lemma 5.7 [Avella-Alaminos and Geiss 2008]. *Let A and B be gentle algebras. If A and B are derived equivalent, then $\phi_A = \phi_B$.*

Example 5.8. Figure 6 shows the quiver of a cluster tilted algebra A of type $\tilde{A}_{18}$, where $r_1 = 2$, $r_2 = 3$, $s_1 = 3$ and $s_2 = 4$ and thus, $r := r_1 + r_2 = 5$ and $s := s_1 + s_2 = 7$.

Define the functions σ and ε for all arrows in Q:

$$\begin{aligned}
\sigma(\alpha_i) &= 1, & \varepsilon(\alpha_i) &= -1 & &\text{for all } i = 1, \dots, 5,\\
\sigma(\alpha_i) &= -1, & \varepsilon(\alpha_i) &= 1 & &\text{for all } i = 6, \dots, 12,\\
\sigma(\beta_{j,1}) &= 1, & \varepsilon(\beta_{j,1}) &= 1 & &\text{for all } j = 1, \dots, 3,\\
\sigma(\beta_{j,2}) &= -1, & \varepsilon(\beta_{j,2}) &= 1 & &\text{for all } j = 1, \dots, 3,\\
\sigma(\gamma_{l,1}) &= -1, & \varepsilon(\gamma_{l,1}) &= -1 & &\text{for all } l = 1, \dots, 4,\\
\sigma(\gamma_{l,2}) &= 1, & \varepsilon(\gamma_{l,2}) &= -1 & &\text{for all } l = 1, \dots, 4.
\end{aligned}$$

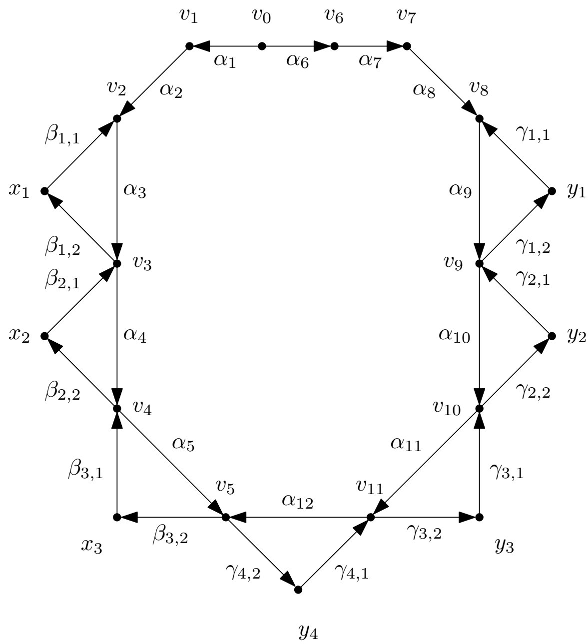


Figure 6. Quiver for Example 5.8.

Then $\mathcal{H}_A$ is formed by h_{v_1}, h_{v_6}, h_{v_7}, $\gamma_{4,2}\alpha_5\alpha_4\alpha_3\alpha_2\alpha_1$, $\beta_{3,2}\alpha_{12}\alpha_{11}\alpha_{10}\alpha_9\alpha_8\alpha_7\alpha_6$, $\beta_{1,1}$, $\beta_{1,2}\beta_{2,1}$, $\beta_{2,2}\beta_{3,1}$, $\gamma_{1,1}$, $\gamma_{1,2}\gamma_{2,1}$, $\gamma_{2,2}\gamma_{3,1}$ and $\gamma_{3,2}\gamma_{4,1}$. The forbidden threads of A are p_{x_1}, p_{x_2}, p_{x_3}, p_{y_1}, p_{y_2}, p_{y_3}, p_{y_4}, α_1, α_2, α_6, α_7, α_8 and all the oriented 3-cycles.

Moreover, we can write

$$\begin{aligned}
\sigma(h_{v_1}) &= -\varepsilon(h_{v_1}) = -\sigma(\alpha_2) = \varepsilon(\alpha_1) = -1,\\
\sigma(h_{v_6}) &= -\varepsilon(h_{v_6}) = -\sigma(\alpha_7) = \varepsilon(\alpha_6) = \quad 1,\\
\sigma(h_{v_7}) &= -\varepsilon(h_{v_7}) = -\sigma(\alpha_8) = \varepsilon(\alpha_7) = \quad 1
\end{aligned}$$

for the trivial permitted threads and

$$\begin{aligned}
\sigma(p_{x_i}) &= \varepsilon(p_{x_i}) = -\sigma(\beta_{i,1}) = -\varepsilon(\beta_{i,2}) = -1 \quad \text{for all } i = 1, 2, 3,\\
\sigma(p_{y_i}) &= \varepsilon(p_{y_i}) = -\sigma(\gamma_{i,1}) = -\varepsilon(\gamma_{i,2}) = \quad 1 \quad \text{for all } i = 1, 2, 3, 4
\end{aligned}$$

for the trivial forbidden threads.

Let $H_0 = h_{v_1}$ and $\Pi_0 = \alpha_1$ with $\varepsilon(h_{v_1}) = -\varepsilon(\alpha_1) = 1$. Then H_1 is the permitted thread that starts in $s(\Pi_0) = v_0$ and $\sigma(H_1) = \sigma(\alpha_6) = -\sigma(\Pi_0) = -1$, that is, $\beta_{3,2}\alpha_{12}\alpha_{11}\alpha_{10}\alpha_9\alpha_8\alpha_7\alpha_6$. Now $\Pi_1 = p_{x_3}$ since it is the forbidden thread that ends in x_3 and $\varepsilon(\Pi_1) = -\varepsilon(H_1) = -\varepsilon(\beta_{3,2}) = -1$. Then $H_2 = \beta_{2,2}\beta_{3,1}$ is the permitted thread starting in x_3 and $\sigma(\Pi_1) = -\sigma(H_2) = -\sigma(\beta_{3,1}) = -1$. Thus, $\Pi_2 = p_{x_2}$ with $\varepsilon(H_2) = \varepsilon(\beta_{2,2}) = -\varepsilon(\Pi_2) = 1$.

In the same way we can define the missing threads and we get

$$\begin{array}{ll}
H_0 = h_{v_1} & \Pi_0^{-1} = \alpha_1^{-1} \\
H_1 = \beta_{3,2}\alpha_{12}\alpha_{11}\alpha_{10}\alpha_9\alpha_8\alpha_7\alpha_6 & \Pi_1^{-1} = p_{x_3} \\
H_2 = \beta_{2,2}\beta_{3,1} & \Pi_2^{-1} = p_{x_2} \\
H_3 = \beta_{1,2}\beta_{2,1} & \Pi_3^{-1} = p_{x_1} \\
H_4 = \beta_{1,1} & \Pi_4^{-1} = \alpha_2^{-1} \\
H_5 = H_0 &
\end{array}$$

$$\to (5, 2)$$

where α_1^{-1} is defined by $s(\alpha_1^{-1}) := t(\alpha_1),\ t(\alpha_1^{-1}) := s(\alpha_1)$ and $(\alpha_1^{-1})^{-1} = \alpha_1$.

If we continue with the algorithm we obtain the second pair $(7, 3) = (s, s_1)$ in the following way:

$$\begin{array}{ll}
H_0 = h_{v_6} & \Pi_0^{-1} = \alpha_6^{-1} \\
H_1 = \gamma_{4,2}\alpha_5\alpha_4\alpha_3\alpha_2\alpha_1 & \Pi_1^{-1} = p_{y_4} \\
H_2 = \gamma_{3,2}\gamma_{4,1} & \Pi_2^{-1} = p_{y_3} \\
H_3 = \gamma_{2,2}\gamma_{3,1} & \Pi_3^{-1} = p_{y_2} \\
H_4 = \gamma_{1,2}\gamma_{2,1} & \Pi_4^{-1} = p_{y_1} \\
H_5 = \gamma_{1,1} & \Pi_5^{-1} = \alpha_8^{-1} \\
H_6 = h_{v_7} & \Pi_6^{-1} = \alpha_7^{-1} \\
H_7 = H_0 &
\end{array}$$

$$\to (7, 3)$$

Finally, we have to add seven pairs $(0, 3)$ for the seven oriented 3-cycles. Thus, we get $\phi_A(5, 2) = 1,\ \phi_A(7, 3) = 1$, and $\phi_A(0, 3) = 7$.

Now we can extend this example to general quivers of cluster tilted algebras of type $\tilde{A}_n$ in normal form.

Proof of Theorem 5.5. We know from Lemma 5.4 that every cluster tilted algebra $A = KQ/I$ of type $\tilde{A}_n$ with parameters $r_1,\ r_2,\ s_1$ and s_2 is derived equivalent to a cluster tilted algebra with a quiver in normal form, as shown in Figure 1, where r_1 is the number of arrows anticlockwise that do not share any arrow with an oriented 3-cycle and s_1 is the number of arrows clockwise that do not share any arrow with an oriented 3-cycle. Moreover, r_2 is the number of oriented 3-cycles that share one arrow α with the nonoriented cycle and α is oriented anticlockwise and s_2 is the number of oriented 3-cycles that share one arrow β with the nonoriented cycle and β is oriented clockwise (see Definition 3.7). Thus, $r := r_1 + r_2$ is the number of anticlockwise arrows of the nonoriented cycle and $s := s_1 + s_2$ is the number of clockwise arrows of the nonoriented cycle.

We consider the quiver Q in normal form with notation as given in Figure 7 and define the functions σ and ε for all arrows in Q:

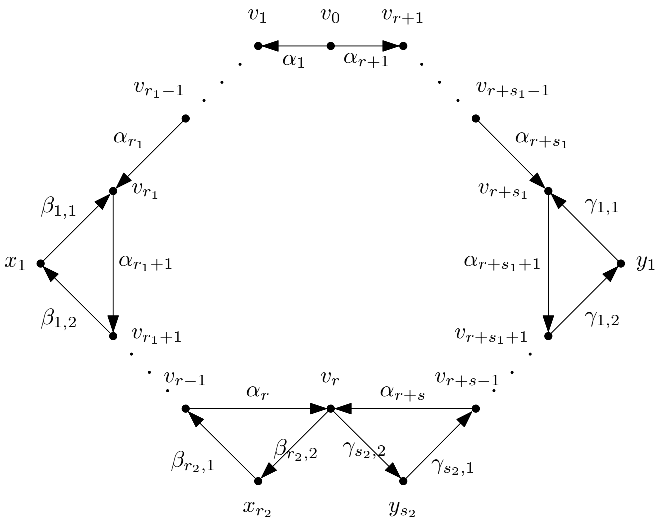


Figure 7. A quiver in normal form.

$$\begin{array}{llll}
\sigma(\alpha_i) = 1, & \varepsilon(\alpha_i) = -1 & \text{for all} & i = 1, \ldots, r, \\
\sigma(\alpha_i) = -1, & \varepsilon(\alpha_i) = 1 & \text{for all} & i = r+1, \ldots, r+s, \\
\sigma(\beta_{j,1}) = 1, & \varepsilon(\beta_{j,1}) = 1 & \text{for all} & j = 1, \ldots, r_2, \\
\sigma(\beta_{j,2}) = -1, & \varepsilon(\beta_{j,2}) = 1 & \text{for all} & j = 1, \ldots, r_2, \\
\sigma(\gamma_{l,1}) = -1, & \varepsilon(\gamma_{l,1}) = -1 & \text{for all} & l = 1, \ldots, s_2, \\
\sigma(\gamma_{l,2}) = 1, & \varepsilon(\gamma_{l,2}) = -1 & \text{for all} & l = 1, \ldots, s_2.
\end{array}$$

Here $\mathcal{H}_A$ is formed by

$$h_{v_1}, \ldots, h_{v_{r_1-1}}, h_{v_{r+1}}, \ldots, h_{v_{r+s_1-1}}, \gamma_{s_2,2}\alpha_r\alpha_{r-1}\ldots\alpha_2\alpha_1, \\ \beta_{r_2,2}\alpha_{r+s}\alpha_{r+s-1}\ldots\alpha_{r+2}\alpha_{r+1}, \beta_{1,1}, \beta_{1,2}\beta_{2,1}, \ldots, \beta_{r_2-1,2}\beta_{r_2,1}, \\ \gamma_{1,1}, \gamma_{1,2}\gamma_{2,1}, \ldots, \gamma_{s_2-1,2}\gamma_{s_2,1}.$$

The forbidden threads of A are $p_{x_1}, \ldots, p_{x_{r_2}}, p_{y_1}, \ldots, p_{y_{s_2}}, \alpha_1, \ldots, \alpha_{r_1}, \alpha_{r+1}, \ldots, \alpha_{r+s_1}$ and all the oriented 3-cycles.

Moreover, we can write

$$\begin{array}{lllll}
\sigma(h_{v_1}) & = -\varepsilon(h_{v_1}) & = -\sigma(\alpha_2) & = \varepsilon(\alpha_1) & = -1, \\
 & \vdots & & & \\
\sigma(h_{v_{r_1-1}}) & = -\varepsilon(h_{v_{r_1-1}}) & = -\sigma(\alpha_{r_1}) & = \varepsilon(\alpha_{r_1-1}) & = -1, \\
\sigma(h_{v_{r+1}}) & = -\varepsilon(h_{v_{r+1}}) & = -\sigma(\alpha_{r+2}) & = \varepsilon(\alpha_{r+1}) & = 1, \\
 & \vdots & & & \\
\sigma(h_{v_{r+s_1-1}}) & = -\varepsilon(h_{v_{r+s_1-1}}) & = -\sigma(\alpha_{r+s_1}) & = \varepsilon(\alpha_{r+s_1-1}) & = 1
\end{array}$$

for the trivial permitted threads and

$$\begin{aligned}
\sigma(p_{x_i}) = \varepsilon(p_{x_i}) = -\sigma(\beta_{i,1}) = -\varepsilon(\beta_{i,2}) &= -1 \quad \text{for all } i = 1, \dots, r_2, \\
\sigma(p_{y_i}) = \varepsilon(p_{y_i}) = -\sigma(\gamma_{i,1}) = -\varepsilon(\gamma_{i,2}) &= 1 \quad \text{for all } i = 1, \dots, s_2
\end{aligned}$$

for the trivial forbidden threads.

Thus, we can apply Algorithm 5.6 as follows:

$$\begin{array}{llll}
H_0 & = h_{v_1} & \Pi_0^{-1} & = \alpha_1^{-1} \\
H_1 & = \beta_{r_2,2}\alpha_{r+s}\alpha_{r+s-1}\dots\alpha_{r+2}\alpha_{r+1} & \Pi_1^{-1} & = p_{x_{r_2}} \\
H_2 & = \beta_{r_2-1,2}\beta_{r_2,1} & \Pi_2^{-1} & = p_{x_{r_2-1}} \\
\vdots & & \vdots & \\
H_{r_2} & = \beta_{1,2}\beta_{2,1} & \Pi_{r_2}^{-1} & = p_{x_1} \\
H_{r_2+1} & = \beta_{1,1} & \Pi_{r_2+1}^{-1} & = \alpha_{r_1}^{-1} \\
H_{r_2+2} & = h_{v_{r_1-1}} & \Pi_{r_2+2}^{-1} & = \alpha_{r_1-1}^{-1} \\
\vdots & & \vdots & \\
H_{r-1} & = h_{v_2} & \Pi_{r-1}^{-1} & = \alpha_2^{-1} \\
H_r & = H_0 & &
\end{array}$$

$$\begin{aligned}
m &= l(\Pi_0) + l(\Pi_{r_2+1}) + l(\Pi_{r_2+2}) + \dots + l(\Pi_{r-1}) \\
&= 1 + \underbrace{1 + 1 + \dots + 1}_{(r-1)-r_2 \text{ times}} \\
&= 1 + (r-1) - r_2 \\
&= r - r_2 \\
&= r_1
\end{aligned}$$

$$\rightarrow (r, r_1)$$

If we continue with the algorithm we obtain the second pair (s, s_1) as follows:

$$\begin{array}{llll}
H_0 & = h_{v_{r+1}} & \Pi_0^{-1} & = \alpha_{r+1}^{-1} \\
H_1 & = \gamma_{s_2,2}\alpha_r\alpha_{r-1}\dots\alpha_2\alpha_1 & \Pi_1^{-1} & = p_{y_{s_2}} \\
H_2 & = \gamma_{s_2-1,2}\gamma_{s_2,1} & \Pi_2^{-1} & = p_{y_{s_2-1}} \\
\vdots & & \vdots & \\
H_{s_2} & = \gamma_{1,2}\gamma_{2,1} & \Pi_{s_2}^{-1} & = p_{y_1} \\
H_{s_2+1} & = \gamma_{1,1} & \Pi_{s_2+1}^{-1} & = \alpha_{r+s_1}^{-1} \\
H_{s_2+2} & = h_{v_{r+s_1-1}} & \Pi_{s_2+2}^{-1} & = \alpha_{r+s_1-1}^{-1} \\
\vdots & & \vdots & \\
H_{s-1} & = h_{v_{r+2}} & \Pi_{s-1}^{-1} & = \alpha_{r+2}^{-1} \\
H_s & = H_0 & &
\end{array}$$

$$\rightarrow (s, s_1)$$

Finally, we have to add $r_2 + s_2$ pairs $(0, 3)$ for the oriented 3-cycles. Thus, we have $\phi_A(r, r_1) = 1$, $\phi_A(s, s_1) = 1$ and $\phi_A(0, 3) = r_2 + s_2$, where $r = r_1 + r_2$ and $s = s_1 + s_2$.

Now, let A and B be two cluster tilted algebras of type $\tilde{A}_n$ with parameters r_1, r_2, s_1, s_2, respectively $\tilde{r}_1$, $\tilde{r}_2$, $\tilde{s}_1$, $\tilde{s}_2$. From above we can conclude that $\phi_A = \phi_B$ if and only if $r_1 = \tilde{r}_1$, $r_2 = \tilde{r}_2$, $s_1 = \tilde{s}_1$ and $s_2 = \tilde{s}_2$ or $r_1 = \tilde{s}_1$, $r_2 = \tilde{s}_2$, $s_1 = \tilde{r}_1$ and $s_2 = \tilde{r}_2$ (which ends up being the same quiver).

Hence, if A is derived equivalent to B, we know from Lemma 5.7 that $\phi_A = \phi_B$ and thus, that the parameters are the same. Otherwise, if A and B have the same parameters, they are both derived equivalent to the same cluster tilted algebra with a quiver in normal form. □

Acknowledgement

The author thanks Thorsten Holm for many helpful suggestions and discussions about the topics of this work.

References

[Assem and Redondo 2009] I. Assem and M. J. Redondo, "The first Hochschild cohomology group of a Schurian cluster-tilted algebra", *Manuscripta Math.* **128**:3 (2009), 373–388. MR 2009m:16018 Zbl 1211.16007

[Assem et al. 2008a] I. Assem, T. Brüstle, and R. Schiffler, "Cluster-tilted algebras and slices", *J. Algebra* **319**:8 (2008), 3464–3479. MR 2009f:16021 Zbl 1159.16011

[Assem et al. 2008b] I. Assem, T. Brüstle, and R. Schiffler, "Cluster-tilted algebras as trivial extensions", *Bull. Lond. Math. Soc.* **40**:1 (2008), 151–162. MR 2009c:16086 Zbl 1182.16009

[Assem et al. 2010] I. Assem, T. Brüstle, G. Charbonneau-Jodoin, and P.-G. Plamondon, "Gentle algebras arising from surface triangulations", *Algebra Number Theory* **4**:2 (2010), 201–229. MR 2011h:16045 Zbl 05704468

[Avella-Alaminos and Geiss 2008] D. Avella-Alaminos and C. Geiss, "Combinatorial derived invariants for gentle algebras", *J. Pure Appl. Algebra* **212**:1 (2008), 228–243. MR 2008i:16014 Zbl 1143.16016

[Buan and Vatne 2008] A. B. Buan and D. F. Vatne, "Derived equivalence classification for cluster-tilted algebras of type A_n", *J. Algebra* **319**:7 (2008), 2723–2738. MR 2009j:16013 Zbl 1155.16010

[Buan et al. 2006] A. B. Buan, R. J. Marsh, M. Reineke, I. Reiten, and G. Todorov, "Tilting theory and cluster combinatorics", *Adv. Math.* **204**:2 (2006), 572–618. MR 2007f:16033 Zbl 1127.16011

[Buan et al. 2007] A. B. Buan, R. J. Marsh, and I. Reiten, "Cluster-tilted algebras", *Trans. Amer. Math. Soc.* **359**:1 (2007), 323–332. MR 2007f:16035 Zbl 1123.16009

[Buan et al. 2008] A. B. Buan, R. J. Marsh, and I. Reiten, "Cluster mutation via quiver representations", *Comment. Math. Helv.* **83**:1 (2008), 143–177. MR 2008k:16026 Zbl 1193.16016

[Caldero et al. 2006] P. Caldero, F. Chapoton, and R. Schiffler, "Quivers with relations and cluster tilted algebras", *Algebr. Represent. Theory* **9**:4 (2006), 359–376. MR 2007f:16036 Zbl 1127.16013

[Derksen and Owen 2008] H. Derksen and T. Owen, "New graphs of finite mutation type", *Electron. J. Combin.* **15**:1 (2008), Research Paper 139. MR 2009j:05102 Zbl 1180.05052

[Derksen et al. 2008] H. Derksen, J. Weyman, and A. Zelevinsky, "Quivers with potentials and their representations, I: Mutations", *Selecta Math.* (*N.S.*) **14**:1 (2008), 59–119. MR 2010b:16021 Zbl 1204.16008

[Fomin and Zelevinsky 2002] S. Fomin and A. Zelevinsky, "Cluster algebras, I: Foundations", *J. Amer. Math. Soc.* **15**:2 (2002), 497–529. MR 2003f:16050 Zbl 1021.16017

[Fomin and Zelevinsky 2003] S. Fomin and A. Zelevinsky, "Cluster algebras, II: Finite type classification", *Invent. Math.* **154**:1 (2003), 63–121. MR 2004m:17011 Zbl 1054.17024

[Fomin et al. 2008] S. Fomin, M. Shapiro, and D. Thurston, "Cluster algebras and triangulated surfaces, I: Cluster complexes", *Acta Math.* **201**:1 (2008), 83–146. MR 2010b:57032 Zbl 05508850

[Happel 1988] D. Happel, *Triangulated categories in the representation theory of finite dimensional algebras*, London Mathematical Society Lecture Note Series **119**, Cambridge University Press, Cambridge, 1988. MR 89e:16035 Zbl 0635.16017

[Holm 2005] T. Holm, "Cartan determinants for gentle algebras", *Arch. Math.* (*Basel*) **85**:3 (2005), 233–239. MR 2006e:16022 Zbl 1084.16009

[Krause 1997] H. Krause, "Stable equivalence preserves representation type", *Comment. Math. Helv.* **72**:2 (1997), 266–284. MR 98k:16018 Zbl 0901.16008

[Labardini-Fragoso 2009] D. Labardini-Fragoso, "Quivers with potentials associated to triangulated surfaces", *Proc. Lond. Math. Soc.* (3) **98**:3 (2009), 797–839. MR 2010b:16033 Zbl 05551833

[Murphy 2010] G. J. Murphy, "Derived equivalence classification of m-cluster tilted algebras of type A_n", *J. Algebra* **323**:4 (2010), 920–965. MR 2011e:16019 Zbl 1207.16014

[Rickard 1989a] J. Rickard, "Derived categories and stable equivalence", *J. Pure Appl. Algebra* **61**:3 (1989), 303–317. MR 91a:16004 Zbl 0685.16016

[Rickard 1989b] J. Rickard, "Morita theory for derived categories", *J. London Math. Soc.* (2) **39**:3 (1989), 436–456. MR 91b:18012 Zbl 0642.16034

[Rickard 1991] J. Rickard, "Derived equivalences as derived functors", *J. London Math. Soc.* (2) **43**:1 (1991), 37–48. MR 92b:16043 Zbl 0683.16030

[Seven 2007] A. I. Seven, "Recognizing cluster algebras of finite type", *Electron. J. Combin.* **14**:1 (2007), Research Paper 3. MR 2008c:05192 Zbl 1114.05103

[Vatne 2010] D. F. Vatne, "The mutation class of D_n quivers", *Comm. Algebra* **38**:3 (2010), 1137–1146. MR 2011c:13049 Zbl 1196.16011

Communicated by Andrei Zelevinsky
Received 2010-03-04 Revised 2010-07-26 Accepted 2010-08-31

bastian@math.uni-hannover.de *Leibniz Universität Hannover, Institut für Algebra, Zahlentheorie und Diskrete Mathematik, Welfengarten 1, D-30167 Hannover, Germany*

ALGEBRA AND NUMBER THEORY 5:5(2011)

Kazhdan–Lusztig polynomials and drift configurations

Li Li and Alexander Yong

The coefficients of the Kazhdan–Lusztig polynomials $P_{v,w}(q)$ are nonnegative integers that are upper semicontinuous relative to Bruhat order. Conjecturally, the same properties hold for h-polynomials $H_{v,w}(q)$ of local rings of Schubert varieties. This suggests a parallel between the two families of polynomials. We prove our conjectures for Grassmannians, and more generally, covexillary Schubert varieties in complete flag varieties, by deriving a combinatorial formula for $H_{v,w}(q)$. We introduce *drift configurations* to formulate a new and compatible combinatorial rule for $P_{v,w}(q)$. From our rules we deduce, for these cases, the coefficient-wise inequality $P_{v,w}(q) \preceq H_{v,w}(q)$.

1. Introduction

Overview. This paper studies two families of polynomials $\{P_{v,w}(q)\}$ and $\{H_{v,w}(q)\}$ defined for pairs of permutations v, w in the symmetric group S_n (more generally, any Weyl group W). The former family consists of the celebrated *Kazhdan–Lusztig polynomials*, introduced in [Kazhdan and Lusztig 1979] to study representations of Hecke algebras. There it was conjectured that $P_{v,w}(q) \in \mathbb{Z}_{\geq 0}[q]$. This was later established by the same authors [1980] by interpreting $P_{v,w}(q)$ as the Poincaré polynomial for Goresky–MacPherson's local intersection cohomology for the torus fixed point e_v of the Schubert variety X_w in the complete flag variety $\mathrm{Flags}(\mathbb{C}^n)$.

A key contribution to the theory is R. Irving's theorem [1988] that the $P_{v,w}(q)$ are *upper semicontinuous*: if $v' \leq v \leq w$ in Bruhat order, then $P_{v,w}(q) \preceq P_{v',w}(q)$, where "$\preceq$" means that, for each i, the coefficient of q^i in $P_{v,w}(q)$ is weakly smaller than the coefficient of q^i in $P_{v',w}(q)$. Thus, the Kazhdan–Lusztig polynomials are measures of the singularities of Schubert varieties whose coefficient growth tracks the worsening pathology of singularities as one moves along torus invariant $\mathbb{P}^1$'s towards the "most singular" point $e_{\mathrm{id}} \in X_w$. In particular, $P_{v,w}(q) = 1$ if and only if $e_v \in X_w$ is a (rationally) smooth point.

MSC2000: primary 14M15; secondary 05E15, 20F55.

Keywords: Kazhdan–Lusztig polynomials, Hilbert series, Schubert varieties.

Conversely, the desire for insight into the combinatorics of Kazhdan–Lusztig polynomials naturally leads to the basic problem of understanding where and how the singularities of Schubert varieties worsen. In view of this converse problem, the growth of any semicontinuous singularity measure of Schubert varieties is of interest. One seeks concrete comparisons of different measures; see, e.g., [Woo and Yong 2008] and the references therein.

Specifically, a well-studied semicontinuous measure is given by the *Hilbert–Samuel multiplicity* $\mathrm{mult}_{e_v}(X_w)$. However, while this contains useful local data about X_w, even more is carried by the $\mathbb{Z}$-graded Hilbert series of $\mathrm{gr}_{\mathfrak{m}_{e_v}}\mathbb{O}_{e_v,X_w}$, the associated graded ring of the local ring $\mathbb{O}_{e_v,X_w}$,

$$\mathrm{Hilb}(\mathrm{gr}_{\mathfrak{m}_{e_v}}\mathbb{O}_{e_v,X_w}, q) = \frac{H_{v,w}(q)}{(1-q)^{\ell(w)}},$$

where $\ell(w) = \dim(X_w)$ is the *Coxeter length* of w. In particular, $\mathrm{mult}_{e_v}(X_w) = H_{v,w}(1)$.

Conjecturally, each *h-polynomial* $H_{v,w}(q)$ is also in $\mathbb{Z}_{\geq 0}[q]$, and moreover is upper semicontinuous, just as is the case for Kazhdan–Lusztig polynomials. These conjectures suggest that the growth of the coefficients of the two families of polynomials is somehow correlated. In this paper, we offer an examination in the Grassmannian case, and more generally in the case of covexillary Schubert varieties inside $\mathrm{Flags}(\mathbb{C}^n)$. There the nonnegativity and semicontinuity conjectures are proved by deriving a new combinatorial rule for $H_{v,w}(q)$. In addition, by introducing *drift configurations* as a model for the Kazhdan–Lusztig polynomials in these settings (after [Lascoux and Schützenberger 1981] and [Lascoux 1995]), we prove the inequality $P_{v,w}(q) \preceq H_{v,w}(q)$. This combinatorial discovery further indicates the link between the two families; no alternative explanation via algebraic or geometric methods seems available at present.

Summarizing, the main thesis of this paper is that there exists a parallel between $\{P_{v,w}(q)\}$ and $\{H_{v,w}(q)\}$. Our basis for this perspective comes from proofs of compatible and positive combinatorial rules for the two families of polynomials.

Statements of the main conjecture and theorems. Recapitulating, this paper formulates, and constructs supporting combinatorics for, the following conjecture:

Conjecture 1.1. *The h-polynomials $H_{v,w}(q)$ have nonnegative integral coefficients. In addition, they are upper semicontinuous; i.e., if $v' \leq v$ in Bruhat order then $H_{v,w}(q) \preceq H_{v',w}(q)$.*

The nonnegativity claim would actually be immediate if $\mathrm{gr}_{\mathfrak{m}_{e_v}}\mathbb{O}_{e_v,X_w}$ is Cohen–Macaulay (see page 604). However, this latter assertion seems to be a folklore conjecture. Although $\mathbb{O}_{e_v,X_w}$ is itself Cohen–Macaulay [Ramanathan 1985], this property might be lost when degenerating to $\mathrm{gr}_{\mathfrak{m}_{e_v}}\mathbb{O}_{e_v,X_w}$. On the other hand, the results detailed in this paper and in [Li and Yong 2011] also support the Cohen–Macaulayness conjecture. In particular, the latter would follow from that paper's Conjecture 8.5, a stronger claim asserting that Stanley–Reisner simplicial complexes of certain Gröbner degenerations of Kazhdan–Lusztig varieties are vertex decomposable.

The semicontinuity claim is itself a strengthening of the nonnegativity claim since the smoothness of X_w at e_w implies $H_{w,w}(q) = 1$. Furthermore, although the betti numbers of $\mathrm{gr}_{\mathfrak{m}_{e_v}} \mathbb{O}_{e_v, X_w}$ are semicontinuous, the coefficients of $H_{v,w}(q)$ are an involved, signed expression in terms of those numbers. Therefore, this semicontinuity phenomenon seems substantive.

The natural projection map

$$\pi : \mathrm{Flags}(\mathbb{C}^n) \twoheadrightarrow \mathrm{Gr}_k(\mathbb{C}^n) : \quad (\langle 0\rangle \subset F_1 \subset \cdots \subset F_k \subset \cdots \subset F_{n-1} \subset \mathbb{C}^n) \mapsto F_k,$$

where $\mathrm{Gr}_k(\mathbb{C}^n)$ is the Grassmannian of k-dimensional planes in $\mathbb{C}^n$, is a fibration: local properties of torus fixed points $e_\mu \in X_\lambda \subseteq \mathrm{Gr}_k(\mathbb{C}^n)$ for Young diagrams λ, $\mu \subseteq k \times (n-k)$, are equivalent to local properties of $e_v \in X_w \subseteq \mathrm{Flags}(\mathbb{C}^n)$ where v, $w \in S_n$ are maximal Coxeter length representatives of λ, μ where the latter are thought of as cosets of $S_n/(S_k \times S_{n-k})$; see, e.g., [Brion 2004, Example 1.2.3]. These v and w are *cograssmannian*, i.e., they have a unique ascent, at position k: $v(k) < v(k+1)$ and $w(k) < w(k+1)$.

Lifting Grassmannian problems to $\mathrm{Flags}(\mathbb{C}^n)$ has the advantage of allowing one to embed them within the wider class of *covexillary Schubert varieties* X_w, i.e., where w is 3412-avoiding: there are no indices $i_1 < i_2 < i_3 < i_4$ such that $w(i_1)$, $w(i_2)$, $w(i_3)$, $w(i_4)$ are in the same relative order as 3412. This class appears more tractable than general flag Schubert varieties since it shares many of the same features as Grassmannian Schubert varieties. However, there is a salient difference: Grassmannian Schubert varieties are locally defined by equations that are homogeneous with respect to the standard grading that assigns each variable degree one. In general, this is not true in the covexillary case. This homogeneity means that taking associated graded of the local ring essentially does nothing, and so $\mathrm{gr}_{\mathfrak{m}_{e_v}} \mathbb{O}_{e_v, X_w}$ is automatically Cohen–Macaulay; see, e.g., [Li and Yong 2011, Section 1] and page 604.

The covexillary condition has already attracted significant attention; see, e.g., [Lakshmibai and Sandhya 1990; Lascoux 1995; Manivel 2001; Knutson and Miller 2005; Knutson et al. 2008; Knutson et al. 2009; Li and Yong 2011]. In particular, Section 2.4 of [Knutson and Miller 2005] connects the condition to *ladder determinantal ideals* studied in commutative algebra. Our three main theorems below concern the covexillary setting, providing our main cases of support towards both our main thesis and Conjecture 1.1.

One of our results is to prove the following link between $H_{v,w}(q)$ and $P_{v,w}(q)$:

Theorem 1.2. *For w covexillary,*

$$P_{v,w}(q) \preceq H_{v,w}(q) \quad \textit{and} \quad \deg P_{v,w}(q) = \deg H_{v,w}(q).$$

While the Grassmannian case *per se* is new and supports our thesis, the covexillary generality also further highlights the amenability of covexillary Schubert varieties. Our proof of Theorem 1.2 is based on a new formula for covexillary Kazhdan–Lusztig polynomials. An earlier rule was given by A. Lascoux [1995], generalizing his earlier Grassmannian rule with M.-P. Schützenberger [Lascoux and Schützenberger 1981]. (For more recent treatments of the Grassmannian case see [Shigechi and Zinn-Justin 2010; Jones and Woo 2010], for example.) Our formulation of a covexillary rule is in terms of *drift configurations*. It is entirely graphical and is perhaps more handy to compute.

To state our rule we use standard combinatorics of the symmetric group (see, e.g., [Manivel 2001, Chapter 2]) as well as some terminology introduced in [Li and Yong 2011]. (The reader may wish to compare Examples 1.5 and 1.6 below with what follows.) Let $w \in S_n$ be covexillary. Superimpose the *graph* $G(v)$ of v drawn with dots $\circ$ in positions $(n - v(j) + 1, j)$ on top of the *diagram*

$$D(w) = \big\{(i, j) : i < n - w(j) + 1 \text{ and } j < w^{-1}(n - i + 1)\big\} \subset [n] \times [n].$$

Throughout, we use the convention that rows are indexed from bottom to top, and columns are indexed from left to right. Move each box $\mathfrak{e}$ of the *essential set*

$$\mathscr{E}(w) = \big\{(i, j) \in D(w) : (i + 1, j), (i, j + 1) \notin D(w)\big\}$$

diagonally southwest by the number of dots of $G(v)$ weakly southwest of $\mathfrak{e}$. Call the resulting boxes $\{\mathfrak{e}'\}$, and define $B(v, w)$ to be the smallest Young diagram that contains $\{\mathfrak{e}'\}$ and $(1, 1)$ (we use French convention for our Young diagrams). The *shape* $\lambda(w)$ of w is obtained by sorting the vector counting the number of boxes in nonempty rows of $D(w)$ into decreasing order. Now, draw $\lambda(w)$ in the southwest corner of $B(v, w)$.

Declare that any corner of $\lambda(w)$ is 0-*special*. Let $\mathrm{arm}(b)$ (respectively, $\mathrm{leg}(b)$) refer to the boxes in $\lambda(w)$ strictly to the right (above) of b and in the same row (column). Inductively, a box $b \in \lambda(w)$ is z-*special*, for $z \in \mathbb{N}$ if it is maximally northeast subject to

- $|\mathrm{leg}(b)| = |\mathrm{arm}(b)|$; and
- none of the boxes of $\{b\} \cup \mathrm{arm}(b) \cup \mathrm{leg}(b)$ are y-special for any $y < z$.

A box is *special* if it is z-special for some z. The *continent* of a special box b is the set of $x \in \lambda(w)$ such that b is the maximally northeast special box that is weakly southwest of x. The union of continents is

$$\mathrm{Pangaea}(v, w) \subseteq \lambda(w)$$

(the set difference being an immovable reference continent).

Definition 1.3. A *drift configuration* $\mathfrak{D}$ is a nonoverlapping configuration of continents inside $B(v, w)$, such that

- each special box is diagonally weakly northeast of its position in Pangaea(v, w), and
- relative southwest-northeast positions of special cells are maintained.

Let drift(v, w) be the set of all such $\mathfrak{D}$ and let wt$(\mathfrak{D})$ be the total distance traveled by the continents from Pangaea(v, w). Consider the generating series

$$Q_{v,w}(q) = \sum_{\mathfrak{D}\in \mathrm{drift}(v,w)} q^{\mathrm{wt}(\mathfrak{D})}.$$

Theorem 1.4. *Suppose that $v, w \in S_n$ and w is covexillary. Then*:

(I) $P_{v,w}(q) = Q_{v,w}(q)$.

(II) *If we instead take every box of $\lambda(w)$ to be a separate "country", each of which "drifts" according to the rules of Definition 1.3, the total number of drift configurations is* $\mathrm{mult}_{e_v}(X_w)$; *hence*

$$P_{v,w}(1) \leq \mathrm{mult}_{e_v}(X_w),$$

as is manifest from (I).

(III) *There is a vertex decomposable* (*thus shellable*) *simplicial complex* $\mathrm{KL}_{v,w}$ *that is homeomorphic to a ball or a sphere, and whose facets are labeled by* $\mathfrak{D} \in \mathrm{drift}(v, w)$.

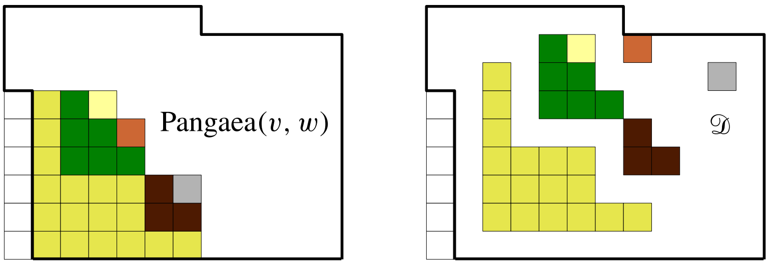
Our proof of (I) is a bijection with A. Lascoux's rule (which descends to a bijection with the rule of [Lascoux and Schützenberger 1981] for Grassmannians). The multiplicity rule from (II) just restates the theorem from [Li and Yong 2011] (compare the Grassmannian rule of [Ikeda and Naruse 2009]). Although the inequality of (II) is a consequence of Theorem 1.2, we are emphasizing that our rule from (I) is compatible with our multiplicity rule and makes the inequality transparent. Actually, whether such an inequality might exist was first asked to us (independently) by S. Billey and A. Woo. Afterwards, H. Naruse informed us that he has a proof for all cominuscule G/P. These questions and results provided us initial motivation for our work towards Theorem 1.4. Note that as with the more general inequality of Theorem 1.2, this inequality is not true in general. For example, $P_{13425,34512}(1) = 3$ while $\mathrm{mult}_{e_{13425}}(X_{34512}) = 2$.

Statement (III) is derived from [Knutson et al. 2008]. It points out a further resemblance to the combinatorics of $\mathrm{mult}_{e_v}(X_w)$ in [Li and Yong 2011], where a similar complex also appears.

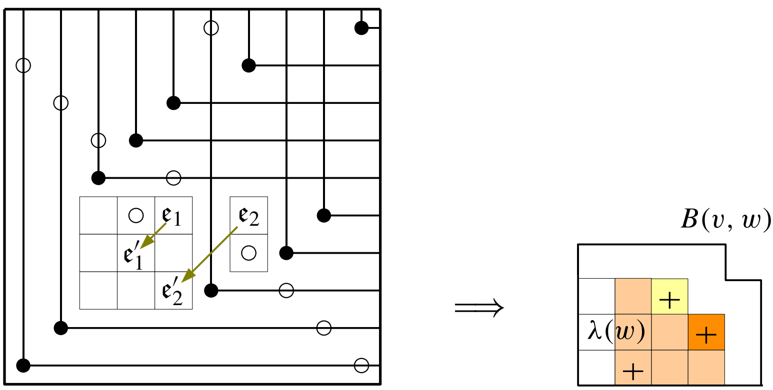
Example 1.5. The left diagram depicts Pangaea(v, w), where $v = \mathrm{id}$ and

$$w = \underline{20}\ \underline{19}\ \underline{18}\ \underline{11}\ \underline{10}\ 9\ 8\ \underline{12}\ \underline{17}\ \underline{16}\ 7\ 6\ \underline{15}\ \underline{14}\ \underline{13}\ 5\ 4\ 3\ 2\ 1.$$

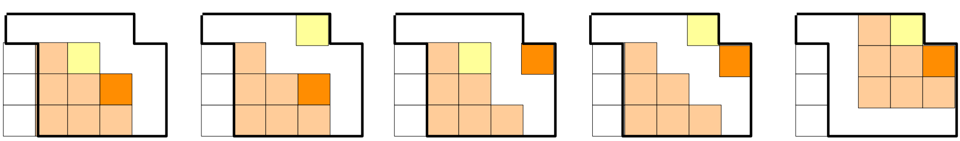
It has six continents, shown in different colors. The right diagram shows a particular drift $\mathfrak{D} \in \mathrm{drift}(v, w)$; its weight is 14.



Example 1.6. Let $w = \underline{10}\,954382761$, $v = 234651789\,\underline{10}$. Here $\lambda(w) = (4, 4, 3)$. The left figure shows $D(w)$, with $G(w)$ overlaid as black dots and $G(v)$ as open circles.



Starting from $D(w)$ and the overlaid $\circ$'s of $G(v)$, we derive $B(v, w)$, shown on the right. The special boxes are marked by $+$'s. We have $\mathscr{E}(w) = \{\mathfrak{e}_1, \mathfrak{e}_2\}$ (being the maximally northeast boxes of each connected component of $D(w)$) move to $\{\mathfrak{e}'_1, \mathfrak{e}'_2\}$, as determined by the $\circ$'s of $G(v)$. These are the five drift configurations:



We can write $Q_{v,w}(q) = 1 + 2q + q^2 + q^3$. $\square$

Our proof of Theorem 1.2 also depends on a new (and the first manifestly positive) combinatorial rule for covexillary $H_{v,w}(q)$. It additionally implies special cases of the nonnegativity and upper semicontinuity conjectures. Identify a partition $\lambda = (\lambda_1 \geq \cdots \geq \lambda_\ell > 0)$ with its Young diagram (in French notation). Recall that a Young tableau T of shape λ is *semistandard* if it is weakly increasing along rows and strictly increasing up columns. Given a vector $\boldsymbol{b} = (b_1, \ldots, b_\ell)$, we say T is *flagged* by $\boldsymbol{b}$ if each entry in row i is at most b_i. Let $\mathrm{SSYT}(\lambda, \boldsymbol{b})$ denote the set of semistandard Young tableaux flagged by $\boldsymbol{b}$. A (nonempty) set-valued filling is semistandard if each tableau obtained by choosing a singleton from each set is semistandard [Buch 2002]. Similarly, we define *flagged set-valued semistandard tableaux*, and the set $\mathrm{SetSSYT}(\lambda, \boldsymbol{b})$ [Knutson et al. 2008].

Define $U \in \mathrm{SetSSYT}(\lambda, \boldsymbol{b})$ to be *lower saturated* if no smaller number can be added to any box $U(i, j)$ while maintaining semistandardness. In symbols, each $U(i, j)$ is of the form

$$[\alpha, \beta] := \{\alpha, \alpha+1, \ldots, \beta-1, \beta\},$$

for some α, β (depending on i, j), where

$$\alpha = \max\bigl\{\max U(i, j-1), 1+\max U(i-1, j)\bigr\}.$$

Our convention for lower saturated tableaux is that $U(i, 0) = 1$ for all $i > 0$ and $U(0, j) = 0$ for all $j > 0$. Let

$$\mathrm{Lower}(\lambda, \boldsymbol{b}) \subseteq \mathrm{SetSSYT}(\lambda, \boldsymbol{b})$$

denote this subset of lower saturated tableaux.

Define the *saturation* $\mathrm{sat}(T) \in \mathrm{Lower}(\lambda, \boldsymbol{b})$ of $T \in \mathrm{SSYT}(\lambda, \boldsymbol{b})$ to be

$$\mathrm{sat}(T)(i, j) = \bigl[\max\{T(i, j-1), 1+T(i-1, j)\}, T(i, j)\bigr].$$

For $U \in \mathrm{SetSSYT}(\lambda, \boldsymbol{b})$, let

$$\mathrm{ex}(U) = |U| - |\lambda|,$$

where $|U|$ refers to the number of entries of U and $|\lambda| = \lambda_1 + \lambda_2 + \cdots$.

Finally, if $T \in \mathrm{SSYT}(\lambda, \boldsymbol{b})$ set

$$\mathrm{depth}(T) := \mathrm{ex}(\mathrm{sat}(T)) = |\mathrm{sat}(T)| - |T|. \tag{1-1}$$

If $\lambda(w) = (\lambda(w)_1 \geq \cdots \geq \lambda(w)_\ell > 0)$, define

$$\boldsymbol{b} = \boldsymbol{b}(\Theta_{v,w}) = (b_1, \ldots, b_\ell) \tag{1-2a}$$

by

$$b_i = \max\bigl\{m : B(v, w)_m \geq \lambda(w)_i + m - i\bigr\}. \tag{1-2b}$$

This is the maximum distance that the rightmost box in row i can drift diagonally northeast within $B(v, w)$ (ignoring the presence of other boxes).

Theorem 1.7. *Let $w \in S_n$ be covexillary. Then*

$$H_{v,w}(q) = \sum_{T \in \mathrm{SSYT}(\lambda(w), \boldsymbol{b}(\Theta_{v,w}))} q^{\mathrm{depth}(T)} = \sum_{U \in \mathrm{Lower}(\lambda(w), \boldsymbol{b}(\Theta_{v,w}))} q^{\mathrm{ex}(U)}.$$

Moreover, Conjecture 1.1 is true under the hypothesis.

Example 1.8. For $n = 5$, $w = 52341$, $v = 12345$. There are five semistandard tableaux of shape $(2, 1)$ and flagged by $(2, 3)$:

$$\begin{array}{|c|c|}\cline{1-1} 2 & \multicolumn{1}{c}{} \\ \hline 1 & 1 \\ \hline \end{array} \quad \begin{array}{|c|c|}\cline{1-1} 3 & \multicolumn{1}{c}{} \\ \hline 1 & 1 \\ \hline \end{array} \quad \begin{array}{|c|c|}\cline{1-1} 2 & \multicolumn{1}{c}{} \\ \hline 1 & 2 \\ \hline \end{array} \quad \begin{array}{|c|c|}\cline{1-1} 3 & \multicolumn{1}{c}{} \\ \hline 1 & 2 \\ \hline \end{array} \quad \begin{array}{|c|c|}\cline{1-1} 3 & \multicolumn{1}{c}{} \\ \hline 2 & 2 \\ \hline \end{array}$$

Their saturations are

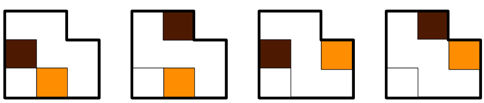
$$\begin{array}{|c|c|}\cline{1-1} 2 & \multicolumn{1}{c}{} \\ \hline 1 & 1 \\ \hline \end{array} \quad \begin{array}{|c|c|}\cline{1-1} 2,3 & \multicolumn{1}{c}{} \\ \hline 1 & 1 \\ \hline \end{array} \quad \begin{array}{|c|c|}\cline{1-1} 2 & \multicolumn{1}{c}{} \\ \hline 1 & 1,2 \\ \hline \end{array} \quad \begin{array}{|c|c|}\cline{1-1} 2,3 & \multicolumn{1}{c}{} \\ \hline 1 & 1,2 \\ \hline \end{array} \quad \begin{array}{|c|c|}\cline{1-1} 3 & \multicolumn{1}{c}{} \\ \hline 1,2 & 2 \\ \hline \end{array}$$

The corresponding ex values are

$$0, 1, 1, 2, 1.$$

Thus by Theorem 1.7, $H_{v,w}(q) = 1 + 3q + q^2$. □

Example 1.9. Continuing Example 1.8, there are four drift configurations of the two continents:



The Kazhdan–Lusztig polynomial is $P_{v,w}(q) = 1 + 2q + q^2$. We see that $P_{v,w}(q) \preceq H_{v,w}(q)$, in agreement with Theorem 1.2. □

Organization and contents. In Section 2, we state some preliminaries and further discuss Conjecture 1.1. We then prove Theorem 1.7. In Section 3, we briefly recall, for comparison, basics about Kazhdan–Lusztig theory. We then prove Theorem 1.2 while temporarily assuming Theorem 1.4(I). Section 4 is devoted to the construction of the simplicial complex of Theorem 1.4(II) and proof of its asserted properties. We furthermore define polynomials generalizing $Q_{v,w}(q)$ that naturally arise from this complex. In Section 5 we prove Theorem 1.4(I). We end that section with two comments (Remarks 5.5 and 5.6) about further properties of $P_{v,w}(q)$ that can be deduced from the rule. In Section 6, we give a formula for a different "q-analogue" of $\mathrm{mult}_{e_v}(X_w)$ than $H_{v,w}(q)$. In Section 7, we offer some final remarks.

2. Hilbert series of the local ring $\mathcal{O}_{e_v, X_w}$

2.1. ***Preliminaries.*** We use the usual identification $\mathrm{Flags}(\mathbb{C}^n) = GL_n/B$ where B is the Borel subgroup consisting of invertible upper triangular matrices. Thus GL_n acts on $\mathrm{Flags}(\mathbb{C}^n)$ by left multiplication, as does B, and the torus T of invertible diagonal matrices. For each $v \in S_n$, let e_v denote the associated T-fixed point. The *Schubert cell* is $X_w^\circ := Be_w$, while its Zariski closure is the *Schubert variety* $X_w = \overline{X_w^\circ}$, an irreducible variety of dimension $\ell(w)$. We have that $e_v \in X_w$ if and only if $v \leq w$ in *Bruhat order*. A neighborhood of each point $p \in X_w$ is isomorphic to a neighborhood of some e_v, by the action of B. Hence, it suffices to restrict attention to T-fixed points. Let B_- be the opposite Borel subgroup of invertible lower triangular matrices. If we set $\Omega_v^\circ = B_- vB/B$ to be the *opposite Schubert cell*, then up to crossing by affine space, a local neighborhood of $e_v \in X_w$ is given by the *Kazhdan–Lusztig variety* $\mathcal{N}_{v,w} = X_w \cap \Omega_v^\circ$ [Kazhdan and Lusztig 1979, Lemma A.4].

Suppose p is a point on a scheme Y. Let $\mathrm{gr}_{\mathfrak{m}_p}\mathcal{O}_{p,Y}$ denote the *associated graded ring* of the local ring $\mathcal{O}_{p,Y}$ with respect to its maximal ideal $\mathfrak{m}_p$, i.e.,

$$\mathrm{gr}_{\mathfrak{m}_p}\mathcal{O}_{p,Y} = \bigoplus_{i \geq 0} m_p^i / m_p^{i+1}.$$

Since $\mathrm{gr}_{\mathfrak{m}_p}\mathcal{O}_{p,Y}$ picks up a $\mathbb{Z}$-grading, it now makes sense to discuss its Hilbert series. One can always express this series in the form

$$\mathrm{Hilb}(\mathrm{gr}_{\mathfrak{m}_p}\mathcal{O}_{p,Y}, q) = \frac{H_{p,Y}(q)}{(1-q)^{\dim Y}}$$

where $H_{p,Y}(q) \in \mathbb{Z}[q]$ is the *h-polynomial* associated to $p \in Y$. It follows from standard facts that $H_{p,Y}(1) = \mathrm{mult}_p(Y)$; see, e.g., [Kreuzer and Robbiano 2005, Theorem 5.4.15]. Hence $H_{p,Y}(q) = 1$ if and only if Y is smooth at p. In addition, note $H_{p,Y}(0) = 1$, since this is the dimension of the zero graded piece of $\mathrm{gr}_{\mathfrak{m}_p}\mathcal{O}_{p,Y}$, i.e., the dimension of the field $\mathcal{O}_{p,Y}/\mathfrak{m}_p$.

Now, for any $v, w \in S_n$, we define $H_{v,w}(q) \in \mathbb{Z}[q]$ to be the h-polynomial associated to $e_v \in X_w$. At present, there is no purely combinatorial formula (even nonpositive or recursive) for computing $H_{v,w}(q)$. However, instead one can utilize the explicit coordinates and equations for the ideal $I_{v,w}$ to define $\mathcal{N}_{v,w} = \mathrm{Spec}\left(\mathbb{C}[z^{(v)}]/I_{v,w}\right)$, as done in [Woo and Yong 2008, Section 3.2]. Then one can Gröbner degenerate $\mathcal{N}_{v,w}$ to a scheme theoretic union of coordinate subspaces $\mathcal{N}'_{v,w}$, using any of the term orders $\prec_{v,w,\pi}$ from [Li and Yong 2011, Section 3]. As explained in Theorem 3.1 (and its proof) of that reference, the stated Gröbner degenerations degenerate not only $\mathcal{N}_{v,w}$ but also its projectivized tangent cone $\mathrm{Proj}(\mathrm{gr}_{\mathfrak{m}_{e_v}}\mathcal{O}_{e_v,X_w})$. Therefore the h-polynomial of $\mathcal{N}'_{v,w}$ equals $H_{v,w}(q)$.

2.2. ***Conjectures.*** Let us now return to the discussion of Conjecture 1.1. Using the method for computing $H_{v,w}(q)$ summarized above, we obtained exhaustive checks for $n \leq 7$ of the following claim, restated from the introduction:

Nonnegativity conjecture. $H_{v,w}(q) \in \mathbb{Z}_{\geq 0}[q]$.

In [Li and Yong 2011, Conjecture 8.5] we conjectured that within the family of term orders $\prec_{v,w,\pi}$, at least one gives a Gröbner limit scheme $\mathcal{N}'_{v,w}$ that is reduced, equidimensional and whose Stanley–Reisner simplicial complex $\Delta_{v,w}$ is a vertex-decomposable ball or sphere. This implies in particular that $\Delta_{v,w}$ is shellable and thus Cohen–Macaulay. If this conjecture were true, it would follow that $\mathrm{gr}_{\mathfrak{m}_{e_v}} \mathbb{O}_{e_v,X_w}$ is Cohen–Macaulay. Thus the nonnegativity conjecture would hold by, e.g., [Bruns and Herzog 1993, Corollary 4.1.10].

In the case that $I_{v,w}$ is a homogeneous ideal, with respect to the standard grading that assigns each variable degree 1, since $\mathbb{O}_{e_v,X_w}$ is Cohen–Macaulay [Ramanathan 1985], it follows that the associated graded ring is Cohen–Macaulay; see [Bruns and Herzog 1993, Exercise 2.1.27(c)], for example. Hence nonnegativity follows in this case. A. Knutson [2009, p. 25] has shown that this homogeneity occurs whenever w is 321-avoiding. Moreover, in [Woo and Yong 2009, Section 5] it was explained how "parabolic moving" reduces a large percentage of cases (for $n \leq 10$) to the homogeneous case. However, not every case can be so reduced, including those in the covexillary class. Thus, these cases provide further support for the nonnegativity conjecture, separate from Theorem 1.7.

Upper semicontinuity conjecture. *If $v' \leq v \leq w$ in Bruhat order, then*

$$H_{v,w}(q) \preceq H_{v',w}(q).$$

Unfortunately, even if we knew $\mathrm{gr}_{\mathfrak{m}_{e_v}} \mathbb{O}_{e_v,X_w}$ to be Cohen–Macaulay, we do not know any way to express these coefficients in homological terms that would make the upper semicontinuity conjecture transparent. It should be noted that the proof of this property for Kazhdan–Lusztig polynomials in [Irving 1988] was not achieved using the geometry of Schubert varieties. However, see the geometric argument for the more general result [Braden and MacPherson 2001, Theorem 3.6].

Although any proof of the above conjectures is desired, ideally one would also like combinatorial explanations of the properties.

Let us pause to collect some further facts for small n in the following computational result. For (D) below we refer the reader to [Woo and Yong 2008, Section 2.1] for the definition of *interval pattern avoidance* of $[x, y] \in S_\infty \times S_\infty$. There we explain that the existence of an *interval pattern embedding* guarantees $\mathcal{N}_{x,y} \cong \mathcal{N}_{\tilde{w},w}$, where $[x, y] \cong [\tilde{w}, w]$ is an isomorphism of posets of Bruhat intervals in S_∞. Thus, if the inequality $P_{x,y}(q) \preceq H_{x,y}(q)$ fails, so must $P_{\tilde{w},w}(q) \preceq H_{\tilde{w},w}(q)$.

Proposition 2.1. (A) $\deg H_{v,w}(q) \leq \deg P_{v,w}(q)$ *for* $v \leq w \in S_n$ *and* $n \leq 6$.

(B) $\deg H_{v,w}(q) \leq \frac{1}{2}(\ell(w)-\ell(v)-1)$ *for* $v < w \in S_n$ *and* $n \leq 7$.

(C) *The coefficients of* $H_{v,w}(q)$ *form a unimodal sequence for* $v, w \in S_n$ *and* $n \leq 7$.

(D) $P_{v,w}(q) \preceq H_{v,w}(q)$ *holds for all* $v \leq w \in S_n$ *and* $n \leq 6$, *if and only if* w *interval pattern avoids*

$$[14235, 45123], \quad [31524, 53412], \quad [14325, 45312],$$
$$[13425, 34512], \quad [24153, 45231], \quad [154326, 564312].$$

(Note that the first and fourth intervals, and the second and fifth intervals are related by taking inverses. For all $n \geq 1$, the inequality fails whenever w contains one of these intervals.)

Proof and discussion. Each of the assertions were verified using Macaulay 2. For (A) and (B) note that $\deg P_{v,w}(q) \leq \frac{1}{2}(\ell(w)-\ell(v)-1)$ is a standard fact about Kazhdan–Lusztig polynomials; see item (iii) on page 608.

For (D), computation shows that $P_{v,w}(q) = H_{v,w}(q)$ for $n \leq 4$, so the inequality holds in that situation. We checked that each of the intervals $[x, y]$ listed corresponds to a failure of the inequality for $n \leq 5$. For $n = 6$ we computationally verified the claim (there are 36 cases $w \in S_6$ where the inequality fails for some $v \leq w$, and of those only one cannot be blamed on the $n = 5$ cases). The $n > 6$ case follows from general properties of interval pattern embeddings recalled above. □

One might conjecture that both (A) and its weak form (B) hold for all n. However with (A), experience has shown that data for $n \leq 6$ is soft evidence for any conjecture that involves Kazhdan–Lusztig polynomials. Note that if (A) is true, one cannot have $P_{v,w}(q) \preceq H_{v,w}(q)$ unless $\deg H_{v,w}(q) = \deg P_{v,w}(q)$, which is indeed what we show when w is covexillary.

In view of (C), it is also natural to guess that unimodality is true in general. One warning however is that the stronger assertion that the coefficients of $H_{v,w}(q)$ are *log-concave* is false, as the example below shows:

Example 2.2. Let $w = 5671234$ and $v = 1352476$. Computation using Macaulay 2 shows there is a choice of $\prec_{v,w,\pi}$ such that $\mathcal{N}'_{v,w}$ is Cohen–Macaulay (but not Gorenstein), and that $H_{1352476,5671234}(q) = 1 + 2q + q^2 + q^3$, which is not log-concave. □

By contrast, see the related work [Rubey 2005], which shows log-concavity holds in a special ladder determinantal case (note that w is not covexillary in our counterexample).

Even knowing Cohen–Macaulayness of $\mathrm{gr}_{\mathfrak{m}_{e_v}} \mathcal{O}_{e_v, X_w}$ does not, in and of itself, prove unimodality. In fact, R. Stanley [1989, Conjecture 4(a)] had conjectured unimodality for a general graded Cohen–Macaulay domain R over a field which is generated by R_1. Actually, he even conjectured the stronger claim of log-concavity, although counterexamples to the stronger claim were later found by G. Niesi and L. Robbiano; see [Brenti 1994, Section 5]. (Example 2.2 gives a different counterexample to Stanley's log-concavity conjecture.)

It should also be mentioned that in contrast, the Kazhdan–Lusztig polynomials are not in general unimodal and in fact P. Polo [1999] proved that every nonnegative integral polynomial with constant coefficient 1 is some $P_{v,w}(q)$.

While Theorem 1.7 allows us to prove the nonnegativity, upper semicontinuity and degree properties for covexillary X_w, a solution to the following problem has eluded us:

Problem 2.3. Give a combinatorial proof (e.g., using Theorem 1.7) for the unimodality conjecture, when w is covexillary (or even cograssmannian) by establishing a sequence of explicit injections and surjections of the relevant Young tableaux.

Concerning (D), we do not expect the characterization to be valid for all n. Instead, one aims to expand this list into a (human-readable) classification, via a finite list of families of patterns to avoid, as is the case for many other properties studied in [Woo and Yong 2008].

Using the analogy with Kazhdan–Lusztig theory, numerous further problems, which had been previously considered for $P_{v,w}(q)$ but not $H_{v,w}(q)$, make sense. To name a few: Is $H_{v,w}(q)$ determined by the poset isomorphism class of the interval $[v, w]$ in Bruhat order? (This is an analogue of a conjecture of G. Lusztig.) Can one give a combinatorial algorithm for computing $H_{v,w}(q)$? Better yet, can one find a positive combinatorial rule for $H_{v,w}(q)$, thus establishing the nonnegativity conjecture?

2.3. *Proof of Theorem 1.7*. Continuing the definitions before the statement of Theorem 1.7, set

$$\sup : \mathrm{SetSSYT}(\lambda, \boldsymbol{b}) \to \mathrm{SSYT}(\lambda, \boldsymbol{b})$$

by sending U to T where $T(i, j) = \max U(i, j)$. The following is clear:

Lemma 2.4. *The maps*

$$\mathrm{sat} : \mathrm{SSYT}(\lambda, \boldsymbol{b}) \to \mathrm{Lower}(\lambda, \boldsymbol{b}) \quad \textit{and} \quad \sup|_{\mathrm{Lower}(\lambda,\boldsymbol{b})} : \mathrm{Lower}(\lambda, \boldsymbol{b}) \to \mathrm{SSYT}(\lambda, \boldsymbol{b})$$

are mutually inverse bijections.

Let us recall some definitions and terminology utilized in [Li and Yong 2011]. Define $r^w_b = r^w_{(i,j)}$ to be the number of $\bullet$ of $G(w)$ weakly southwest of the box $b = (i, j)$. Given $v \le w$ and w covexillary, $\Theta_{v,w} \in S_n$ is defined there to be the unique permutation such that $\lambda(w) = \lambda(\Theta_{v,w})$ and

$$\mathscr{E}(\Theta_{v,w}) = \begin{Bmatrix} \mathfrak{e}' \,:\, \mathfrak{e}' \text{ is obtained by moving each } \mathfrak{e} \text{ in} \\ \mathscr{E}(w) \text{ diagonally southwest by } r^v_{\mathfrak{e}} \text{ units.} \end{Bmatrix}$$

The permutation $\Theta_{v,w}$ was proved to be itself covexillary.

Define $B(w)$ to be the smallest Young diagram with southwest corner in position $(1, 1)$ that contains all of $\mathscr{E}(w)$. Set

$$B(v, w) = B(\Theta_{v,w}).$$

If $\lambda(w) = (\lambda(w)_1 \geq \cdots \geq \lambda(w)_\ell > 0)$, define $\boldsymbol{b} = \boldsymbol{b}(w) = (b_1, \ldots, b_\ell)$ by

$$b_i = \max\{m : \ B(w)_m \geq \lambda(w)_i + m - i\}.$$

This agrees with, and slightly reformulates, the definitions of $B(v, w)$ and $\boldsymbol{b}$ from the introduction.

In [Li and Yong 2011, Theorem 6.6] we proved that

$$\mathrm{Hilb}(\mathrm{gr}_{\mathfrak{m}_{e_v}} \mathbb{O}_{e_v,X_w}, q) = G_{\lambda(w)}(q)/(1-q)^{\binom{n}{2}},$$

where

$$G_{\lambda(w)}(q) = \sum_{k \geq |\lambda(w)|} (-1)^{k-|\lambda(w)|}(1-q)^k \times \#\mathrm{SetSSYT}(\lambda(w), \boldsymbol{b}, k)$$

and $\#\mathrm{SetSSYT}(\lambda(w), \boldsymbol{b}, k)$ is the number of flagged set-valued semistandard Young tableaux of shape $\lambda(w)$ with flag $\boldsymbol{b} = \boldsymbol{b}(\Theta_{v,w})$ which use exactly k entries.

Since the local ring $\mathbb{O}_{e_v,X_w}$ is of dimension $\ell(w) = \binom{n}{2} - |\lambda(w)|$, we rewrite

$$\mathrm{Hilb}(\mathrm{gr}_{\mathfrak{m}_{e_v}} \mathbb{O}_{e_v,X_w}, q) = \frac{H_{v,w}(q)}{(1-q)^{\ell(w)}},$$

where

$$H_{v,w}(q) = \sum_{U \in \mathrm{SetSSYT}(\lambda(w),\boldsymbol{b})} (q-1)^{\mathrm{ex}(U)}.$$

We need to show that

$$\sum_{U \in \mathrm{SetSSYT}(\lambda(w),\boldsymbol{b})} (q-1)^{\mathrm{ex}(U)} = \sum_{T \in \mathrm{SSYT}(\lambda(w),\boldsymbol{b})} q^{\mathrm{depth}(T)} \tag{2-1}$$

by proving that, for every $T \in \mathrm{SSYT}(\lambda(w), \boldsymbol{b})$,

$$\sum_{U \in \mathrm{sup}^{-1}(T)} (q-1)^{\mathrm{ex}(U)} = q^{\mathrm{depth}(T)}.$$

There are $\mathrm{depth}(T)$ elements in $\mathrm{sat}(T)$ but not in T. We can delete any subset of those elements from $\mathrm{sat}(T)$ and obtain $T' \in \mathrm{sup}^{-1}(T)$ (so $\#\,\mathrm{sup}^{-1}(T) = 2^{\mathrm{depth}(T)}$). Hence the left-hand side is equal to

$$(1 + (q-1))^{\mathrm{depth}(T)} = q^{\mathrm{depth}(T)},$$

and therefore the equality (2-1) follows. Thus, the first equality of the theorem holds and the second is clear from Lemma 2.4.

The nonnegativity claim is manifest from the combinatorial rule; however, let us also give a geometric proof. In [Li and Yong 2011] we proved that for covexillary w, $\mathcal{N}_{v,w}$ degenerates, under a choice of $\prec_{v,w,\pi}$ to a Cohen–Macaulay limit scheme $\mathcal{N}'_{v,w}$. Hence, nonnegativity of $H_{v,w}(q)$ follows from [Bruns and Herzog 1993, Corollary 4.1.10] and the discussion on page 603.

For the upper semicontinuity claim, fix $w \in S_n$ and suppose $v' \leq v \leq w$. Consider an essential box $\mathfrak{e} \in \mathscr{E}(w)$. In the construction of $\mathscr{E}(\Theta_{v,w})$, the essential box $\mathfrak{e}$ is moved diagonally southwest by $r^v_{\mathfrak{e}}$ units. Since $v' \leq v$, a standard characterization of Bruhat order shows $r^{v'}_{\mathfrak{e}} \leq r^v_{\mathfrak{e}}$. Thus, each essential box $\mathfrak{e}$ moves further southwest in to its position in $\mathscr{E}(\Theta_{v,w})$ than it does for $\mathscr{E}(\Theta_{v',w})$. Therefore,

$$B(v, w) \subseteq B(v', w),$$

and hence,

$$\boldsymbol{b}(\Theta_{v,w}) = (b_1, \ldots, b_\ell) \leq \boldsymbol{b}(\Theta_{v',w}) = (b'_1, \ldots, b'_\ell),$$

in the sense that $b_i \leq b'_i$ for every i. Consequently, $\mathrm{SSYT}(\lambda, \boldsymbol{b}) \subseteq \mathrm{SSYT}(\lambda, \boldsymbol{b}')$, which clearly implies $H_{v,w}(q) \preceq H_{v',w}(q)$, as desired. □

3. Kazhdan–Lusztig theory

The Hecke algebra. Let $R = \mathbb{Z}[q^{\frac{1}{2}}, q^{-\frac{1}{2}}]$ be the ring of Laurent polynomials over $\mathbb{Z}$ in the indeterminate $q^{\frac{1}{2}}$. The *Hecke algebra* $\mathscr{H}_{n-1}$ of S_n is the algebra over R with basis $\{T_w : w \in S_n\}$ and relations

$$T_{s_i} T_w = T_{s_i w} \quad \text{if } \ell(s_i w) > \ell(w),$$
$$T_{s_i}^2 = (q-1) T_{s_i} + q T_{\mathrm{id}}.$$

There is an involution $\iota : \mathscr{H}_{n-1} \to \mathscr{H}_{n-1}$ defined by $\iota(q^{\frac{1}{2}}) = q^{-\frac{1}{2}}$ and $\iota(T_w) = T^{-1}_{w^{-1}}$.

It was proved in [Kazhdan and Lusztig 1979] that there exists a basis $\{\mathscr{C}'_w\}$ of $\mathscr{H}_{n-1}$ that is uniquely determined by the conditions

$$\iota(\mathscr{C}'_w) = \mathscr{C}'_w \quad \text{and} \quad C'_w = (q^{-\frac{1}{2}})^{\ell(w)} \sum_{v \leq w} P_{v,w}(q) T_v,$$

where

(i) $P_{w,w}(q) = 1$,

(ii) $P_{v,w}(q) = 0$ if $v \not\leq w$, and

(iii) $P_{v,w}(q) \in \mathbb{Z}[q]$ is of degree at most $\frac{1}{2}(\ell(w) - \ell(v) - 1)$ if $v < w$.

The existence of this basis was established by an explicit recursion for the *Kazhdan–Lusztig polynomials* $P_{v,w}(q)$, which we omit. Our source for these facts is [Billey and Lakshmibai 2000, Chapter 6], to which we refer the reader to for further details.

Conditions (i) and (ii) also hold for the $H_{v,w}(q)$, while (iii) conjecturally holds (compare Proposition 2.1 and the discussion thereafter). It is mildly tempting to think about another basis of the Hecke algebra defined by replacing $P_{v,w}(q)$ by $H_{v,w}(q)$ in the above definition of C'_w. While this other basis has a unimodular transition matrix with the Kazhdan–Lusztig basis, it doesn't possess any of the other nice properties, such as positive structure constants or invariance under the involution ι.

Proof of Theorem 1.2. Recall that in what follows, we are assuming the formula for $P_{v,w}(q)$ from Theorem 1.4 that we prove in Section 5.

Given any box $(i, j) \in \lambda(w)$, let $(\hat{i}, j)$ be the topmost box in column j.

Let $\boldsymbol{b} = \boldsymbol{b}(\Theta_{v,w})$ be defined by Equations (1-2) (or see the proof of Theorem 1.7, page 606). Define

$$\Psi : \mathrm{drift}(v, w) \to \mathrm{SSYT}(\lambda(w), \boldsymbol{b})$$

by sending a drift configuration $\mathfrak{D}$ to the semistandard tableau T, as follows. For each special box $(i, j) \in \lambda(w)$ we fill $(\hat{i}, j)$ with the entry $(\hat{i} + d)$, where d is the distance moved in $\mathfrak{D}$ by the continent associated to (i, j), from $\mathrm{Pangaea}(v, w)$. Note that the value of this entry is the height of the box $(\hat{i}, j)$ after drifting in the drift configuration $\mathfrak{D}$. Now fill in the remaining empty boxes of $\lambda(w)$ by working down columns, from right to left, according to the prescription

$$T(i, j) = \min\bigl\{T(i+1, j)-1,\ T(i-1, j+1)+1\bigr\}. \tag{3-1}$$

By convention, set

$$T(i, j) = \begin{cases} \infty & \text{if } i > 0 \text{ and } (i, j) \notin \lambda(w), \text{ or if } j > m, \\ 0 & \text{if } i = 0 \text{ and } j \leq m, \end{cases} \tag{3-2}$$

where m is the number of columns in $\lambda(w)$.

Example 3.1. For the five drift configurations $\mathfrak{D}$ in Example 1.6, the corresponding $\Psi(\mathfrak{D})$ are as follows, where the boxes $(\hat{i}, j)$ corresponding to special boxes are underlined.

3	$\underline{3}$	$\underline{3}$	
2	2	2	$\underline{2}$
1	1	1	1

3	$\underline{3}$	$\underline{4}$	
2	2	2	$\underline{2}$
1	1	1	1

3	$\underline{3}$	$\underline{3}$	
2	2	2	$\underline{3}$
1	1	1	2

3	$\underline{3}$	$\underline{4}$	
2	2	3	$\underline{3}$
1	1	1	2

3	$\underline{4}$	$\underline{4}$	
2	2	3	$\underline{3}$
1	1	1	2

We will also need the $\mathrm{sat}(\Psi(\mathfrak{D}))$, which here are

3	3	3	
2	2	2	2
1	1	1	1

3	3	3,4	
2	2	2	2
1	1	1	1

3	3	3	
2	2	2	3
1	1	1	1,2

3	3	4	
2	2	2,3	3
1	1	1	1,2

3	3,4	4	
2	2	2,3	3
1	1	1	1,2

Lemma 3.2. *Suppose* $\mathfrak{D} \in \mathrm{drift}(v, w)$ *and* $T = \Psi(\mathfrak{D})$. *Then*:

(i) T *is a semistandard Young tableau* (*i.e.*, Ψ *is well-defined*).

(ii) Ψ *is an injection.*

(iii) *If the* j*-th column of* $\lambda(w)$ *has no special box*, *then* $T(i, j) = i$ *for all* $1 \leq i \leq \widehat{i}$.

(iv) $\mathrm{wt}(\mathfrak{D}) = \mathrm{ex}(\mathrm{sat}(T)) = \mathrm{depth}(T)$.

Proof. (i) Since each corner of $\lambda(w)$ is special, it is assigned a finite number. Hence (3-1) assigns each box of $\lambda(w)$ a finite number. The column semistandardness conditions are immediate from (3-1). We now establish the row semistandardness condition $T(i, j) \leq T(i, j+1)$, considering the two cases that can occur.
Case 1: (i, j) *is atop a special box.* That is, there is a special box (i_0, j) with $i = \widehat{i_0}$. Then if $(i, j+1)$ is in $\lambda(w)$, it is atop another special box: Suppose not. Then let the arm and leg length of (i, j) be $\mathscr{L}$. Note that since $\lambda(w)$ is a Young diagram, $(i-\mathscr{L}+1, j+\mathscr{L}+1) \notin \lambda(w)$. Thus there is a smallest integer k such that $1 \leq k \leq \mathscr{L}$ and $(i-k+1, j+k+1) \notin \lambda(w)$. For this k note that $(i-k+1, j+1)$ has equal arm and leg length equal, no other special boxes are above it (by assumption) and no boxes to strictly to its right can be special (their leg lengths are strictly longer than their arm lengths). Hence $(i-k+1, j+1)$ is special, but this is a contradiction.

Now that we know that both (i, j) and $(i, j+1)$ are atop special boxes, hence $T(i, j)$ and $T(i, j+1)$ are the heights of the boxes (i, j) and $(i, j+1)$ in the drift configuration $\mathfrak{D}$. From this interpretation, it is clear that $T(i, j) \leq T(i, j+1)$.
Case 2: (i, j) *is not atop a special box.* In this situation, by (3-1),

$$T(i, j) \leq T(i-1, j+1)+1 \leq T(i, j+1).$$

(ii) This is immediate since different drift configurations will lead to different initial fillings, of the boxes $(\widehat{i}, j)$ where (i, j) is a special box.

(iii) First note that $(\widehat{i}, j+1), (\widehat{i}-1, j+2), (\widehat{i}-2, j+3), \ldots, (1, j+\widehat{i})$ must lie in $\lambda(w)$. Otherwise suppose $k \in \mathbb{Z}_{\geq 0}$ is the smallest integer that $(\widehat{i}-k, j+k+1)$ is not in $\lambda(w)$. Since the j-th column does not contain a special box, $(\widehat{i}, j)$ is not a corner, so $(\widehat{i}, j+1)$ must lie in $\lambda(w)$, and we have $k \geq 1$. Since k is the smallest integer where the failure occurs, $(\widehat{i}-k+1, j+k)$ must lie in $\lambda(w)$, and therefore $(\widehat{i}-k, j+k)$ lies in $\lambda(w)$. The conclusion that $(\widehat{i}-k, j)$ is deduced is a similar manner as in Case 1 of (i).

Now applying (3-1) repeatedly, we have

$$T(\widehat{i}, j) \leq T(\widehat{i}-1, j+1)+1 \leq T(\widehat{i}-2, j+2)+2 \leq \cdots \leq T(1, j+\widehat{i}-1)+\widehat{i}-1,$$

and each of the boxes being considered actually lie in $\lambda(w)$, because of what we just argued. Since $T(1, j+\widehat{i}-1) = 1$ (which holds because $(1, j+\widehat{i}) \in \lambda(w)$ so (3-1) is assigned using the boundary value $T(0, j+\widehat{i}) = 0$), we have $T(\widehat{i}, j) \leq \widehat{i}$, which forces by the fact T is semistandard that $T(i, j) = i$ for $1 \leq i \leq \widehat{i}$.

(iv) The second equality here is just the definition; see (1-1). We establish the first equality. Consider the j-th column of $\lambda(w)$.

Case 1: this column contains a special box (i, j). The column contains $\widehat{i}$ boxes and so each of the numbers $1, 2, \ldots, \widehat{i}+d$ appears exactly once in this column of $\mathrm{sat}(T)$, by the definition of sat and Ψ. Hence the number of extra entries of $\mathrm{sat}(T)$ in column j is equal to $(\widehat{i}+d)-\widehat{i}=d$, which is the same as the distance moved by the continent of (i, j).

Case 2: the column contains no special box. By (iii), there are no extra entries in this column.

Summing up the number of extra entries in each column j of $\mathrm{sat}(T)$, we conclude that $\mathrm{ex}(\mathrm{sat}(T))$ is equal to $\mathrm{wt}(\mathfrak{D})$, as desired. □

Therefore,

$$P_{v,w}(q)=\sum_{\mathfrak{D}\in\mathrm{drift}(v,w)} q^{\mathrm{wt}(\mathfrak{D})}=\sum_{\mathfrak{D}\in\mathrm{drift}(v,w)} q^{\mathrm{depth}(\Psi(\mathfrak{D}))}\preceq\sum_{T\in\mathrm{SSYT}(\lambda(w),\boldsymbol{b})} q^{\mathrm{depth}(T)}=H_{v,w}(q).$$

Here the first equality holds by Theorem 1.4(I), the second equality is by (iv), the "$\preceq$" is by (ii), and the final equality is by Theorem 1.7.

It remains to prove that

$$\deg H_{v,w}(q)=\deg P_{v,w}(q).$$

Since we have already proved that $P_{v,w}(q)\preceq H_{v,w}(q)$ which implies $\deg P_{v,w}(q)\le \deg H_{v,w}(q)$, we need only to prove that $\deg H_{v,w}(q)\le \deg P_{v,w}(q)$. To do so, we will need the following lemma.

Lemma 3.3. *An element $T\in\mathrm{SSYT}(\lambda(w),\boldsymbol{b})$ is in the image of $\Psi:\mathrm{drift}(v,w)\to\mathrm{SSYT}(\lambda(w),\boldsymbol{b})$ if and only if both of the following conditions are true*:

(a) *For any box (i, j) that is not equal to $(\widehat{i'}, j)$ for a special box (i', j),* (3-1) *holds under the convention* (3-2).

(b) *If (i, j) and (i', j') are any two special boxes with (i, j) weakly southwest of (i', j'), then*
$$T(\widehat{i}, j)-\widehat{i}\le T(\widehat{i'}, j')-\widehat{i'}.$$

Proof. Let $\mathfrak{D}\in\mathrm{drift}(v,w)$. We show that $\Psi(\mathfrak{D})$ satisfies (a) and (b). The condition (a) holds by the definition of Ψ. The condition (b) follows since $T(\widehat{i}, j)-\widehat{i}$ equals the distance drifted by the continent containing (i, j), $T(\widehat{i'}, j')-\widehat{i'}$ equals the distance drifted by the continent containing (i', j'), and the continent associated to (i, j) cannot move further northeast than the continent associated to (i', j').

Conversely, we now show that every $T\in\mathrm{SSYT}(\lambda(w),\boldsymbol{b})$ satisfying (a) and (b) is in the image of Ψ. Consider the (putative) drift configuration $\mathfrak{D}$ defined as follows. To each continent of $\mathfrak{D}$ associated to a special box (i, j), shift it northeast by $T(\widehat{i}, j)-\widehat{i}$ units. We first prove that each continent fits inside $B(v, w)$: Consider the continent with special box (i, j). If part of the continent is shifted out of the boundary $B(v, w)$, then by (b) there is some northeast corner of $\lambda(w)$ (i.e., a 1×1 continent) that has been pushed out of $B(v, w)$ by that part of the continent. Hence the corresponding T is not in $\mathrm{SSYT}(\lambda(w), \boldsymbol{b})$, a contradiction.

Now, condition (b) guarantees that $\mathcal{D}$ can in fact be obtained without continents overlapping. Hence $\mathcal{D} \in \mathrm{drift}(v, w)$. Finally, by (a), we have $\Psi(\mathcal{D}) = T$. ☐

Given $T \in \mathrm{SSYT}(\lambda(w), \boldsymbol{b})$, consider this condition:

$$\begin{array}{l}\text{Some box } (i, j) \text{ in } \lambda(w) \text{ is not a northeast}\\ \text{corner and is such that (3-1) does not hold.}\end{array} \tag{3-3}$$

Suppose (3-3) holds for $T = T_0$. Suppose also that (i, j) is chosen such that j is smallest, with ties broken by taking i smallest.

A brief outline of the remainder of the proof follows. Starting from T_0, we construct a sequence $T_1, T_2, \ldots \in \mathrm{SSYT}(\lambda(w), \boldsymbol{b})$ with increasing depth until we arrive at a T_k that fails (3-3). This T_k is proved to be in the image of Ψ. Then we show that $\mathcal{D} := \Psi^{-1}(T_k) \in \mathrm{drift}(v, w)$ satisfies $\mathrm{wt}(\mathcal{D}) \geq \mathrm{depth}(T_0)$. From this the result follows; see (3-8).

So let $T_1 \in \mathrm{SSYT}(\lambda(w), \boldsymbol{b})$ be the augmentation of T_0 obtained by setting

$$T_1(i, j) = \min\{T_0(i+1, j) - 1, T_0(i-1, j+1) + 1\} \tag{3-4}$$

and letting all other entries in T_1 be the same as in T_0.

Now we show that $T_1 \in \mathrm{SSYT}(\lambda(w), \boldsymbol{b})$. To do this, we need to check the semistandardness conditions

$$T_1(i, j-1) \leq T_1(i, j) \leq T_1(i, j+1), \tag{3-5}$$

$$T_1(i-1, j) < T_1(i, j) < T_1(i+1, j). \tag{3-6}$$

We first check (3-5). The second inequality is trivial from (3-4). For the first inequality, we have

$$\begin{aligned}T_0(i, j-1) &\leq T_0(i+1, j-1) - 1 \leq T_0(i+1, j) - 1,\\ T_0(i, j-1) &\leq T_0(i-1, j) + 1 \leq T_0(i-1, j+1) + 1.\end{aligned}$$

(The second of those lines uses the minimality of our choice of (i, j).) Hence

$$T_1(i, j-1) = T_0(i, j-1) \leq \min\bigl\{T_0(i+1, j) - 1, T_0(i-1, j+1) + 1\bigr\} = T_1(i, j).$$

Similarly for (3-6): the second inequality is trivial from (3-4), whereas for the first inequality, we have

$$\begin{aligned}T_0(i-1, j) &< T_0(i, j) \leq T_0(i+1, j) - 1,\\ T_0(i-1, j) &\leq T_0(i-1, j+1) < T_0(i-1, j+1) + 1,\end{aligned}$$

and hence

$$T_1(i-1,j) = T_0(i-1,j) < \min\bigl\{T_0(i+1,j)-1,\, T_0(i-1,j+1)+1\bigr\} = T_1(i,j).$$

Next, we claim that

$$\operatorname{depth}(T_1) \geq \operatorname{depth}(T_0).$$

The difference in depth between T_1 and T_0 can only be blamed on the boxes in positions (i,j), $(i,j+1)$ and $(i+1,j)$. Without loss of generality, let us assume that each of the latter two boxes actually lie in $\lambda(w)$ (at least one of $(i,j+1)$ or $(i+1,j)$ is in $\lambda(w)$ since (i,j) is assumed to not be a northeast corner; analyzing the resulting cases is similar and easier). Taking this into account leads to

$$\begin{aligned}
&\operatorname{depth}(T_1) - \operatorname{depth}(T_0) = \\
&T_1(i,j) - T_0(i,j) + \min\bigl\{T_1(i,j+1) - T_1(i,j),\; T_1(i,j+1) - T_1(i-1,j+1) - 1\bigr\} \\
&\qquad\qquad - \min\bigl\{T_0(i,j+1) - T_0(i,j),\; T_0(i,j+1) - T_0(i-1,j+1) - 1\bigr\} \\
&\qquad\qquad + \min\bigl\{T_1(i+1,j) - T_1(i+1,j-1),\; T_1(i+1,j) - T_1(i,j) - 1\bigr\} \\
&\qquad\qquad - \min\bigl\{T_0(i+1,j) - T_0(i+1,j-1),\; T_0(i+1,j) - T_0(i,j) - 1\bigr\}.
\end{aligned}$$

Recall that T_0 and T_1 coincide outside of (i,j). For simplicity, set

$$y := T_r(i+1,j), \quad z := T_r(i,j+1), \quad u := T_r(i+1,j-1), \quad v := T_r(i-1,j+1),$$

for $r = 0, 1$. Also let

$$x := T_0(i,j), \qquad x' := T_1(i,j) = \min(y-1,\, v+1).$$

Using $\min(a,b) = \frac{1}{2}(a+b-|a-b|)$, this gives

$$\begin{aligned}
&\operatorname{depth}(T_1) - \operatorname{depth}(T_0) \\
&\quad = x' - x + \min(z-x',\, z-v-1) - \min(z-x,\, z-v-1) \\
&\qquad\qquad + \min(y-x'-1,\, y-u) - \min(y-x-1,\, y-u) \\
&\quad = x' - x + \tfrac{1}{2}(2z - x' - v - 1 - |x'-v-1|) - \tfrac{1}{2}(2z - x - v - 1 - |x-v-1|) \\
&\qquad\qquad + \tfrac{1}{2}(2y - x' - u - 1 - |x'-u+1|) - \tfrac{1}{2}(2y - x - u - 1 - |x-u+1|) \\
&\quad = \tfrac{1}{2}\bigl((|x-u+1| + |x-v-1|) - (|x'-u+1| + |x'-v-1|)\bigr) \\
&\quad = \tfrac{1}{2}\bigl(f(x) - f(x')\bigr),
\end{aligned}$$

where

$$f(a) := |a-u+1| + |a-v-1|.$$

It is elementary that $f(a)$ takes the minimal value throughout the (real) interval

$$[\min(v+1,\, u-1),\; \max(v+1,\, u-1)].$$

Notice that x' is in this interval: $x' \geq \min(v+1, u-1)$ since $y \geq u$. On the other hand, $x' \leq v+1 \leq \max(v+1, u-1)$. Since f attains its minimum at x', we have $f(x) - f(x') \geq 0$, so $\mathrm{depth}(T_1) \geq \mathrm{depth}(T_0)$ as required.

Repeating this procedure so long as the undesirable property (3-3) still holds, we obtain successively $T_0, T_1, T_2, T_3, \ldots$. We claim that after a finite number of iterations (3-3) finally fails for some T_k, $k \geq 0$. To see this, let the vector

$$\boldsymbol{u}(T) = (u_1, u_2, \ldots, u_{|\lambda(w)|})$$

measure how far $T \in \mathrm{SSYT}(\lambda(w), \boldsymbol{b})$ is from failing (3-3): Order the boxes in $\lambda(w)$ from left to right, and in each column from bottom up. For example, in Example 1.6, the order is

3	6	9	
2	5	8	11
1	4	7	10

For each $1 \leq i \leq |\lambda(w)|$, define u_i to be 0 if the i-th box is a northeast corner or if (3-1) holds; otherwise let $u_i = 1$. Then $\boldsymbol{u}(T) = (0, 0, \ldots, 0)$ means that we are in the good case that (3-3) fails. We define a pure reverse lex order on $\{0, 1\}^{|\lambda(w)|}$: given $\boldsymbol{u}, \boldsymbol{u}' \in \{0, 1\}^{|\lambda(w)|}$, we say that $\boldsymbol{u} > \boldsymbol{u}'$ if

$$u_{|\lambda(w)|} = u'_{|\lambda(w)|}, \quad u_{|\lambda(w)|-1} = u'_{|\lambda(w)|-1}, \quad \ldots, \quad u_{i+1} = u'_{i+1}, \quad u_i > u'_i,$$

for some i. It is straightforward to check that $\boldsymbol{u}(T_t) > \boldsymbol{u}(T_{t+1})$ at each step t, so the procedure must eventually terminate, say at step k, with $\boldsymbol{u}(T_k) = (0, 0, \ldots, 0)$, as desired.

Let $T = T_k$ be the output of the procedure above. We want to apply Lemma 3.3 to conclude that $T_k(i, j)$ is in the image of Ψ. We must verify conditions (a) and (b) of the lemma.

Since (3-3) fails, every box that is not a northeast corner has (3-1) holding. In particular, this includes every box described by (a), and so (a) holds.

To check (b), let $\mathscr{L} := \widehat{i} - i$ be the leg length of (i, j). Since (i, j) is special, $\mathscr{L} = |\mathrm{arm}(i, j)|$; moreover, we can apply the argument in the proof of Lemma 3.2(iii) to the subset of the Young diagram $\lambda(w)$ consisting of those boxes strictly above row i and weakly to the right of column j, and conclude that the following boxes lie in $\lambda(w)$:

$$(\widehat{i}, j+1), \quad (\widehat{i}-1, j+2), \quad \ldots, \quad (\widehat{i}-\mathscr{L}+1, j+\mathscr{L}).$$

In particular, the boxes

$$(\widehat{i}, j), \quad (\widehat{i}-1, j+1), \quad (\widehat{i}-2, j+2), \quad \ldots, \quad (\widehat{i}-\mathscr{L}, j+\mathscr{L})$$

are not the northeast corners of $\lambda(w)$; hence (3-1) holds for them by the construction of $T = T_k$. By (3-1), we have

$$T(\widehat{i}-m,\, j+m) \geq T(\widehat{i},\, j) - m, \qquad \text{for } m = 0, 1, \ldots, \mathscr{L}. \tag{3-7}$$

Since $(\widehat{i'}, j')$ is to the right of $(\widehat{i'}, j+(\widehat{i}-\widehat{i'}))$, we have

$$T(\widehat{i'}, j') \geq T(\widehat{i'}, j+(\widehat{i}-\widehat{i'})) = T(\widehat{i}-(\widehat{i}-\widehat{i'}), j+(\widehat{i}-\widehat{i'})) \geq T(\widehat{i}, j) - (\widehat{i}-\widehat{i'}),$$

where the last inequality holds because of (3-7) for $m = \widehat{i}-\widehat{i'}$, and since the hypothesis that (i, j) is weakly southwest of (i', j') implies $\widehat{i}-\widehat{i'} \leq \mathscr{L}-1$. Thus,

$$T(\widehat{i}, j) - \widehat{i} \leq T(\widehat{i'}, j') - \widehat{i'}.$$

Therefore condition (b) holds.

Concluding, there exists $\mathfrak{D} \in \operatorname{drift}(v, w)$ such that $\Psi(\mathfrak{D}) = T_k$ and $\operatorname{wt}(\mathfrak{D}) = \operatorname{depth}(T_k)$. Then

$$\operatorname{wt}(\mathfrak{D}) = \operatorname{depth}(T_k) \geq \operatorname{depth}(T_{k-1}) \geq \cdots \geq \operatorname{depth}(T_0) \tag{3-8}$$

and so $\deg P_{v,w}(q) \geq \deg H_{v,w}(q)$, as was to be shown. $\square$

4. A ball of drift configurations

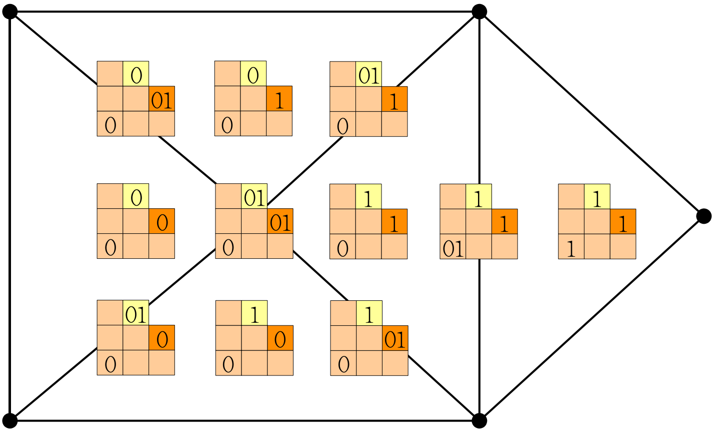
***Construction of* $\mathbf{KL}_{v,w}$.** In order to emphasize the combinatorial relations of drift configurations to Young tableaux, consider an equivalent formulation of drift configurations: A *semistandard* (*ordinary*) *drift tableau* T bijectively associated to $\mathfrak{D}$ is a filling of each continent C of Pangaea(v, w) by the distance C has moved from Pangaea(v, w).

Similarly, a *set-valued drift tableau* is a filling of each continent by some nonempty set of nonnegative integers; it is *semistandard* if any ordinary drift tableau it contains (in the obvious sense) is semistandard. It is *limit semistandard* if it contains at least one semistandard (ordinary) drift tableau. The *empty-face drift tableau* $\mathscr{E}_{v,w}$ is the set-valued drift tableau that is the union of all semistandard ordinary ones.

Define $\mathrm{KL}_{v,w}$ to be the simplicial complex whose faces are indexed by limit semistandard drift tableau and where face containment is by reverse containment of drift tableau. In particular, the vertices are labeled by limit semistandard tableaux $(b \not\mapsto y)$ obtained by removing precisely one entry y from a set $\mathscr{E}_{v,w}(b)$ of the box $b \in \lambda(w)$, provided $|\mathscr{E}_{v,w}(b)| > 1$. (It will be convenient to also consider *phantom vertices* which are those $(b \not\mapsto y)$ where $|\mathscr{E}_{v,w}(b)| = 1$; these become honest vertices after coning over $\mathrm{KL}_{v,w}$.)

This gives an example of a tableau complex in the sense of [Knutson et al. 2008]. We illustrate the case discussed in Example 1.6, showing the interior faces of the

2-dimensional complex $\mathrm{KL}_{23465178 9\underline{10}, \underline{10}954382761}$:



The claims in Theorem 1.4 about the structure of $\mathrm{KL}_{v,w}$ then follow immediately from [Knutson et al. 2008, Theorem 2.8]. We conclude that the interior faces of $\mathrm{KL}_{v,w}$ are labeled by semistandard set-valued drift tableaux while the exterior faces are labeled by nonsemistandard but limit semistandard tableaux. Also the codimension of a face $\mathcal{D}$ is $|\mathcal{D}| - \#\text{continents}$, the number of "extra" entries of $\mathcal{D}$.

***K*-polynomials of $\mathbf{KL}_{v,w}$.** Let us take this opportunity to formalize a connection between the K-polynomials of $\mathrm{KL}_{v,w}$ and $P_{v,w}(q)$. We will utilize facts collected about general tableau complexes from [Knutson et al. 2008, Section 4]. Let V be the set of vertices of a simplicial complex Δ and set $R = \Bbbk[\Delta]$ to be the polynomial ring in variables $x_{\mathfrak{v}}$ for $\mathfrak{v} \in V$. This is the ambient ring for the *Stanley–Reisner ideal* $I_\Delta = \langle \prod_{\mathfrak{v}\in F} x_{\mathfrak{v}} : F \text{ is not a face of } \Delta \rangle$ of Δ, and R/I_Δ is the *Stanley–Reisner ring*. We use the alphabet $\boldsymbol{t}_{\mathfrak{v}} = \{t_{\mathfrak{v}} : \mathfrak{v} \in V\}$ for the finely graded Hilbert series $\mathrm{Hilb}(R/I_\Delta; \boldsymbol{t})$ and K-polynomials $\mathcal{K}(R/I_\Delta, \boldsymbol{t})$.

Let us define a family of polynomials for $v \le w$, where w is covexillary. We will see this is a hybrid of the K-polynomial of $\mathrm{KL}_{v,w}$ and the Kazhdan–Lusztig polynomial $P_{v,w}(q)$:

$$\mathfrak{P}_{v,w}(\beta; \boldsymbol{t}) = \sum_{\mathcal{D}\in \mathrm{SVDT}(v,w)} \beta^{|\mathcal{D}|-\#\text{continents}(v,w)} \prod_{b\in\lambda(w)} \prod_{y\in\mathcal{D}(b)} (1 - t_{(b \not\mapsto y)}), \qquad (4\text{-}1)$$

where $\mathrm{SVDT}(v, w)$ is the set of set-valued drift tableaux associated to drift configurations in $\mathrm{drift}(v, w)$, $|\mathcal{D}|$ is the number of entries in $\mathcal{D}$, and $\#\text{continents}(v, w)$ is the number of continents in $\mathrm{Pangaea}(v, w)$. There are a number of interesting specializations of this polynomial. Here we do not assume $|\mathcal{E}_{v,w}(b)| > 1$, i.e., $(b \not\mapsto y)$ might be a phantom vertex.

By the ballness/sphereness claim of $\mathrm{KL}_{v,w}$ from Theorem 1.4, together with [Knutson et al. 2008, Theorem 4.3], it follows that

$$\mathfrak{P}_{v,w}(-1;\boldsymbol{t}) = \mathcal{K}(R/I_{\mathrm{KL}_{\mathrm{v,w}}};\boldsymbol{t}). \tag{4-2}$$

One can consider a vertex decomposition of any complex Δ at a vertex $\mathfrak{v}$. This is given by $\Delta = \mathrm{del}_{\mathfrak{v}}(\Delta) \cup \mathrm{star}_{\mathfrak{v}}(\Delta)$, where $\mathrm{del}_{\mathfrak{v}}(\Delta) = \{F \in \Delta : \mathfrak{v} \notin F\}$ is the *deletion* of $\mathfrak{v}$ and $\mathrm{star}_{\mathfrak{v}}(\Delta) = \{F \in \Delta : F \cup \{\mathfrak{v}\} \in \Delta\}$ is the *star* of $\mathfrak{v}$. Automatically one has, for $\mathfrak{v} = (b \not\mapsto y)$,

$$\begin{aligned}\mathcal{K}&(R/I_{\mathrm{KL}_{v,w}};\boldsymbol{t})\\ &= t_{(b\not\mapsto y)}\mathcal{K}(R/I_{\mathrm{del}_{(b\not\mapsto y)}(\mathrm{KL}_{v,w})};\boldsymbol{t}) + (1-t_{(b\not\mapsto y)})\mathcal{K}(R/I_{\mathrm{star}_{(b\not\mapsto y)}(\mathrm{KL}_{v,w})};\boldsymbol{t}).\end{aligned} \tag{4-3}$$

By tracing the specializations below, one should eventually interpret recursions from [Lascoux and Schützenberger 1981] for $P_{v,w}(q)$ using (4-3) and thus vertex decompositions of $\mathrm{KL}_{v,w}$. We do not pursue this here.

Consider

$$\mathfrak{P}_{v,w}(-1;t_{(b\not\mapsto y)} \mapsto 1-x_y) = \sum_{\mathcal{D}\in\mathrm{SVDT}(v,w)} (-1)^{|\mathcal{D}|-\#\mathrm{continents}(v,w)}\boldsymbol{x}^{\mathcal{D}}, \tag{4-4}$$

where

$$\boldsymbol{x}^{\mathcal{D}} = \prod_{i\geq 0} x_i^{\#i\text{'s appearing in }\mathcal{D}}.$$

Another specialization is given by

$$\mathfrak{P}_{v,w}(0;t_{(b\not\mapsto y)} \mapsto 1-x_y) = \sum_{\mathcal{D}\in\mathrm{SSDT}(v,w)} \boldsymbol{x}^{\mathcal{D}}, \tag{4-5}$$

where $\mathrm{SSDT}(v, w)$ is the set of ordinary, semistandard drift tableau associated to v, w. (In setting $\beta = 0$ we take the convention that $0^0 = 1$ in (4-1).)

Finally, by considering the principal specialization of (4-5) we have

$$\mathfrak{P}_{v,w}(0;t_{(b\not\mapsto y)} \mapsto 1-q^y) = P_{v,w}(q).$$

5. The proof of Theorem 1.4(I)

Proof that $Q_{v,w}(q) = P_{v,w}(q)$. We give a weight-preserving bijection between $\mathrm{drift}(v, w)$ and the trees weight-enumerated by Lascoux's rule [1995] for $P_{v,w}(q)$. We mostly follow the presentation of his rule found in [Billey and Lakshmibai 2000, 6.3.29].

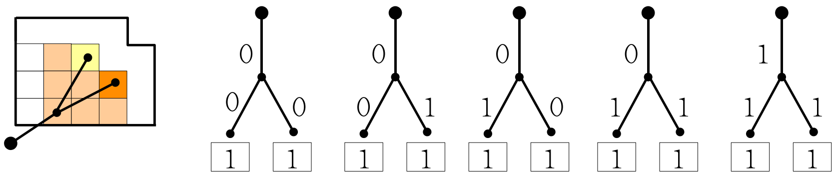
Given $\mathcal{D} \in \mathrm{drift}(v, w)$, construct a rooted, edge-labeled tree $\mathcal{T}$ as follows. Associate to each continent C a non-root vertex of $\mathcal{T}$. Moreover if the special box b of C is southwest of the special box b' of an adjacent continent C', then we draw an edge between the corresponding vertices. If there is no special box strictly southwest of b, then the corresponding vertex is joined to the root of $\mathcal{T}$.

Thus, each 1×1 continent $C = \{(h, \lambda(w)_h)\}$ (equivalently, those that come from northeast corners of $\lambda(w)$) corresponds to a leaf p of $\mathcal{T}$. Now we bound the edge incident to p by $b_h - h$, where

$$b_h = \max\{m : B(v, w)_m \geq \lambda(w)_h + m - h\}.$$

Let $DL(\mathcal{T})$ be the set of all edge labelings of $\mathcal{T}$ by nonnegative integers such that the labels weakly increase from root to leaf. For any edge labeled tree $\mathcal{G}$ let $|\mathcal{G}|$ be the sum of the edge labels of $\mathcal{G}$.

As an example, here are the edge-labeled trees for the drift configurations in Example 1.6. (The framed number below each leaf is the bound for that leaf.)



Lemma 5.1. *There is a bijection* $\Phi : \mathrm{drift}(v, w) \to DL(\mathcal{T})$ *such that*

$$\mathrm{wt}(\mathcal{D}) = |\Phi(\mathcal{D})|.$$

Proof. Define $\Phi(\mathcal{D})$ to be the edge labeling of $\mathcal{T}$ such that the edge associated to a continent C (i.e., the edge whose child end is the vertex associated to C) is labeled by the distance that C has drifted in $\mathcal{D}$. That the labels are weakly increasing in $\Phi(\mathcal{D})$ is implied by the condition that the continents do not overlap in $\mathcal{D}$. Note that if C is a 1×1 continent then $b_h - h$ is the largest distance that C can drift inside $B(v, w)$; this accounts for the leaf bound. (For an example, see diagram immediately above.) It is then easy to check that Φ is the desired bijection. □

Lascoux's rule constructs a tree $\mathcal{T}'$ as follows: For the partition $\lambda(w)$, the *parenthesis-word* is a word using "(" and ")" and obtained by walking with east and south steps along the northeast border of $\lambda(w)$. We record a "(" for each east step and a ")" for each south step. Now pair left and right parentheses starting from the closest pairs "( )". Each pair corresponds to a vertex of the tree; the closest pairs are associated to leaves and a pair encloses its children. Unpaired parentheses do not contribute to the tree. This process results in a directed forest. Finally, we introduce an additional root and attach an edge to the root of each tree in the forest.

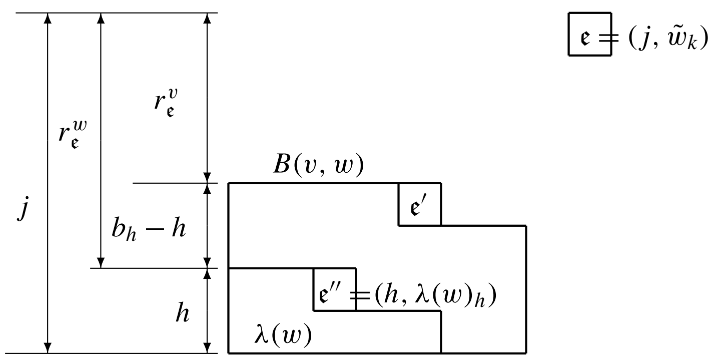
Lemma 5.2. *There is a graph isomorphism* $\delta : \mathcal{T} \to \mathcal{T}'$. *Under this isomorphism, if* v *corresponds to a* 1×1 *continent associated to a corner* c *of* $\lambda(w)$, *then* $\delta(v)$ *corresponds to a closest parenthesis pair associated to the same corner* c.

Proof. Each leaf of $\mathcal{T}$ corresponds to a corner c of $\lambda(w)$. On the other hand, this corner gives rise to a closest pair "( )" in Lascoux's construction, which corresponds to a leaf of $\mathcal{T}'$. Thus we can construct a bijection between the leaves of the two trees, which we now argue extends to the bijection δ between the two trees themselves.

A continent C is a *z-continent* if it is defined by a z-special box b. Fix a vertex $v \in \mathcal{T}$ associated to such a continent. By construction, each child of v is a vertex $\{v'\}$ associated to a y-continent C' adjacent and northeast of C in Pangaea(v, w), where $y < z$. Since $b \in C$ is a special box, by using the fact that $|\mathrm{arm}(b)| = |\mathrm{leg}(b)|$ we have that the column b is in corresponds to a "(" and the row b is in corresponds to a ")", where these two parentheses are paired with one another in the parenthesis word. Clearly, this pair gives a vertex $v' \in \mathcal{T}'$, and all vertices of $\mathcal{T}'$ arise this way. That is, there is a bijection at the level of vertices $\delta : \mathcal{T} \to \mathcal{T}'$. Moreover, that the children of $\delta(v)$ are exactly $\{\delta(v')\}$ (for children v' of v) is also immediate from the constructions of $\mathcal{T}$ and $\mathcal{T}'$ $\square$

Lascoux's rule similarly defines increasing edge labelings $EL(\mathcal{T})$ on $\mathcal{T}$ as we did for $DL(\mathcal{T})$. It remains to check that these labelings are the same as the ones in $DL(\mathcal{T})$. For this, we only need to show that the bound attached to the leaves are the same. In [Billey and Lakshmibai 2000, 6.3.29, Step 2], for each given leaf, a bigrassmannian permutation is determined in three substeps, from which Lascoux's leaf bounds are determined. We now explain these steps. (For readers comparing what follows with that reference, note that Billey and Lakshmibai's x is our $\tilde{w} = w^{-1}w_0$, while their w is our $\tilde{v} = v^{-1}w_0$.)

The reader may find the following diagram useful for the description of Lascoux's labeling process:



Substep (1): leaves p of $\mathcal{T}$ correspond to distinct numbers in the code of $\tilde{w}$. The code $(c_1, \ldots, c_n)$ of $\tilde{w}$ is given by

$$c_i = \#\{j > i \mid \tilde{w}_j < \tilde{w}_i\} = \#\{\text{boxes of } D(w) \text{ in row } i\}.$$

Recall $\lambda(w)$ is the result of sorting this code into decreasing order. A leaf p of $\mathcal{T}$ corresponds to a corner $\mathfrak{e}'' = (h, \lambda(w)_h)$ of $\lambda(w)$. Associate $\lambda(w)_h$ to p. This $\lambda(w)_h$ is equal to c_i for some i. Clearly a different c_i is assigned to each p.

Substep (2): $\lambda(w)_h$ gives a crossing of $\tilde{w}$. By definition, a *crossing* of $\tilde{w}$ is a 4-tuple $(i, j, j+1, k)$ satisfying

$$\tilde{w}_{j+1} \leq \tilde{w}_k < \tilde{w}_i \leq \tilde{w}_j, \quad \tilde{w}_i = \tilde{w}_k + 1 \quad \text{for } i \leq j < k; \tag{5-1}$$

see [Lascoux and Schützenberger 1996]. Now given the $\mathfrak{e}''$ associated to p, there is a unique essential box $\mathfrak{e}$ in $D(w)$ that is diagonally northeast of $\mathfrak{e}''$. We define j and k by declaring that the coordinates of $\mathfrak{e}$ are $(j, \tilde{w}_k)$. Let i be such that $\tilde{w}_i = \tilde{w}_k + 1$.

We claim that $(i, j, j+1, k)$ forms a crossing. Let us first check the weak inequalities of $\tilde{w}_{j+1} \leq \tilde{w}_k < \tilde{w}_i \leq \tilde{w}_j$ (the strict inequality being true by definition). For the rightmost inequality, we have $\tilde{w}_j = w^{-1}w_0(j) = w^{-1}_{n-j+1}$, which in words is the column position of the $\bullet$ of $G(w)$ that necessarily must be to the right of $\mathfrak{e}$, which itself is in column $\tilde{w}_k$. In other words $\tilde{w}_k \leq \tilde{w}_j$. Now, for the leftmost inequality, note $\tilde{w}_{j+1} = w^{-1}w_0(j+1) = w^{-1}(n-j)$ which is the column position of the $\bullet$ of $G(w)$ in row $j+1$. Since $\mathfrak{e}$ is an essential box, that $\bullet$ must be weakly to the left, i.e., $\tilde{w}_{j+1} \leq \widetilde{w}_k$, as desired. It remains to check $i \leq j$ and $j < k$. For the former inequality, we compute $w\widetilde{w}_i = n - i + 1$ which is the row position of the $\bullet$ of $G(w)$ in column $\tilde{w}_i$. Since $\mathfrak{e}$ is an essential box, the $\bullet$ is weakly below the $\mathfrak{e}$, i.e., $i \leq j$. Similarly, for the latter inequality, we consider $w\widetilde{w}_k = n-k+1$, which is the position of the $\bullet$ of $G(w)$ in column $\tilde{w}_k$. This must be strictly above the $\mathfrak{e}$, i.e., $j < k$.

Now associate the crossing $(i, j, j+1, k)$ to p (and hence $\lambda(w)_h$). Actually, the description in [Billey and Lakshmibai 2000] gives a different way to assign a crossing to p. However, it is straightforward to check that their crossing is same as the one described above.

Substep (3): each crossing gives a maximal bigrassmannian $[a, b, c, d]$ below $\tilde{w}$. Here $[a, b, c, d]$ denotes

$$(1, \ldots, a,\ a+c+1, \ldots, a+c+b,\ a+1, \ldots, a+c,\\ a+c+b+1, \ldots, a+b+c+d) \in S_n.$$

Lascoux's rule associates to $(i, j, j+1, k)$ a maximal bigrassmannian

$$[z, j-z, \tilde{w}_k - z, n - \tilde{w}_k - j + z],$$

where

$$z = \#\{p < j : \tilde{w}_p < \tilde{w}_k\}.$$

Notice that z is the number of $\bullet$'s in $G(w)$ weakly southwest of $\mathfrak{e} = (j, \widetilde{w}_k)$, i.e.,

$$z = r^w_{\mathfrak{e}}. \tag{5-2}$$

This concludes substep (3) of step 2 of [Billey and Lakshmibai 2000].

Lascoux's rule then assigns to p the leaf bound

$$\text{distance}([z,\, j-z,\, \tilde{w}_k - z,\, n-\tilde{w}_k - j + z],\, \tilde{v}),$$

where

$$\text{distance}([a,b,c,d],\tilde{v}) = \max\{r \geq 0 \mid [a-r,\, b+r,\, c+r,\, d-r] \leq \tilde{v}\},$$

and where "$\leq$" refers to Bruhat order on S_n. This completes the description of Lascoux's algorithm.

Recall that $r^{v}_{(a+b,a+c)}$ equals the number of dots of $G(v)$ weakly southwest of $(a+b,\, a+c)$. The proof of the following fact is straightforward to argue (and also follows from the deeper developments in [Lascoux and Schützenberger 1996]):

Lemma 5.3. *For any bigrassmannian permutation $[a,b,c,d]$ and permutation $\tilde{v}$ in S_n, the inequality $[a,b,c,d] \leq \tilde{v}$ is equivalent to $r^{v}_{(a+b,a+c)} \leq a$, where $v = w_0\tilde{v}^{-1}$.* □

Proposition 5.4. *The leaf bounds on $DL(\mathcal{T})$ and $EL(\mathcal{T})$ are the same.*

Proof. By Lemma 5.3,

$$\begin{aligned} [z-r,\, j-z+r,\, \tilde{w}_k - z + r,\, n - \tilde{w}_k - j + z - r] \leq \tilde{v} &\iff \\ r^{v}_{(z-r)+(j-z+r),(z-r)+(\tilde{w}_k - z + r)} \leq z - r &\iff \\ r^{v}_{(j,\tilde{w}_k)} \leq z - r &\iff \\ r^{v}_{\mathfrak{e}} \leq z - r. & \end{aligned} \tag{5-3}$$

Hence, the maximal r such that any of the inequalities (5-3) hold is

$$r = z - r^{v}_{\mathfrak{e}} = r^{w}_{\mathfrak{e}} - r^{v}_{\mathfrak{e}},$$

where we have used (5-2).

In terms of drift configurations, r is the largest distance that a corner $\mathfrak{e}'' = (h, \lambda(w)_h)$ can be moved diagonally northeast and remain in $B(v,w)$ (see [Li and Yong 2011, Lemma 5.7]). By the definition of $B(v,w)$, $b_h = j - r^{v}_{\mathfrak{e}}$. It is also easy to check that $j = h + r^{w}_{\mathfrak{e}}$ (again by the same lemma). Then

$$b_h - h = j - r^{v}_{\mathfrak{e}} - h = (j-h) - r^{v}_{\mathfrak{e}} = r^{w}_{\mathfrak{e}} - r^{v}_{\mathfrak{e}} = r.$$

This completes the proof of the proposition. □

By Lascoux's rule,

$$P_{w_0\tilde{v}, w_0\tilde{w}}(q) (= P_{w_0 v^{-1} w_0, w_0 w^{-1} w_0}(q) = P_{v,w}(q)) = \sum q^{|T|},$$

where the sum is over $EL(T)$ and $|T|$ is the total sum of the edge labels. Since we have established the desired weight-preserving bijection, the claim $Q_{v,q}(q) = P_{v,w}(q)$ then follows.

Remark 5.5. There are two basic symmetries of Kazhdan–Lusztig polynomials: $P_{v,w}(q) = P_{w_0v^{-1}w_0,w_0w^{-1}w_0}(q)$ and $P_{v,w}(q) = P_{v^{-1},w^{-1}}(q)$. The first symmetry is manifest in our rule and drift$(w_0v^{-1}w_0, w_0w^{-1}w_0)$ is obtained by transposing the drift configurations of drift(v, w). For the second, it is an exercise to prove that $\lambda(w) = \lambda(w^{-1})$ and $B(v, w) = B(v^{-1}, w^{-1})$, so drift$(v^{-1}, w^{-1}) =$ drift(v, w).

Remark 5.6. From Theorem 1.4(I) it is not hard to show the following. For $w, v \in S_n$ where w is covexillary and $v \le w$, let k be the number of special boxes of $\lambda(w)$ and let $m = \lfloor \frac{n-k+1}{2} \rfloor$. If $[m]_q = 1+q+\cdots+q^{m-1}$, then $[q^i]P_{v,w}(q) \le [q^i]([m]_q)^k$ for all i. In particular, $P_{v,w}(1) \le m^k$.

6. Another q-analogue of multiplicity

We can think of $H_{v,w}(q)$ as a q-analogue of Hilbert–Samuel multiplicity, in the sense that $H_{v,w}(1) = \mathrm{mult}_{e_v}(X_w)$. Let us point out that in the covexillary setting, there is another q-analogue available. As in Theorem 1.4(II), regard each box of $\lambda(w)$ as a separate country; the "drift configurations" are precisely the *pipe dreams* $P \in \mathrm{Pipes}(v, w)$ in [Li and Yong 2011]. Now let

$$\widetilde{\mathrm{wt}}(P) = q^d,$$

where d is the total of the distance drifted by the countries, and set

$$\widetilde{H}_{v,w}(q) = \sum_{\mathrm{P} \in \mathrm{Pipes(v,w)}} \widetilde{\mathrm{wt}}(P).$$

In the following theorem we use the standard q-notation:

$$[a]_q = 1+q+\cdots+q^{a-1} \quad \text{and} \quad \binom{a}{b}_q = \frac{[a]_q[a-1]_q\cdots[a-b+1]_q}{[b]_q\cdots[1]_q}.$$

Theorem 6.1. $\widetilde{H}_{v,w}(q) = q^{-\sum_{i\ge1}(i-1)\lambda_i} \det\left(\binom{b_i+\lambda_i-i+j-1}{\lambda_i-i+j}_q\right)_{1\le i,j\le \ell(\lambda)}$, *where $\ell(\lambda)$ is the number of nonzero parts of λ and $\boldsymbol{b} = \boldsymbol{b}(\Theta_{v,w})$.*

Proof. For brevity, we refer the reader to the setup of [Li and Yong 2011, Sections 5.2 and 6.2]. Notice that

$$s_{\lambda,\boldsymbol{b}}(1, q, q^2, q^3, \ldots) = \det\left(\binom{b_i+\lambda_i-i+j-1}{\lambda_i-i+j}_q\right)_{1\le i,j\le \ell(\lambda)}$$

where the left-hand side of the equality is the principal specialization of the (single) flagged Schur polynomial for shape $\lambda(w)$ with flag $\boldsymbol{b} = \boldsymbol{b}(\Theta_{v,w})$.

Given a pipe dream $P \in \mathrm{Pipes}(v, w)$ that corresponds to a flagged semistandard Young tableau T, write

$$\mathrm{wt}_x(P) := \mathrm{wt}_x(T)$$

to mean the usual multivariate weight assigned to T (that is, the one such that $s_{\lambda,\boldsymbol{b}}(x_1, x_2, x_3, \ldots) = \sum_T \mathrm{wt}_x(T)$). Let $\mathrm{wt}'_q(P)$ be the principal specialization of $\mathrm{wt}_x(P)$ given by $x_i \mapsto q^{i-1}$ and finally set

$$\mathrm{wt}_q(P) = q^{-\sum_{i\geq 1}(i-1)\lambda_i} \times \mathrm{wt}'_q(P).$$

It remains to show that $\mathrm{wt}_q(P) = \widetilde{\mathrm{wt}}(P)$ for each P. To do this, let us induct on $\widetilde{\mathrm{wt}}(P) \geq 0$. The base case that $\widetilde{\mathrm{wt}}(P) = 0$, i.e., where P is the starting configuration holds since $\mathrm{wt}'_q(P) = q^{\sum_{i\geq 1}(i-1)\lambda_i}$.

Now suppose $\widetilde{\mathrm{wt}}(P) > 0$. Then there is a P' such that a move of the form

$$\begin{matrix} \cdot & \cdot \\ + & \cdot \end{matrix} \quad \mapsto \quad \begin{matrix} \cdot & + \\ \cdot & \cdot \end{matrix}$$

in some 2×2 subsquare of $[n]\times[n]$ brought us to P (and no other $+$ in P' has changed). Thus, we can compare $\mathrm{wt}_x(P')$ and $\mathrm{wt}_x(P)$: the latter only differs from the former in that some factor of x_i changed to x_{i+1} (where i and $i+1$ are the rows changed by the move above). Hence applying induction we have

$$\mathrm{wt}_q(P) = \mathrm{wt}_q(P') \times q = \widetilde{\mathrm{wt}}(P') \times q = \widetilde{\mathrm{wt}}(P),$$

as desired. $\square$

It is clear from Theorem 1.4 that

$$P_{v,w}(q) \preceq \widetilde{H}_{v,w}(q).$$

With the same proof that we used for $H_{v,w}(q)$, one shows that $\widetilde{H}_{v,w}(q)$ is upper semicontinuous. However, in general $\widetilde{H}_{v,w}(q) \neq H_{v,w}(q)$. Moreover, we do not know any algebraic/geometric measure for general Schubert varieties that specializes to $\widetilde{H}_{v,w}(q)$.

7. Concluding remarks

We are presently unaware of any geometric proof of the inequality of Theorem 1.2. For general Y, let us assume, for simplicity of our discussion, that all odd local intersection cohomology groups vanish, and set

$$P_{p,Y}(q) = \sum_{i\geq 0} \dim(\mathcal{H}^{2i}_p(Y)) q^i.$$

Question 7.1. Under what assumptions is either the inequality $P_{p,Y}(q) \preceq H_{p,Y}(q)$ and/or the weaker inequality $P_{p,Y}(1) \leq H_{p,Y}(1) (= \mathrm{mult}_p(Y))$ true?

Our results on $H_{v,w}(q)$ are based on the degeneration, flat over $\mathrm{Spec}(\mathbb{Z})$, given in [Li and Yong 2011]. Hence Theorem 1.7 is valid over a field $\Bbbk$ of arbitrary characteristic and Conjecture 1.1 seems similarly valid. However, the arguments of [Li and Yong 2011] also prove that the projectivized tangent cones of the Kazhdan–Lusztig varieties $\mathcal{N}_{v,w}$ are isomorphic to those for $\mathcal{N}_{\mathrm{id},\Theta_{v,w}}$. It is then not hard to construct some cograssmannian v', w' with the same property. We do not know if $\mathcal{N}_{v,w}$ and any such $\mathcal{N}_{v',w'}$ are actually isomorphic, although a number of useful implications would be a consequence of this fact.

A number of formulae have been obtained for $P_{v,w}(q)$. For example, general, nonpositive formulae have been obtained in [Billera and Brenti 2011] and [Brenti 1998]. Beyond the covexillary case, few positive formulae are known; see, e.g., [Billey and Warrington 2001] (which treats the 321-hexagon avoiding case) and the references therein. It would be interesting to try to extend our main theorems to these other contexts as well.

Finally, we believe many of the ideas of this paper can be extended to other Lie groups. In particular, we expect Theorems 1.2, 1.4 and 1.7 to have analogues for (co)minuscule G/P; cf. [Boe 1988]. However, this requires sufficient technicalities that it is better left to a separate treatment.

Acknowledgements

We thank Sara Billey, Xuhua He, Hiroshi Naruse and Alexander Woo for useful suggestions and questions that inspired this work. We also thank Jonah Blasiak, Allen Knutson, Venkatramani Lakshmibai, Ezra Miller, Greg Warrington and the anonymous referee for helpful comments. AY is partially supported by NSF grants DMS-0601010 and DMS-0901331.

References

[Billera and Brenti 2011] L. J. Billera and F. Brenti, "Quasisymmetric functions and Kazhdan–Lusztig polynomials", *Israel J. Math.* **184** (2011), 317–348. MR 2823980

[Billey and Lakshmibai 2000] S. Billey and V. Lakshmibai, *Singular loci of Schubert varieties*, Progress in Mathematics **182**, Birkhäuser, Boston, MA, 2000. MR 2001j:14065 Zbl 0959.14032

[Billey and Warrington 2001] S. C. Billey and G. S. Warrington, "Kazhdan–Lusztig polynomials for 321-hexagon-avoiding permutations", *J. Algebraic Combin.* **13**:2 (2001), 111–136. MR 2002f:05161 Zbl 0979.05109

[Boe 1988] B. D. Boe, "Kazhdan–Lusztig polynomials for Hermitian symmetric spaces", *Trans. Amer. Math. Soc.* **309**:1 (1988), 279–294. MR 89i:22024 Zbl 0669.17009

[Braden and MacPherson 2001] T. Braden and R. MacPherson, "From moment graphs to intersection cohomology", *Math. Ann.* **321**:3 (2001), 533–551. MR 2003g:14030 Zbl 1077.14522

[Brenti 1994] F. Brenti, "Log-concave and unimodal sequences in algebra, combinatorics, and geometry: an update", pp. 71–89 in *Jerusalem combinatorics '93*, edited by H. Barcelo and G. Kalai, Contemp. Math. **178**, Amer. Math. Soc., Providence, RI, 1994. MR 95j:05026 Zbl 0813.05007

[Brenti 1998] F. Brenti, "Lattice paths and Kazhdan–Lusztig polynomials", *J. Amer. Math. Soc.* **11**:2 (1998), 229–259. MR 98h:20072 Zbl 0904.20033

[Brion 2004] M. Brion, "Lectures on the geometry of flag varieties", lecture notes, Summer school on Schubert Varieties, (Warsaw, 2003), 2004. arXiv math/0410240

[Bruns and Herzog 1993] W. Bruns and J. Herzog, *Cohen–Macaulay rings*, Cambridge Studies in Advanced Mathematics **39**, Cambridge University Press, Cambridge, 1993. MR 95h:13020 Zbl 0788.13005

[Buch 2002] A. S. Buch, "A Littlewood–Richardson rule for the K-theory of Grassmannians", *Acta Math.* **189**:1 (2002), 37–78. MR 2003j:14062 Zbl 1090.14015

[Ikeda and Naruse 2009] T. Ikeda and H. Naruse, "Excited Young diagrams and equivariant Schubert calculus", *Trans. Amer. Math. Soc.* **361**:10 (2009), 5193–5221. MR 2010i:05351 Zbl 05613028

[Irving 1988] R. S. Irving, "The socle filtration of a Verma module", *Ann. Sci. École Norm. Sup.* (4) **21**:1 (1988), 47–65. MR 89h:17015 Zbl 0706.17006

[Jones and Woo 2010] B. Jones and A. Woo, "Kazhdan–Lusztig polynomials for cograssmannian permutations", preprint, 2010.

[Kazhdan and Lusztig 1979] D. Kazhdan and G. Lusztig, "Representations of Coxeter groups and Hecke algebras", *Invent. Math.* **53**:2 (1979), 165–184. MR 81j:20066 Zbl 0499.20035

[Kazhdan and Lusztig 1980] D. Kazhdan and G. Lusztig, "Schubert varieties and Poincaré duality", pp. 185–203 in *Geometry of the Laplace operator* (Honolulu, 1979), Proc. Sympos. Pure Math. **36**, Amer. Math. Soc., Providence, R.I., 1980. MR 84g:14054 Zbl 0461.14015

[Knutson 2009] A. Knutson, "Frobenius splitting, point counting, and degeneration", preprint, 2009. arXiv 0911.4941

[Knutson and Miller 2005] A. Knutson and E. Miller, "Gröbner geometry of Schubert polynomials", *Ann. of Math.* (2) **161**:3 (2005), 1245–1318. MR 2006i:05177

[Knutson et al. 2008] A. Knutson, E. Miller, and A. Yong, "Tableau complexes", *Israel J. Math.* **163** (2008), 317–343. MR 2009b:05264 Zbl 1145.05055

[Knutson et al. 2009] A. Knutson, E. Miller, and A. Yong, "Gröbner geometry of vertex decompositions and of flagged tableaux", *J. Reine Angew. Math.* **630** (2009), 1–31. MR 2010m:13017

[Kreuzer and Robbiano 2005] M. Kreuzer and L. Robbiano, *Computational commutative algebra*, vol. 2, Springer, Berlin, 2005. MR 2006h:13036 Zbl 1090.13021

[Lakshmibai and Sandhya 1990] V. Lakshmibai and B. Sandhya, "Criterion for smoothness of Schubert varieties in $\mathrm{Sl}(n)/B$", *Proc. Indian Acad. Sci. Math. Sci.* **100**:1 (1990), 45–52. MR 91c:14061 Zbl 0714.14033

[Lascoux 1995] A. Lascoux, "Polynômes de Kazhdan–Lusztig pour les variétés de Schubert vexillaires", *C. R. Acad. Sci. Paris Sér. I Math.* **321**:6 (1995), 667–670. MR 96g:05144 Zbl 0886.05117

[Lascoux and Schützenberger 1981] A. Lascoux and M.-P. Schützenberger, "Polynômes de Kazhdan & Lusztig pour les grassmanniennes", pp. 249–266 in *Young tableaux and Schur functors in algebra and geometry* (Toruń, 1980), Astérisque **87**, Soc. Math. France, Paris, 1981. MR 83i:14045

[Lascoux and Schützenberger 1996] A. Lascoux and M.-P. Schützenberger, "Treillis et bases des groupes de Coxeter", *Electron. J. Combin.* **3**:2 (1996), Res. paper 27. MR 98c:05168

[Li and Yong 2011] L. Li and A. Yong, "Some degenerations of Kazhdan–Lusztig polynomials and multiplicities of Schubert varieties", *Adv. Math.* **229**:1 (2011), 633–667.

[Manivel 2001] L. Manivel, *Symmetric functions, Schubert polynomials and degeneracy loci*, SMF/AMS Texts and Monographs **6**, Amer. Math. Soc., Providence, RI, 2001. MR 2002h:05161 Zbl 0998.14023

[Polo 1999] P. Polo, "Construction of arbitrary Kazhdan–Lusztig polynomials in symmetric groups", *Represent. Theory* **3** (1999), 90–104. MR 2000j:14079 Zbl 0968.14029

[Ramanathan 1985] A. Ramanathan, "Schubert varieties are arithmetically Cohen–Macaulay", *Invent. Math.* **80**:2 (1985), 283–294. MR 87d:14044 Zbl 0541.14039

[Rubey 2005] M. Rubey, "The h-vector of a ladder determinantal ring cogenerated by 2×2 minors is log-concave", *J. Algebra* **292**:2 (2005), 303–323. MR 2007a:05147 Zbl 1127.13009

[Shigechi and Zinn-Justin 2010] K. Shigechi and P. Zinn-Justin, "Path representation of maximal parabolic Kazhdan–Lusztig polynomials", preprint, 2010. arXiv 1001.1080

[Stanley 1989] R. P. Stanley, "Log-concave and unimodal sequences in algebra, combinatorics, and geometry", pp. 500–535 in *Graph theory and its applications: East and West* (Jinan, 1986), edited by M. F. Capobianco et al., Ann. New York Acad. Sci. **576**, New York Acad. Sci., New York, 1989. MR 92e:05124 Zbl 0792.05008

[Woo and Yong 2008] A. Woo and A. Yong, "Governing singularities of Schubert varieties", *J. Algebra* **320**:2 (2008), 495–520. MR 2009j:14063 Zbl 1152.14046

[Woo and Yong 2009] A. Woo and A. Yong, "A Gröbner basis for Kazhdan–Lusztig ideals", preprint, 2009. To appear in *Amer. J. Math.* arXiv 0909.0564

Communicated by Victor Reiner
Received 2010-06-19 Revised 2010-08-22 Accepted 2010-10-01

li2345@oakland.edu *Department of Mathematics and Statistics, Oakland University, Rochester, MI 48309, United States*

ayong@math.uiuc.edu *Department of Mathematics, University of Illinois at Urbana–Champaign, Urbana, IL 61801, United States*

ALGEBRA AND NUMBER THEORY 5:5(2011)

Renormalization and quantum field theory

Richard E. Borcherds

The aim of this paper is to describe how to use regularization and renormalization to construct a perturbative quantum field theory from a Lagrangian. We first define renormalizations and Feynman measures, and show that although there need not exist a canonical Feynman measure, there is a canonical orbit of Feynman measures under renormalization. We then construct a perturbative quantum field theory from a Lagrangian and a Feynman measure, and show that it satisfies perturbative analogues of the Wightman axioms, extended to allow time-ordered composite operators over curved spacetimes.

1. Introduction

We give an overview of the construction of a perturbative quantum field theory from a Lagrangian. We start by translating some terms in physics into mathematical terminology.

Definition 1. Spacetime is a smooth finite-dimensional metrizable manifold M, together with a "causality" relation $\leqslant$ that is closed, reflexive, and transitive. We say that two points are *spacelike separated* if they are not comparable; in other words, if neither $x \leqslant y$ nor $y \leqslant x$.

The causality relation $a \leqslant b$ means informally that a occurs before b. The causality relation will often be constructed in the usual way from a Lorentz metric with a time orientation, but since we do not use the Lorentz metric for anything else we do not bother to give M one. The Lorentz metric will later appear implicitly in the choice of a cut propagator, which is often constructed using a metric.

Definition 2. The sheaf of classical fields Φ is the sheaf of smooth sections of some finite-dimensional super vector bundle over spacetime.

This research was supported by a Miller professorship and an NSF grant. I thank the referees for suggesting many improvements.
MSC2000: 22E70.
Keywords: quantum field theory, renormalization, Feynman measure, Hopf algebra, Feynman diagram.

When the sheaf of classical fields is a supersheaf, one uses the usual conventions of superalgebra: in particular the symmetric algebras used later are understood to be symmetric algebras in the superalgebra sense, and the usual superalgebra minus signs should be inserted into formulas whenever the order of two terms is exchanged.

As usual, a global section of a sheaf of things is called a thing, so a classical field φ is a global section of the sheaf Φ of classical fields, and so on. (A subtle point is sometimes things called classical fields in the physics literature are better thought of as sections of the *dual* of the sheaf of classical fields; in practice this distinction does not matter because the sheaf of classical fields usually comes with a bilinear form giving a canonical isomorphism with its dual.)

Definition 3. The sheaf of derivatives of classical fields or simple fields is the sheaf $J\Phi = \mathrm{Hom}(J, \Phi)$, where J is the sheaf of jets of M and the Hom is taken over the smooth functions on M, equal to the inverse limit of the sheaves of jets of finite order of M, as in [Grothendieck 1967, 16.3].

Definition 4. The sheaf of (polynomial) Lagrangians or composite fields $SJ\Phi$ is the symmetric algebra of the sheaf $J\Phi$ of derivatives of classical fields.

Its sections are (polynomial) Lagrangians, in other words polynomial in fields and their derivations, so for example $\lambda\varphi^4 + m^2\varphi^2 + \varphi\partial_i^2\varphi$ is a Lagrangian, but $\sin(\varphi)$ is not.

Perturbative quantum field theories depend on the choice of a Lagrangian L, which is the sum of a free Lagrangian L_F that is quadratic in the fields, and an interaction Lagrangian $L_I \in SJ\Phi \otimes \mathbb{C}[[\lambda_1, \dots, \lambda_n]]$ whose coefficients are infinitesimal, in other words elements of a formal power series ring $\mathbb{C}[[\lambda_1, \dots, \lambda_n]]$ over the reals with constant terms 0.

Definition 5. The sheaf of Lagrangian densities or local actions $\omega SJ\Phi = \omega \otimes SJ\Phi$ is the tensor product of the sheaf $SJ\Phi$ of Lagrangians and the sheaf ω of smooth densities (taken over smooth functions on M).

For a smooth manifold, the (dualizing) sheaf ω of smooth densities (or smooth measures) is the tensor product of the orientation sheaf with the sheaf of differential forms of highest degree, and is noncanonically isomorphic to the sheaf of smooth functions. Densities are roughly "things that can be locally integrated". For example, if M is oriented, then $(\lambda\varphi^4 + m^2\varphi^2 + \varphi\partial_i^2\varphi)d^nx$ is a Lagrangian density.

We use Γ and Γ_c to stand for spaces of global and compactly supported sections of a sheaf. These will usually be spaces of smooth functions (or compactly supported smooth functions) in which case they are topologized in the usual way so that their duals are compactly supported distributions (or distributions) taking values in some sheaf.

Definition 6. A (nonlocal) *action* is a polynomial in local actions, in other words an element of the symmetric algebra $S\Gamma\omega SJ\Phi$ of the real vector space $\Gamma\omega SJ\Phi$ of local actions.

We do not complete the symmetric algebra, so expressions such as $e^{i\lambda L}$ are not in general nonlocal actions, unless we work over some base ring in which λ is nilpotent.

We will use $*$ for complex conjugation and for the antipode of a Hopf algebra and for the adjoint of an operator and for the anti-involution of a $*$-algebra. The use of the same symbol for all of these is deliberate and indicates that they are all really special cases of a universal "adjoint" or "antipode" operation that acts on everything: whenever two of these operations are defined on something they are equal, so can all be denoted by the same symbol.

The quantum field theories we construct depend on the choice of a cut propagator Δ that is essentially the same as the 2-point Wightman distribution

$$\Delta(\varphi_1, \varphi_2) = \int_{x,y} \langle 0|\varphi_1(x)\varphi_2(y)|0\rangle dx dy$$

Definition 7. A propagator Δ is a continuous bilinear map $\Gamma_c\omega\Phi \times \Gamma_c\omega\Phi \to \mathbb{C}$.

- Δ is called *local* if $\Delta(f, g) = \Delta(g, f)$ whenever the supports of f and g are spacelike separated.
- Δ is called *Feynman* if it is symmetric: $\Delta(f, g) = \Delta(g, f)$.
- Δ is called *Hermitian* if $\Delta^* = \Delta$, where Δ^* is defined by $\Delta^*(f^*, g^*) = \Delta(g, f)^*$ (with a change in order of f and g).
- Δ is called *positive* if $\Delta(f^*, f) \geqslant 0$ for all f.
- Δ is called *cut* if it satisfies the following "positive energy" condition: at each point x of M there is a partial order on the cotangent space defined by a proper closed convex cone C_x, such that if (p, q) is in the wave front set of Δ at some point $(x, y) \in M^2$ then $p \leqslant 0$ and $q \geqslant 0$. Also, as a distribution, Δ can be written in local coordinates as a boundary value of something in the algebra generated by smooth functions and powers and logarithms of polynomials (the boundary values taken so that the wave front sets lie in the regions specified above). Moreover if $x = y$ then $p + q = 0$.

A propagator can also be thought of as a complex distribution on $M \times M$ taking values in the dual of the external tensor product $J\Phi \boxtimes J\Phi$. In particular it has a wave front set (see [Hörmander 1990]) at each point of M^2, which is a cone in the imaginary cotangent space of that point. If A and B are in $\Gamma_c\Phi$, then $\Delta(A, B)$ is defined to be a compactly supported distribution on $M \times M$, defined by $\Delta(A, B)(f, g) = \Delta(Af, Bf)$ for f and g in $\Gamma\omega$.

The key point in the definition of a cut propagator is the condition on the wave front sets, which distinguishes the cut propagators from other propagators such as Feynman propagators or advanced and retarded propagators that can have more complicated wave front sets. For most common cut propagators in Minkowski space, this follows from the fact that their Fourier transforms have support in the positive cone. The condition about being expressible in terms of smooth functions and powers and logs of polynomials is a minor technical condition that is in practice satisfied by almost any reasonable example, and is used in the proof that Feynman measures exist.

If $(p_1, \dots p_n)$ is in the imaginary cotangent space of a point of M^n, then we write $(p_1, \dots p_n) \geqslant 0$ if $p_j \geqslant 0$ for all j, and call it positive if it is not zero.

Example 8. Over Minkowski space, most of the usual cut propagators are positive (except for ghost fields), local, and Hermitian. Most of the ideas for the proof of this can be seen for the simplest case of the propagator for massive Hermitian scalar fields. Using translation invariance, we can write $\Delta(x, y) = \Delta(x - y)$ for some distribution Δ on Minkowski spacetime. Then the Fourier transform of this in momentum space is a rotationally invariant measure supported on one of the two components of vectors with $p^2 = m^2$. This propagator is positive because the measure in momentum space is positive. It satisfies the wave front set part of the cut condition because the Fourier transform has support in the positive cone, and explicit calculation shows that it can be written in terms of powers and logs of polynomials. It satisfies locality because it is invariant under rotations that preserve the direction of time, and under such rotations any space-like vector is conjugate to its negative, so $\Delta(x) = \Delta(-x)$ whenever x is spacelike, in other words $\Delta(x, y) = \Delta(y, x)$ whenever x and y are spacelike separated. The corresponding Feynman propagator is given by $1/(p^2 + m^2 + i\varepsilon)$ where the $i\varepsilon$ indicates in which direction one integrates around the poles, so the cut propagator is just the residue of the Feynman propagator along one of the 2 components of the 2-sheeted hyperboloid $p^2 = m^2$.

For other fields such as spinor fields in Minkowski space, the sheaf of classical fields will usually be some sort of spin bundle. The propagators can often be expressed in terms of the the propagator for a scalar field by acting on it with polynomials in momentum multiplied by Dirac's gamma matrices γ^μ, for example $i(\gamma^\mu p_\mu + m)/(p^2 - m^2)$. Unfortunately there are a bewildering number of different notational and sign conventions for gamma matrices.

Compactly supported actions give functions on the space $\Gamma\Phi$ of smooth fields, by integrating over spacetime M. A Feynman measure is a sort of analogue of Haar measure on a finite-dimensional real vector space. We can think of a Haar measure as an element of the dual of the space of continuous compactly supported functions. For infinite-dimensional vector spaces there are usually not enough continuous compactly supported functions, but instead we can define a measure to be an element of the dual of some other space of functions. We will think of Feynman measures as something like elements of the dual of all functions that are given by free field Gaussians times a compactly supported action. In other words a Feynman measure should assign a complex number to each compactly supported action, formally representing the integral over all fields of this action times a Gaussian e^{iL_F}, where we think of the action as a function of classical fields (or rather sections of the dual of the space of classical fields, which can usually be identified with classical fields). Moreover the Feynman measure should satisfy some sort of analogue of translation invariance.

The space $e^{iL_F}S\Gamma_c\omega SJ\Phi$ is a free rank 1 module over $S\Gamma_c\omega SJ\Phi$ generated by the basis element e^{iL_F}, which can be thought of either as a formal symbol or a formal power series. Its elements can be thought of as representing functions of classical fields that are given by a polynomial times the Gaussian e^{iL_F}, and will be the functions that the Feynman measure is defined on. The symmetric algebra $S\Gamma_c\omega SJ\Phi$ is topologized as the direct sum of the spaces $S^n\Gamma_c\omega SJ\Phi$, each of which is topologized by regarding it as a space of smooth test functions over M^n.

For the definition of a Feynman measure we need to extend the propagator Δ to a larger space as follows. We think of the propagator Δ as a map taking $\Gamma_c J\Phi\otimes\Gamma_c J\Phi$ to distributions on $M\times M$. We then extend it a map from $\Gamma_c SJ\Phi\times\Gamma_c SJ\Phi$ to distributions on $M\times M$ by putting $\Delta(a_1\cdots a_n, b_1\cdots b_n)=\sum_{\sigma\in S_n}\Delta(a_1, b_{\sigma(1)})\times\cdots\times\Delta(a_1, b_{\sigma(n)})$ where the sum is over all elements of the symmetric group S_n (and defining it to be 0 for arguments of different degrees). Finally we extend it to a map from $S^m\Gamma_c SJ\Phi\times S^n\Gamma_c SJ\Phi$ to distributions on $M^m\times M^n$ using the "bicharacter" property: in other words $\Delta(AB, C)=\sum\Delta(A, C')\Delta(B, C'')$ where the coproduct of C is $\sum C'\otimes C''$, and similarly for $\Delta(A, BC)$.

Definition 9. A *Feynman measure* is a continuous linear map

$$\omega: e^{iL_F}S\Gamma_c\omega SJ\Phi\to\mathbb{C}.$$

The Feynman measure is said to be associated with the propagator Δ if it satisfies the following conditions:

- Smoothness on the diagonal: Whenever $(p_1,\ldots,p_n)$ is in the wave front set of ω at the point $(x,\ldots,x)$ on the diagonal, then $p_1+\cdots+p_n=0$
- Nondegeneracy: there is a smooth nowhere vanishing function g such that $\omega(e^{iL_F}v)$ is $\int_M gv$ for v in $\Gamma_c\omega S^0 J\Phi=\Gamma_c\omega$.
- Gaussian condition, or weak translation invariance: For $A\in S^m\Gamma_c\omega SJ\Phi$, $B\in S^n\Gamma_c\omega SJ\Phi$, with both sides interpreted as distributions on M^{m+n}, we

have

$$\omega(AB) = \sum \omega(A')\Delta(A'', B'')\omega(B')$$

whenever there is no element in the support of A that is $\leqslant$ some element of the support of B. Here $\sum A' \otimes A'' \in S\Gamma_c\omega SJ\Phi \otimes S\Gamma_c SJ\Phi$ is the image of A under the map $S^m\Gamma_c\omega SJ\Phi \to S^m\Gamma_c\omega SJ\Phi \otimes S^m\Gamma_c SJ\Phi$ induced by the coaction $\omega SJ\Phi \to \omega SJ\Phi \otimes SJ\Phi$ of $SJ\Phi$ on $\omega SJ\Phi$, and similarly for B. The product on the right is a product of distributions, using the extended version of Δ defined just before this definition.

We explain what is going on in this definition. We would like to define the value of the Feynman measure to be a sum over Feynman diagrams, formed by joining up pairs of fields in all possible ways by lines, and then assigning a propagator to each line and taking the product of all propagators of a diagram. This does not work because of ultraviolet divergences: products of propagators need not be defined when points coincide. If these products were defined then they would satisfy the Gaussian condition, which then says roughly that if the vertices are divided into two disjoint subsets a and b, then a Feynman diagram can be divided into a subdiagram with vertices a, a subdiagram with vertices b, and some lines between a and b. The value $\omega(AB)$ of the Feynman diagram would then be the product of its value $\omega(A')$ on a, the product $\Delta(A'', B'')$ of all the propagators of lines joining a and b, and its value $\omega(B')$ on b. The Gaussian condition need not make sense if some point of a is equal to some point of b because if these points are joined by a line then the corresponding propagator may have a bad singularity, but does make sense whenever all points of a are not $\leq$ all points of b. The definition above says that a Feynman measure should at least satisfy the Gaussian condition in this case, when the product is well defined.

Unfortunately the standard notation ω for a dualizing sheaf, such as the sheaf of densities, is the same as the standard notation ω for a state in the theory of operator algebras, which the Feynman measure will be a special case of. It should be clear from the context which meaning of ω is intended.

If ω is a Feynman measure and $A \in e^{iL_F} S^n\Gamma_c\omega SJ\Phi$ then $\omega(A)$ is a complex number, but can also be considered as the compactly supported density on M^n taking a smooth f to $\omega(A)(f) = \omega(Af)$. The integral of this density $\omega(A)$ over spacetime is just the complex number $\omega(A)$.

Since $e^{iL_F} S\Gamma_c\omega SJ\Phi$ is a coalgebra (where elements of $\Gamma_c\omega SJ\Phi$ are primitive and e^{iL_F} is grouplike), the space of Feynman measures is an algebra, whose product is called convolution.

The nondegeneracy condition just excludes some uninteresting degenerate cases, such as the measure that is identically zero, and the function g appearing in it is usually normalized to be 1. The condition about smoothness on the diagonal implies that the product on the right in the Gaussian condition is defined. This is because ω has the property that if an element $(p_1, \dots p_n)$ of the wave front set of some point is nonzero then its components cannot all be positive and cannot all be negative. This shows that the wave front sets are such that the product of distributions is defined.

If A is in $e^{iL_F} S\Gamma_c \omega SJ\Phi$, then $\omega(A)$ can be thought of as a Feynman integral

$$\omega(A) = \int A(\varphi)\mathfrak{D}\varphi,$$

where L_F is a quadratic action with cut propagator Δ, and where A is considered to be a function of fields φ. The integral is formally an integral over all classical fields. The Gaussian condition is a weak form of translation invariance of this measure under addition of classical fields. Formally, translation invariance is equivalent to the Gaussian condition with the condition about supports omitted and cut propagators replaced by Feynman propagators, but this is not well defined because the Feynman propagators can have such bad singularities that their products are sometimes not defined when two spacetime points coincide.

The Feynman propagator Δ_F of a Feynman measure ω is defined to be the restriction of ω to $\Gamma_c\omega\Phi \times \Gamma_c\omega\Phi$. It is equal to the cut propagator at "time-ordered" points $(x, y) \in M^2$ where $x \not\leqslant y$, but will usually differ if $x \leqslant y$. As it is symmetric, it is determined by the cut propagator except on the diagonal of $M \times M$. Unlike cut propagators, Feynman propagators may have singularities on the diagonal whose wave front sets are not contained in a proper cone, so that their products need not be defined.

Any symmetric algebra SX over a module X has a natural structure of a commutative and cocommutative Hopf algebra, with the coproduct defined by making all elements of X primitive (in other words, $\Delta x = x \otimes 1 + 1 \otimes x$ for $x \in X$). In other words, SX is the coordinate ring of a commutative affine group scheme whose points form the dual of X under addition. For general results about Hopf algebras see [Abe 1980]. Similarly $SJ\Phi$ is a sheaf of commutative cocommutative Hopf algebras, with a coaction on itself and the trivial coaction on ω, and so has a coaction on $S\omega SJ\Phi$, preserving the coproduct of $S\omega SJ\Phi$. It corresponds to the sheaf of commutative affine algebraic groups whose points correspond to the sheaf $J\Phi$ under addition.

Definition 10. A renormalization is an automorphism of $S\omega SJ\Phi$ preserving its coproduct and the coaction of $SJ\Phi$. The group of renormalizations is called the ultraviolet group.

The justification for this rather mysterious definition is Theorem 15, which shows that renormalizations act simply transitively on the Feynman measures associated to a given local cut propagator. In other words, although there is no canonical

Feynman measure on the space of classical fields, there is a canonical orbit of such measures under renormalization.

More generally, renormalizations are global sections of the sheaf of renormalizations (defined in the obvious way), but we will make no use of this viewpoint.

The (infinite-dimensional) ultraviolet group really ought to be called the "renormalization group", but unfortunately this name is already used for a quite different 1-dimensional group. The "renormalization group" is the group of positive real numbers, together with an action on Lagrangians by "renormalization group flow". The relation between the renormalization group and the ultraviolet group is that the renormalization group flow can be thought of as a nonabelian 1-cocycle of the renormalization group with values in the ultraviolet group, using the action of renormalizations on Lagrangians that will be constructed later.

The ultraviolet group is indirectly related to the Hopf algebras of Feynman diagrams introduced in [Kreimer 1998] and applied to renormalization in [Connes and Kreimer 2000], though this relation is not that easy to describe. First of all their Hopf algebras correspond to Lie algebras, and the ultraviolet group has a Lie algebra, and these two Lie algebras are related. There is no direct relation between Connes and Kreimer's Lie algebras and the Lie algebra of the ultraviolet group, in the sense that there seems to be no natural homomorphism in either direction. However there seems to be a sort of intermediate Lie algebra that has homomorphisms to both. This intermediate Lie algebra (or group) can be defined using Feynman diagrams decorated with smooth test functions rather than the sheaf $S\omega SJ\Phi$ used here. Unfortunately all my attempts to explain the product of this Lie algebra explicitly have resulted in an almost incomprehensible combinatorial mess so complicated that it is unusable. Roughly speaking, the main differences between the ultraviolet group and the intermediate Lie algebra is that the Lie algebra of the ultraviolet group amalgamates all Feynman diagrams with the same vertices while the intermediate Lie algebra keeps track of individual Feynman diagrams, and the main difference between the intermediate Lie algebra and Kreimer's algebra is that the intermediate Lie algebra is much fatter than Kreimer's algebra because it has infinite-dimensional spaces of smooth functions in it. In some sense Kreimer's algebra could be thought of as a sort of skeleton of the intermediate Lie algebra.

All reasonable Feynman measures for a given free field theory are equivalent up to renormalization, but it is not easy to show that at least one exists. We do this by following the usual method of constructing a perturbative quantum field theory in physics. We first regularize the cut local propagator which produces a meromorphic family of Feynman measures, following Etingof [1999, pages 597–607] in using Bernstein's theorem [1972] on the analytic continuation of powers of a polynomial to construct the regularization. We then use an infinite renormalization to eliminate the poles of the regularized Feynman measure in order of their complexity.

A quantum field theory satisfying the Wightman axioms [Streater and Wightman 2000, §3.1] is determined by its Wightman distributions, which are given by linear maps $\omega_n : T^n\Gamma_c\omega\Phi \to \mathbb{C}$ from the tensor powers of the space of test functions for each n. We will follow H. J. Borchers [1962] in combining the Wightman distributions into a Wightman functional $\omega : T\Gamma_c\omega\Phi \to \mathbb{C}$ on the tensor algebra $T\Gamma_c\omega\Phi$ of the space $\Gamma_c\omega\Phi$ of test functions (which is sometimes called a Borchers algebra or Borchers-Uhlmann algebra or BU-algebra). In order to accommodate composite operators we extend the algebra $T\Gamma_c\omega\Phi$ to the larger algebra $T\Gamma_c\omega SJ\Phi$, and to accommodate time ordered operators we extend it further to $\mathrm{TS}\Gamma_c\omega SJ\Phi$. In this set up it is clear how to accommodate perturbative quantum field theories: we just allow ω to take values in a space of formal power series $\mathbb{C}[\![\boldsymbol{\lambda}]\!] = \mathbb{C}[\![\lambda_1, \lambda_2, \dots]\!]$ rather than $\mathbb{C}$. For regularization ω sometimes takes values in a ring of meromorphic functions. There is one additional change we need: it turns out that the elements of $\Gamma_c\omega SJ\Phi$ do not really represent operators on a space of physical states, but are better thought of as operators that map a space of incoming states to a space of outgoing states, and vice versa. If we identify the space of incoming states with the space of physical states, this means that only products of an even number of operators of $S\Gamma_c\omega SJ\Phi$ act on the space of physical states. So the functional defining a quantum field theory is really defined on the subalgebra $T_0S\Gamma_c\omega SJ\Phi$ of even degree elements.

So the main goal of this paper is to construct a linear map from $T_0S\Gamma_c\omega SJ\Phi$ to $\mathbb{C}[\boldsymbol{\lambda}]$ from a given Lagrangian, and to check that it satisfies analogues of the Wightman axioms.

The space of physical states of the quantum field theory can be reconstructed from ω as follows.

Definition 11. Let $\omega : T \to C$ be an $\mathbb{R}$-linear map between real $*$-algebras.

- ω is called Hermitian if $\omega^* = \omega$, where $\omega^*(a^*) = \omega(a)^*$
- ω is called positive if it maps positive elements to positive elements, where an element of a $*$-algebra is called positive if it is a finite sum of elements of the form a^*a.
- ω is called a state if it is positive and normalized by $\omega(1) = 1$
- The left, right, or 2-sided kernel of ω is the largest left, right or 2-sided ideal closed under * on which ω vanishes.
- The space of physical states of ω is the quotient of T by the left kernel of ω. Its sesquilinear form is $\langle a, b\rangle = \omega(a^*b)$, and its vacuum vector is the image of 1.
- The algebra of physical operators of ω is the quotient of T by the 2-sided kernel of ω.

The algebra of physical operators is a $*$-algebra of operators with a left action on the physical states. If ω is positive or Hermitian then so is the sesquilinear form $\langle\ ,\ \rangle$. When ω is Hermitian and positive and C is the complex numbers the left kernel of ω is the set of vectors a with $\omega(a^*a)=0$, and the definition of the space of physical states is essentially the GNS construction and is also the main step of the Wightman reconstruction theorem. In this case the completion of the space of physical states is a Hilbert space.

The maps ω we construct are defined on the real vector space $T_0S\Gamma_c\omega SJ\Phi$ and will initially be $\mathbb{R}$-linear. It is often convenient to extend them to be $\mathbb{C}[[\boldsymbol{\lambda}]]$-linear maps defined on $T_0S\Gamma_c\omega SJ\Phi\otimes\mathbb{C}[[\boldsymbol{\lambda}]]$, in which case the corresponding space of physical states will be a module over $\mathbb{C}[[\boldsymbol{\lambda}]]$ and its bilinear form will be sesquilinear over $\mathbb{C}[[\boldsymbol{\lambda}]]$.

The machinery of renormalization and regularization has little to do with perturbation theory or the choice of Lagrangian: instead, it is needed even for the construction of free field theories if we want to include composite operators. The payoff for all the extra work needed to construct the composite operators in a free field theory comes when we construct interacting field theories from free ones. The idea for constructing an interacting field theory from a free one is simple: we just apply a suitable automorphism (or endomorphism) of the algebra $T_0S\Gamma_c\omega SJ\Phi$ to the free field state ω to get a state for an interacting field. For example, if we apply an endomorphism of the sheaf $\omega SJ\Phi$ then we get the usual field theories of normal ordered products of operators, which are not regarded as all that interesting. For any Lagrangian L there is an infinitesimal automorphism of $T_0S\Gamma_c\omega SJ\Phi$ that just multiplies elements of $S\Gamma_c\omega SJ\Phi$ by iL, which we would like to lift to an automorphism e^{iL}. The construction of an interacting quantum field theory from a Feynman measure ω and a Lagrangian L is then given by the natural action $e^{-iL}\omega$ of the automorphism e^{-iL} on the state ω. The problem is that e^{iL_I} is only defined if the interaction Lagrangian has infinitesimal coefficients, due to the fact that we only defined ω on polynomials times a Gaussian, so this construction only produces perturbative quantum field theories taking values in rings of formal power series. This is essentially the problem of lifting a Lie algebra elements L_I to a group element e^{iL_I}, which is trivial for operators on finite-dimensional vector spaces, but a subtle and hard problem for unbounded operators such as L_I that are not self adjoint. This construction works provided the interacting part of the Lagrangian not only has infinitesimal coefficients but also has compact support. We show that the more general case of Lagrangians without compact support can be reduced to the case of compact support up to inner automorphisms, at least on globally hyperbolic spacetimes, by showing that infrared divergences cancel.

Up to isomorphism, the quantum field theory does not depend on the choice of a Feynman measure or Lagrangian, but only on the choice of a propagator. In particular, the interacting quantum field theory is isomorphic to a free one. This does not mean that interacting quantum field theories are trivial, because this isomorphism does not preserve the subspace of simple operators, so if one only looks at the restriction to simple operators, as in the Wightman axioms, one no longer gets an isomorphism between free and interacting theories. The difference between interacting and free field theories is that one chooses a different set of operators to be the "simple" operators corresponding to physical fields.

The ultraviolet group also has a nonlinear action on the space of infinitesimal Lagrangians. A quantum field theory is determined by the choice of a Lagrangian and a Feynman measure, and this quantum field theory is unchanged if the Feynman measure and the Lagrangian are acted on by the same renormalization. This shows why the choice of Feynman measure is not that important: if one chooses a different Feynman measure, it is the image of the first by a unique renormalization, and by applying this renormalization to the Lagrangian one still gets the same quantum field theory.

Roughly speaking, we show that these quantum field theories $e^{iL_I}\omega$ satisfy the obvious generalizations of Wightman axioms whenever it is reasonable to expect them to do so. For example, we will show that locality holds by showing that the state vanishes on the "locality ideal" of Definition 32, the quantum field theory is Hermitian if we start with Hermitian cut propagators and Lagrangians, and we get a (positive) state if we start with a positive (non-ghost) cut propagator. We cannot expect to get Lorentz invariant theories in general as we are working over a curved spacetime, but if we work over Minkowski space and choose Lorentz invariant cut propagators then we get Lorentz invariant free quantum field theories. In the case of interacting theories Lorentz invariance is more subtle, even if the Lagrangian is Lorentz invariant. Lorentz invariance depends on the cancellation of infrared divergences as we have to approximate the Lorentz invariant Lagrangian by non-Lorentz-invariant Lagrangians with compact support, and we can only show that infrared divergences cancel up to inner automorphisms. This allows for the possibility that the vacuum is not Lorentz invariant, in other words Lorentz invariance may be spontaneously broken by infrared divergences, at least if the theory has massless particles. (It seems likely that if there are no massless particles then infrared divergences cancel and we recover Lorentz invariance, but I have not checked this in detail.)

In the final section we discuss anomalies. Fujikawa [1979] observed that anomalies arise from the lack of invariance of Feynman measures under a symmetry group, and we translate his observation into mathematical language.

The definitions above generalize to the relative case where spacetime is replaced by a morphism $X \to Y$, whose fibers can be thought of as spacetimes parametrized by Y. For example, the sheaf of densities ω is replaced by the dualizing sheaf or complex $\omega_{X/Y}$. We will make no serious use of this generalization, though the section on regularization could be thought of as an example of this where Y is the spectrum of a ring of meromorphic functions.

2. The ultraviolet group

We describe the structure of the ultraviolet group, and show that it acts simply transitively on the Feynman measures associated with a given propagator.

Theorem 12. *The map taking a renormalization $\rho : S\omega SJ\Phi \to S\omega SJ\Phi$ to its composition with the natural map $S\omega SJ\Phi \to S^1\omega S^0 J\Phi = \omega$ identifies renormalizations with the elements of* $\mathrm{Hom}(S\omega SJ\Phi, \omega)$ *that vanish on $S^0\omega SJ\Phi$ and that are isomorphisms when restricted to $\omega = S^1\omega S^0 J\Phi$.*

Proof. This is a variation of the dual of the fact that endomorphisms ρ of a polynomial ring $R[x]$ correspond to polynomials $\rho(x)$, given by the image of the polynomial x under the endomorphism ρ. It is easier to understand the dual result first, so suppose that C is a cocommutative Hopf algebra and ω is a vector space (with C acting trivially on ω). Then the symmetric algebra $S\omega C = S(\omega \otimes C)$ is a commutative algebra acted on by C, and its endomorphisms (as a commutative algebra) correspond exactly to elements of $\mathrm{Hom}(\omega, S\omega C)$ because any such map lifts uniquely to a C-invariant map from ω to ωC, which in turn lifts to a unique algebra homomorphism from $S\omega C$ to itself by the universal property of symmetric algebras. This endomorphism is invertible if and only if the map from ω to $\omega = S^1\omega C^0$ is invertible, where C^0 is the vector space generated by the identity of C.

To prove the theorem, we just take the dual of this result, with C now given by $SJ\Phi$. There is one small modification we need to make in taking the dual result: we need to add the condition that the element of $\mathrm{Hom}(S\omega C, \omega)$ vanishes on $S^0\omega C$ in order to get an endomorphism of $S\omega C$; this is related to the fact that endomorphisms of the polynomial ring $R[x]$ correspond to polynomials, but continuous endomorphisms of the power series ring $R[\![x]\!]$ correspond to power series with vanishing constant term. $\square$

The ultraviolet group preserves the increasing filtration $S^{\leqslant m}\omega SJ\Phi$ and so has a natural decreasing filtration by the groups $G_{\geqslant n}$, consisting of the renormalizations that fix all elements of $S^{\leqslant n}\omega SJ\Phi$. The group $G = G_{\geqslant 0}$ is the inverse limit of the groups $G/G_{\geqslant n}$, and the commutator of $G_{\geqslant m}$ and $G_{\geqslant n}$ is contained in $G_{\geqslant m+n}$, so in particular $G_{\geqslant 1}$ is an inverse limit of nilpotent groups $G_{\geqslant 1}/G_{\geqslant n}$. The group $G_{\geqslant n}$ is a semidirect product $G_{\geqslant n+1}G_n$ of its normal subgroup $G_{\geqslant n+1}$ with the group G_n, consisting of elements represented by elements of $\mathrm{Hom}(S\omega SJ\Phi, \omega)$ that are the identity on $S^1\omega SJ\Phi$ if $n > 0$, and vanish on $S^m\omega SJ\Phi$ for $m > 1$, $m \neq n+1$.

Lemma 13. *The group G is $\dots G_2G_1G_0$ in the sense that any element of G can be written uniquely as an infinite product $\dots g_2g_1g_0$ with $g_i \in G_i$, and conversely any such infinite product converges to an element of G.*

Proof. The convergence of this product follows from the facts that all elements g_i preserve any space $S^{\leqslant m}\omega SJ\Phi$, and all but a finite number act trivially on it. The fact that any element can be written uniquely as such an infinite product follows from the fact that $G/G_{\geqslant n}$ is essentially the product $G_{n-1}\dots G_2G_1G_0$. □

The natural map

$$S\Gamma\omega SJ\Phi \to \Gamma S\omega SJ\Phi$$

is not an isomorphism, because on the left the symmetric algebra is taken over the reals, while on the right it is essentially taken over smooth functions on M.

Lemma 14. *The action of renormalizations on $\Gamma S\omega SJ\Phi$ lifts to an action on $S\Gamma_c\omega SJ\Phi$ that preserves the coproduct, the coaction of $\Gamma SJ\Phi$, and the product of elements with disjoint support.*

Proof. A renormalization is given by a linear map from $\Gamma_c S\omega SJ\Phi$ to $\Gamma_c\omega$, which by composition with the map $S\Gamma_c\omega SJ\Phi \to \Gamma S\omega SJ\Phi$ and the "integration over M" map $\Gamma_c\omega \to \mathbb{R}$ lifts to a linear map from $S\Gamma_c\omega SJ\Phi$ to $\mathbb{R}$. This linear map has the special property that the product of any two elements with disjoint support vanishes, because it is multilinear over the ring of smooth functions. As in Theorem 12, the linear map gives an automorphism of $S\Gamma_c\omega SJ\Phi$ preserving the coproduct and the coaction of $\Gamma SJ\Phi$. As the linear map vanishes on products of disjoint support, the corresponding renormalization preserves products of elements with disjoint support. □

In general, renormalizations do not preserve products of elements of $S\Gamma_c\omega SJ\Phi$ that do not have disjoint support; the ones that do are those in the subgroup G_0.

Theorem 15. *The group of complex renormalizations acts simply transitively on the Feynman measures associated with a given cut local propagator.*

Proof. We first show that renormalizations ρ act on Feynman measures ω associated with a given local cut propagator. We have to show that renormalizations preserve nondegeneracy, smoothness on the diagonal, and the Gaussian property. The first two of these are easy to check, because the value of $\rho(\omega)$ on any element is given by a finite sum of values of ω on other elements, so is smooth along the diagonal.

To check that renormalizations preserve the Gaussian property

$$\omega(AB) = \sum \omega(A')\Delta(A'', B'')\omega(B')$$

we recall that renormalizations ρ preserve products with disjoint support and also commute with the coaction of $SJ\Phi$. Since A and B have disjoint supports we have $\rho(AB) = \rho(A)\rho(B)$. Since ρ commutes with the coaction of $SJ\Phi$, the image of $\rho(A)$ under the coaction of $SJ\Phi$ is $\sum \rho(A') \otimes A''$, and similarly for B. Combining these facts with the Gaussian property for $\rho(A)\rho(B)$ shows that

$$\omega(\rho(AB)) = \sum \omega(\rho(A'))\Delta(A'', B'')\omega(\rho(B'));$$

in other words, the renormalization ρ preserves the Gaussian property.

To finish the proof, we have to show that for any two normalized smooth Feynman measures ω and ω' with the same cut local propagator, there is a unique complex renormalization g taking ω to ω'. We will construct $g = \dots g_2 g_1 g_0$ as an infinite product, with the property that $g_{n-1} \dots g_0 \omega$ coincides with ω' on $e^{iL_F} S^{\leqslant n} \Gamma_c \omega SJ\Phi$. Suppose that $g_0, \dots, g_{n-1}$ have already been constructed. By changing ω to $g_{n-1} \dots g_0 \omega$ we may as well assume that they are all 1, and that ω and ω' coincide on $e^{iL_F} S^{\leqslant n} \Gamma_c \omega SJ\Phi$. We have to show that there is a unique $g_n \in G_n$ such that $g_n \omega$ and ω' coincide on $e^{L_F} S^{n+1} \Gamma_c \omega SJ\Phi$.

The difference $\omega - \omega'$, restricted to $e^{iL_F} S^{n+1} \Gamma_c \omega SJ\Phi$, is a continuous linear function on $e^{iL_F} S^{n+1} \Gamma_c \omega SJ\Phi$, which we think of as a distribution. Moreover, since both ω and ω' are determined off the diagonal by their values on elements of smaller degree by the Gaussian property, this distribution has support on the diagonal of M^{n+1}. Since ω and ω' both have the property that their wave front sets on the diagonal are orthogonal to the diagonal, the same is true of their difference $\omega - \omega'$, so the distribution is given by a map $e^{iL_F} S^{n+1} \Gamma_c \omega SJ\Phi \to \omega$. By Theorem 12 this corresponds to some renormalization $g_n \in G_n$, which is the unique element of G_n such that $g_n \omega$ and ω' coincide on $e^{iL_F} S^{n+1} \Gamma_c \omega SJ\Phi$. □

3. Existence of Feynman measures

We now show (see Theorem 21) the existence of at least one Feynman measure associated to any cut local propagator, by using regularization and renormalization. Regularization means that we construct a Feynman measure over a field of meromorphic functions, which will usually have poles at the point we are interested in, and renormalization means that we eliminate these poles by acting with a suitable meromorphic renormalization.

Lemma 16. *If $f_1, \dots, f_m$ are polynomials in several variables, then there are nonzero* (*Bernstein-Sato*) *polynomials b_i and differential operators D_i such that*

$$b_i(s_1, \dots, s_m) f_1(z)^{s_1} \dots f_m(z)^{s_m} = D_i(z) \left(f_i(z) f_1(z)^{s_1} \dots f_m(z)^{s_m} \right).$$

Proof. Bernstein's proof [1972] of this theorem for the case $m = 1$ also works for any m after making the obvious minor changes, such as replacing the field of rational functions in one variable s_1 by the field of rational functions in m variables. □

Corollary 17. *If $f_1, \ldots, f_m$ are polynomials in several variables then for any choice of continuous branches of the multivalued functions, $f_1(z)^{s_1} \ldots f_m(z)^{s_m}$ can be analytically continued from the region where all s_j have positive real part to a meromorphic distribution-valued function for all complex values of $s_1, \ldots, s_m$.*

Proof. This follows by using the functional equation of Lemma 16 to repeatedly decrease each s_j by 1, just as in Bernstein's proof for the case $m = 1$. □

Theorem 18. *Any cut local propagator Δ has a regularization, in other words a Feynman measure with values in a ring of meromorphic functions whose cut propagator at some point is Δ.*

Proof. The following argument is inspired by the one in [Etingof 1999]. By using a locally finite smooth partition of unity, which exists since we assume that spacetime is metrizable, we can reduce to showing that a regularization exists locally. If a local propagator is smooth, it is easy to construct a Feynman measure for it, just by defining it as a sum of products of Feynman propagators. Now suppose we have a meromorphic family of local propagators Δ_d depending on real numbers d_i, given in local coordinates by a finite sum of boundary values of terms of the form

$$s(x, y)p_1(x, y)^{d_1} \ldots p_k(x, y)^{d_k} \log(p_{k+1}(x, y)) \ldots$$

where s is smooth in x and y, and the p_i are polynomials, and where we choose some branch of the powers and logarithms in each region where they are nonzero. In this case the Feynman measure can also be defined as a meromorphic function of d for all real d. To prove this, we can forget about the smooth function s as it is harmless, and we can eliminate the logarithmic terms by writing $\log(p)$ as dp^t/dt at $t = 0$. For any fixed number of fields with derivatives of fixed order, the corresponding distribution is defined when all variables d_i have sufficiently large real part, because the product of the propagators is smooth enough to be defined in this case. But this distribution is given in local coordinates by the product the d_i'th powers of polynomials of x and y. By Bernstein's Corollary 17 these products can be continued as a meromorphic distribution-valued function of the d_i to all complex d_i.

This gives a Feynman measure with values in the field of meromorphic functions in several variables, and by restricting functions to the diagonal we get a Feynman measure whose value are meromorphic functions in one variable. □

Example 19 (dimensional regularization). Over Minkowski space of dimension d, there is a variation of the construction of a meromorphic Feynman measure, which is very similar to dimensional regularization. In dimensional regularization, one formally varies the dimension of spacetime, to get Feynman diagrams that are meromorphic functions of the dimension of spacetime. One way to make sense out of this is to keep the dimension of spacetime fixed, but vary the propagator of the free field theory, by considering it to be a meromorphic function of a complex number d. The propagator for a Hermitian scalar field, considered as a distribution of z in Minkowski space, can be written as a linear combination of functions of the form

$$\frac{K_{d/2-1}\,(c\sqrt{(z,z)}\,)}{\sqrt{(z,z)}^{d/2-1}},$$

where $K_\nu(z)$ is a multivalued modified Bessel function of the third kind, and where we take a suitable choice of branch (depending on whether we are considering a cut or a Feynman propagator). A similar argument using Bernstein's theorem shows that this gives a Feynman measure that is analytic in d for d with large real part and that can be analytically continued as a meromorphic function to all complex d. This gives an explicit example of a meromorphic Feynman measures for the usual propagators in Minkowski space.

Theorem 20. *Any meromorphic Feynman measure can be made holomorphic by acting on it with a meromorphic renormalization.*

Proof. This is essentially the result that a bare quantum field theory can be made finite by an infinite renormalization. Suppose that ω is a meromorphic Feynman measure. Using the same idea as in Theorem 15 we will construct a meromorphic renormalization $g = \dots g_2 g_1 g_0$ as an infinite product, but this time we choose $g_n \in G_n$ to kill the singularities of order $n+1$. The key point is to prove that these lowest order singularities are "local", meaning that they have support on the diagonal. (In the special case of translation-invariant theories on Minkowski spacetime this becomes the usual condition that they are "polynomials in momentum", or more precisely that their Fourier transforms are essentially polynomials in momentum on the subspace with total momentum zero). The locality follows from the Gaussian property of ω, which determines ω at each order in terms of smaller orders except on the diagonal. In particular if ω is nonsingular at all orders at most n, then the singular parts of the order $n+1$ terms all have support on the diagonal. Since the difference is smooth along the diagonal, we can find some $g_n \in G_n$ that kills off the order $n+1$ singularities, as in Theorem 15. Since renormalizations preserve the Gaussian property we can keep on repeating this indefinitely, killing off the singularities in order of their order. □

The famous problem of "overlapping divergences" is that the counterterms for individual Feynman diagrams used for renormalization sometimes contain non-polynomial (logarithmic) terms in the momentum, which bring renormalization to a halt unless they miraculously cancel when summed over all Feynman diagrams.

This problem is avoided in the proof above because by using the ultraviolet group we only need to handle the divergences of lowest order at each step, where it is easy to see that the logarithmic terms cancel.

Theorem 21. *Any cut local propagator has an associated Feynman measure.*

Proof. This follows from Theorem 18, which uses regularization to show that there is a meromorphic Feynman measure, and Theorem 20 which uses renormalization to show that the poles of this can be eliminated. □

4. Subgroups of the ultraviolet group

There are many additional desirable properties that one can impose on Feynman measures, such as being Hermitian, or Lorentz invariant, or normal ordered, and there is often a subgroup of the ultraviolet group that acts transitively on the measures with the given property. We give several examples of this.

Example 22. A Feynman measure can be normalized so that on $S^1\Gamma_c\omega S^0 J\Phi = \Gamma_c\omega$ its value is given by integrating over spacetime (in other words $g = 1$ in Definition 9), by acting on it by a unique element of the ultraviolet group consisting of renormalizations in G_0 that are trivial on $\omega S^{>0}J\Phi$. This group can be identified with the group of nowhere vanishing smooth complex functions on spacetime. The complementary normal subgroup of the ultraviolet group consists of the renormalizations that fix all elements of $\omega S^0 J\Phi = \omega$, and this acts simply transitively on the normalized Feynman measures. In practice almost any natural Feynman measure one constructs is normalized.

Example 23 (normal ordering). In terms of Feynman diagrams, "normal ordering" means roughly that Feynman diagrams with an edge from a vertex to itself are discarded. We say that a Feynman measure is normally ordered if it vanishes on $\Gamma_c\omega S^{>0}J\Phi$. Informally, $\omega S^{>0}J\Phi$ corresponds to Feynman diagrams with just one point and edges from this point to itself. We will say that a renormalization is normally ordered if it fixes all elements of $\omega S^{>0}J\Phi$. The subgroup of normally ordered renormalizations acts transitively on the normally ordered Feynman measures. The group of all renormalizations is the semidirect product of its normal subgroup $G_{>0}$ of normally-ordered renormalizations with the subgroup G_0 preserving all products. For any renormalization, there is a unique element of G_0 that takes it to a normally ordered renormalization. The Feynman measures constructed by regularization (in particular those constructed by dimensional regularization) are usually normally ordered if the spacetime has positive dimension, but are usually not for 0-dimensional spacetimes. This is because the propagators tend to contain a factor such as $(x - y)^{-2d}$ which vanishes for large $-d$ when $x = y$, and so vanishes on Feynman diagrams with just one point for all d by analytic continuation.

So for most purposes we can restrict to normally-ordered Feynman measures and normally-ordered renormalizations, at least for spacetimes of positive dimension.

Example 24 (normalization of Feynman propagators). In general a renormalization fixes the cut propagator but can change the Feynman propagator, by adding a distribution with support on the diagonal. However there is often a canonical choice of Feynman propagator: the one with a singularity on the diagonal of smallest possible order, which will often also be a Green function for some differential operator. We can add the condition that the Feynman propagator of a Feynman measure should be this canonical choice; the subgroup of renormalizations fixing the Feynman propagator, consisting of renormalizations fixing $S^2\omega J\Phi$, acts simply transitively on these Feynman measures.

Example 25. [simple operators] More generally, there is a subgroup consisting of renormalizations ρ such that $\rho(aB) = \rho(a)\rho(B)$ whenever a is simple (involving only one field), but where B is arbitrary. This stronger condition is useful because it says (roughly) that simple operators containing only one field do not get renormalized; see the discussion in Section 6. We can find a set of Feynman measures acted on simply transitively by this group by adding the condition that

$$\omega(aB) = \sum \Delta_F(aB_1)\omega(B_2)$$

whenever a is simple and $\sum B_1 \otimes B_2$ is the coproduct of B. This relation holds whenever a and B have disjoint supports by definition of a Feynman measure, so the extra condition says that it also holds even when they have overlapping supports. The key point is that the product of distributions above is always defined because any nonzero element of the wave front set of Δ_F is of the form $(p, -p)$. This would not necessarily be true if a were not simple because we would get products of more than 1 Feynman propagator whose singularities might interfere with each other. In terms of Feynman diagrams, this says that vertices with just one edge are harmless: more precisely, with this normalization, adding a vertex with just one edge to a Feynman diagram has the effect of multiplying its value by the Feynman propagator of the edge. As this condition extends the Gaussian property to more Feynman diagrams, it can also be thought of as a strengthening of the translation invariance property of the Feynman measure.

Example 26 (Dyson condition). Classically, Lagrangians were called renormalizable if all their coupling constants have nonnegative mass dimension. The filtration on Lagrangian densities by mass dimension induces a similar filtration on Feynman measures and renormalizations. The Feynman measures of mass dimension $\leqslant 0$ are acted on simply transitively by the renormalizations of mass dimension $\leqslant 0$. This is useful, because the renormalizations of mass dimension at most 0 act on the spaces of Lagrangian densities of mass dimension at most 0, and these often form finite-dimensional spaces, at least if some other symmetry conditions such as Lorentz invariance are added. For example, in dimension 4 the density has dimension -4, so the (Lorentz-invariant) terms of the Lagrangian density of mass dimension at most 0 are given by (Lorentz invariant) terms of the Lagrangian of mass dimension at most 4, such as φ^4, φ^2, $\partial\varphi\partial\varphi$, and so on: the usual Lorentz-invariant even terms whose coupling constants have mass dimension at least 0. For example, we get a three-dimensional space of theories of the form $\lambda\varphi^4 + m\varphi^2 + z\partial\varphi\partial\varphi$ in this way, giving the usual φ^4 theory in 4 dimensions.

Example 27 (boundary terms). The Feynman measures constructed in Section 3 have the property that they vanish on "boundary terms". This means that we quotient the space of local Lagrangians $\Gamma_c \omega S J \Phi$ by its image under the action of smooth vector fields such as $\partial/\partial x_i$, or in other words we replace a spaces of n-forms by the corresponding de Rham cohomology group. These measures are acted on simply transitively by renormalizations corresponding to maps that vanish on boundary terms. This is useful in gauge theory, because some symmetries such as the BRST symmetry are only symmetries up to boundary terms.

Example 28 (symmetry invariance). Given a group (or Lie algebra) G such as a gauge group acting on the sheaf Φ of classical fields and preserving a given cut propagator, the subgroup of G-invariant renormalizations acts simply transitively on the G-invariant Feynman measures with given cut propagator. In general there need not exist any G-invariant Feynman measure associated with a given cut local propagator, though if there is then G-invariant Lagrangians lead to G-invariant quantum field theories. The obstructions to finding a G-invariant measure are cohomology classes called anomalies, and are discussed further in Section 7.

Example 29 (Lorentz invariance). An important case of invariance under symmetry is that of Poincare invariance for flat Minkowski space. In this case the spacetime M is Minkowski space, the Lie algebra G is that of the Poincare group of spacetime translations and Lorentz rotations, and the cut propagator is one of the standard ones for free field theories of fields of finite spin. Then dimensional regularization is invariant under G, so we get a Feynman measure invariant under the Poincare group, and in particular there are no anomalies for the Poincare algebra. The elements of the ultraviolet group that are Poincare invariant act simply transitively on the Feynman measures for this propagator that are Poincare invariant. If we pick any such measure, then we get a map from invariant Lagrangians to invariant quantum field theories.

Example 30 (Hermitian conditions). The group of complex renormalizations has a real form, consisting of the subgroup of (real) renormalizations. This acts simply transitively on the Hermitian Feynman measures associated with a given cut local propagator. The Hermitian Feynman measures (or propagators) are not the real-valued ones, but satisfy a more complicated Hermitian condition described in Definition 36.

5. The free quantum field theory

We extend the Feynman measure $\omega : e^{iL_F}S\Gamma_c\omega SJ\Phi \to \mathbb{C}$, which is something like a measure on classical fields, to $\omega : Te^{iL_F}S\Gamma_c\omega SJ\Phi \to \mathbb{C}$. This extension, restricted to the even degree subalgebra $T_0e^{iL_F}S\Gamma_c\omega SJ\Phi$, is the free quantum field theory. We check that it satisfies analogues of the Wightman axioms.

Formulas involving coproducts can be confusing to write down and manipulate. They are much simpler for the "grouplike" elements g satisfying $\Delta(g) = g \otimes g$, $\eta(g) = 1$, which form a group in any cocommutative Hopf algebra. One problem is that most of the Hopf algebras we use do not have enough grouplike elements over fields: in fact for symmetric algebras the only grouplike element is the identity. However they have plenty of grouplike elements if we add some nilpotent elements to the base field, such as $\exp(\lambda a)$ for any primitive a and nilpotent λ (in characteristic 0). We will adopt the convention that when we talk about grouplike elements, we are tacitly allowing extensions of the base ring by nilpotent elements.

Recall that $Te^{iL_F}S\Gamma_c\omega SJ\Phi$ is the tensor algebra of $e^{iL_F}S\Gamma_c\omega SJ\Phi$, with the product denoted by $\otimes$ to avoid confusing it with the product of $S\Gamma_cSJ\Phi$. We denote the identity of $S\Gamma_cSJ\Phi$ by 1, and the identity of $\mathrm{T}S\Gamma_c\omega SJ\Phi$ by 1_T. The involution $*$ is defined by $(A_1 \otimes \cdots \otimes A_n)^* = A_n^* \otimes \cdots \otimes A_1^*$, and $*$ is -1 on $\Gamma_c\omega SJ\Phi$.

Theorem 31. *If $\omega : e^{iL_F}S\Gamma_c\omega SJ\Phi \to C$ is a Feynman measure, there is a unique extension of ω to $Te^{iL_F}S\Gamma_c\omega SJ\Phi$ such that:*

- (*Gaussian condition*) *If $A, B_1, \ldots, B_m$ are grouplike then*

$$\begin{aligned} &e^{-iL_F}\omega(A \otimes B_m \otimes \cdots \otimes B_1) \\ &\qquad = \sum e^{-iL_F}\omega(A \otimes 1 \otimes \cdots \otimes 1)\Delta(A, B_m \ldots B_1)e^{-iL_F}\omega(B_m \otimes \cdots \otimes B_1). \end{aligned}$$

  (Both sides are considered as densities, as in Definition 9.)

- $e^{-iL_F}\omega(A \otimes A \otimes 1 \otimes \cdots \otimes 1) = 1$ *for A grouplike* (Cutkosky condition; see ['t Hooft 1994, Section 6]).

Proof. We first check that all the products of distributions are well defined by examining their wave front sets. All the distributions appearing have the property that their wave front sets have no positive or negative elements. This follows by induction on the complexity of an element: if all smaller elements have this property, it implies that the products defining it are well defined, and also implies that it has the same property.

The existence and uniqueness of ω follow because the Cutkosky condition defines it on elements of the form $A \otimes 1 \otimes 1 \otimes \cdots \otimes 1$ in terms of those of the form $A \otimes 1 \otimes \cdots \otimes 1$, and the Gaussian condition then determines it on all elements. □

We can also define ω directly as follows. When the propagator is sufficiently regular then the Gaussian condition means that we can write ω on $e^{iL_F} S\Gamma_c \omega S J \Phi$ as a sum over all ways of joining up the fields of an element of $e^{iL_F} S\Gamma_c \omega S J \Phi$ in pairs, where we take the propagator of each pair and multiplying these together. This is of course essentially the usual sum over Feynman diagrams. A minor difference is that we do not distinguish between "internal" vertices associated with a Lagrangian and integrated over all spacetime, and "external" vertices associated with a field and integrated over a compact set: all vertices are associated with a composite operator that may be a Lagrangian or a simple field or a more general composite operator, and all vertices are integrated over compact sets as all coefficients are assumed to have compact support.

Similarly we can define the extension of ω to $Te^{iL_F} S\Gamma_c \omega S J \Phi$ by writing the distributions defining ω as a sum over more complicated Feynman diagrams whose vertices are in addition labeled by nonnegative integers, in such a way that

- the propagators from A_i to A_i are Feynman propagators,
- the propagators from A_i to A_j for $i < j$ are cut propagators Δ, with positive wave front sets on i and negative wave front sets on j, and
- the diagram is multiplied by a factor of $(-1)^{\deg(A_2 A_4 A_6 \dots)}$ (in other words, we apply $*$ to $A_2, A_4, \dots$.)

In general, if the propagator is not sufficiently regular (so that products of propagators might not be defined when some points coincide), we can construct ω by regularization and renormalization as in Section 3, which preserves the conditions defining ω.

Now we show that ω satisfies the locality property of quantum field theories (operators with spacelike-separated supports commute) by showing that it vanishes on the following locality ideal.

Definition 32. We denote by $T_0 S\Gamma_c \omega S J \Phi$ the subalgebra of even degree elements of $TS\Gamma_c \omega S J \Phi$. The locality ideal is the 2-sided ideal of $T_0 S\Gamma_c \omega S J \Phi$ spanned by the coefficients of elements of the form

$$\cdots \otimes Y_1 \otimes ABD \otimes DBC \otimes X_n \otimes \cdots \otimes X_1 - \cdots \otimes Y_1 \otimes AD \otimes DC \otimes X_n \otimes \cdots \otimes X_1$$

(for $A, C \in S\Gamma_c \omega S J \Phi$ and $B, D \in S\Gamma_c \omega S J \Phi[[\lambda]]$ with B, D grouplike) if n is even and there are no points in the support of B that are $\leqslant$any points in the support of A or C, or if n is odd and there are no points in the support of B that are $\geqslant$any points in the support of A or C.

The algebra $T_0 e^{iL_F} S\Gamma_c \omega SJ\Phi$ and its locality ideal are defined in the same way.

Remark 33. The map ω on $T_0 e^{iL_F} S\Gamma_c \omega SJ\Phi$ depends on the choice of Feynman measure. We can define a canonical map independent of the choice of Feynman measure by taking the underlying $*$-algebra to have elements represented by pairs (ω, A) for a Gaussian measure ω and $A \in T_0 e^{iL_F} S\Gamma_c \omega SJ\Phi$, where we identify (ω, A) with $(\rho\omega, \rho A)$ for any renormalization ρ. The canonical state, also denoted by ω, then takes an element represented by (ω, A) to $\omega(A)$.

Theorem 34. *ω vanishes on the locality ideal.*

Proof. We use the notation of Definition 32. We prove this for elements with n even; the case n odd is similar. We can assume that the propagator Δ is sufficiently regular, as we can obtain the general case from this by regularization and renormalization. We will first do the special case when $D = 1$. We can assume that $B = b_1 \dots b_k$ is homogeneous of some order k and write B_I for $\prod_{j\in I} b_j$. If $k = 0$ then the result is obvious as B is constant and both sides are the same, so we can assume that $k > 0$. We show that if $k > 0$ then ω vanishes on

$$\sum_{I\cup J=\{1,\dots k\}} (-1)^{|I|} \cdots \otimes Y_1 \otimes AB_I \otimes B_J C \otimes X_n \otimes \cdots \otimes X_1$$

by showing that the terms cancel out in pairs. This is because if j is the index for which the support of b_j is maximal then ω has the same value on

$$\cdots \otimes Y_1 \otimes AB_I b_j \otimes B_J C \otimes X_n \otimes \cdots \otimes X_1$$

and

$$\cdots \otimes Y_1 \otimes AB_I \otimes b_j B_J C \otimes X_n \otimes \cdots \otimes X_1.$$

Now we do the case of general D. We can assume that the support of D is either $\leqslant$ all points of the support of B or there are no points of it that are $\leqslant$ any points in the support of A or C. In the first case the result follows from the special case $D = 1$ by replacing A and C by AD and CD. In the second case it follows from 2 applications of the special case $D = 1$, replacing B by D and BD, that both terms are equal to $\cdots \otimes Y_1 \otimes A \otimes C \otimes X_n \otimes \cdots \otimes X_1$ and are therefore equal. $\square$

This proof, in the special case that ω vanishes on $B\otimes B - 1\otimes 1$ for B grouplike, is more or less the proof of unitarity of the S-matrix using the "largest time equation" given in ['t Hooft 1994, Section 6]. The locality ideal is not the largest ideal on which ω vanishes, as ω also vanishes on $A \otimes 1 \otimes 1 \otimes B - A \otimes B$; in other words we can cancel pairs $1 \otimes 1$ wherever they occur.

Theorem 35. *Elements of $T_0 S\Gamma_c SJ\Phi$ with spacelike-separated supports commute modulo the locality ideal.*

Proof. It is sufficient to prove this for grouplike degree 2 elements, as if two even degree elements have spacelike-separated supports then they are polynomials in degree 2 elements with spacelike separated supports. We will work modulo the locality ideal. Suppose that the supports of the grouplike elements $W\otimes X\otimes Z$ and Y are spacelike-separated. Then applying Theorem 34 twice gives

$$W\otimes X\otimes YZ = WY\otimes XY\otimes YZ = WY\otimes X\otimes Z$$

Applying this 4 times for various values of W, X, Y, and Z shows that if $A\otimes B$ and $C\otimes D$ are grouplike and have spacelike separated supports, then

$$\begin{aligned}A\otimes B\otimes C\otimes D &= AC\otimes B\otimes I\otimes D = AC\otimes I\otimes I\otimes BD = AC\otimes D\otimes I\otimes B\\ &= C\otimes D\otimes A\otimes B,\end{aligned}$$

so $A\otimes B$ and $C\otimes D$ commute. □

Now we study when the quantum field theory ω is Hermitian, and show that we can find a Hermitian quantum field theory associated to any Hermitian local cut propagator, and show that the group of real renormalizations acts transitively on them.

Definition 36. We say that a Feynman measure ω is Hermitian if its extension to $TS\Gamma_c\omega SJ\Phi$ is Hermitian when restricted to the even subalgebra $T_0S\Gamma_c\omega SJ\Phi$.

Lemma 37. *If the local cut propagator Δ is Hermitian, then it has a Hermitian Feynman measure associated with it.*

Proof. We can assume that the regularization of Δ is also Hermitian, by replacing it by the average of itself and its Hermitian conjugate. We can check directly that the meromorphic family of Feynman measures associated to this Hermitian regularization is Hermitian on $T_0S\Gamma_c\omega SJ\Phi$ (but not on the whole of $TS\Gamma_c\omega SJ\Phi$); in other words $\omega(A_n\otimes\cdots\otimes A_1)=\omega(A_1^*\otimes\cdots\otimes A_n^*)^*$ if n is even. For example, we get a sign factor of $-1^{\deg(A_2)+\deg(A_4)+\cdots}$ in the definition of ω on the first term, a sign factor of $-1^{\deg(A_1)+\deg(A_3)+\cdots}$ form the definition of ω for the second term, whose quotient is the factor $-1^{\deg(A_1)+\deg(A_2)+\cdots}$ coming from the action of $*$ on $A_n\otimes\cdots\otimes A_1$ because n is even. We can then renormalize using real renormalizations to eliminate the poles, and the resulting Feynman measure will be Hermitian. □

Lemma 38. *If a Feynman measure ω is Hermitian and ρ is a complex renormalization, then $\rho(\omega)$ is Hermitian if and only if ρ is real. In particular the subgroup of (real) renormalizations acts simply transitively on the Hermitian Feynman measures associated with a given cut local propagator.*

Proof. This follows from $\rho(\omega)^*=\rho^*(\omega^*)$, and the fact that complex renormalizations act simply transitively on Feynman measures associated with a given cut local propagator. □

Next we show that ω is a state (in other words the space of physical states is positive definite) when the cut propagator Δ is positive, by using a representation of the physical states as a space of distributions. We define the space H_n of n-particle states to be the space of continuous linear maps $S^n\Gamma\omega\Phi \to \mathbb{C}$ (considered as compactly supported symmetric distributions on M^n) whose wave front sets have no positive or negative elements, with a sesquilinear form given by

$$\langle a, b\rangle = \int_{x,y\in M^n} a(x_1,\dots)\prod_j \Delta(x_j, y_j)b(y_j,\dots)^* dxdy.$$

This is similar to the usual definition of the inner product on the space of states of a free field theory, except that we are using distributions rather than smooth functions. We check this is well defined. To show the product of distributions in the integral is defined we need to check that no sum of nonzero elements of the wave front sets is zero, and this follows because nonzero elements of the wave front set of the product of propagators are of the form (p, q) with $p > 0$ and $q < 0$, but a and b by assumption have no positive or negative elements in their wave front sets. The integral over M^n is well defined because a and b have compact support.

Lemma 39. *There is a map f from $T_0S\Gamma_c\omega SJ\Phi$ to the orthogonal direct sum $\bigoplus H_n$ with*

$$\omega(AB) = \langle f(A^*), f(B)\rangle.$$

Proof. By Theorem 31, $\omega(AB)$ is given by

$$\sum \omega(A')\Delta(A'', B'')\omega(B')$$

where $\sum A' \otimes A''$ is the image of A under the coaction of $\Gamma_c SJ\Phi$. This is equal to $\langle f(A^*), f(B)\rangle$ if we define $f(A)$ as follows. Suppose that

$$A = A_{11}A_{12}\cdots \otimes A_{21}A_{22}\dots,$$

and let the image of A_{jk} under the coaction of $\Gamma_c SJ\Phi$ be $\sum A'_{jk} \otimes A''_{jk}$. Then $\omega(A'_{11}A'_{12}\cdots \otimes A'_{21}A'_{22}\dots.)$ can be regarded as a distribution on M^n, where n is the total number of elements A_{jk}. On the other hand, $A''_{11}A''_{12}\dots A''_{21}A''_{22}\dots$ is a function on M^m, where m is the sum of the degree of the elements A''_{jk}, in other words the number of fields occurring in them. There is also a map from m to n, which induces a map from M^n to M^m, and so by push-forward of densities a map from densities on M^n to densities on M^m. The image $f(A)$ is then given by taking the push-forward from M^n to M^m of the compactly supported distribution $\omega(A'_{11}A'_{12}\cdots \otimes A'_{21}A'_{22}\dots.)$ on M^n, multiplying by the function $A''_{11}A''_{12}\dots A''_{21}A''_{22}\dots$ on M^m, symmetrizing the result, and repeating this for each summand of $\sum A'_{jk} \otimes A''_{jk}$. ∎

Corollary 40. *If the cut local propagator* Δ *is positive, then*

$$\omega : Te^{iL_F} S\Gamma_c \omega SJ\Phi \to \mathbb{C}$$

is a state.

Proof. This follows from the previous lemma, because if Δ is positive then so is the sesquilinear form $\langle\ ,\ \rangle$ on H_n, and therefore $\omega(A^*A) = \langle f(A), f(A)\rangle \geqslant 0$. $\square$

6. Interacting quantum field theories

We construct the quantum field theory of a Feynman measure and a compactly supported Lagrangian, by taking the image of the free field theory ω under an automorphism e^{iL_I} where L_I is the interaction part of the Lagrangian. This automorphism is only well defined if the interaction Lagrangian L_I has infinitesimal coefficients, so the interacting quantum field theories we construct are perturbative theories taking values in rings of formal power series $\mathbb{C}[\boldsymbol{\lambda}] = \mathbb{C}[\lambda_1, \dots]$ in the coupling constants $\lambda_1, \dots$. (By "infinitesimal" we mean elements of formal power series rings with vanishing constant term.) We then lift the construction to all actions (possible without compact support) by showing that infrared divergences cancel up to inner automorphisms.

Lemma 41. *The Hopf algebra* $S\Gamma_c\omega SJ\Phi$ *acts on the algebra* $T_0 S\Gamma_c\omega SJ\Phi$, *and maps the locality ideal to itself. Group-like Hermitian elements of the Hopf algebra* $S\Gamma_c\omega SJ\Phi[\![\boldsymbol{\lambda}]\!]$ *preserve the subset of positive elements, and therefore act on the space of states of* $T_0 S\Gamma_c\omega SJ\Phi[\![\boldsymbol{\lambda}]\!]$.

Proof. Group-like elements are algebra automorphisms, and if they are also Hermitian they commute with the involution $*$. In particular grouplike Hermitian elements preserve the set of positive elements (generated by positive linear combinations of elements of the form a^*a), and so map positive linear forms to positive linear forms. $\square$

Definition 42. The quantum field theory of a Lagrangian $L = L_F + L_I$, where L_I has compact support and infinitesimal coefficients, is $e^{-iL}\omega : T_0 S\Gamma_c\omega SJ\Phi \to \mathbb{C}[\![\boldsymbol{\lambda}]\!]$.

The Hopf algebra $S\Gamma_c\omega SJ\Phi$ acts on the vector space $S\Gamma_c\omega SJ\Phi$ by multiplication, so grouplike elements of the form $e^{iL_F+iL_I}$ take $S\Gamma_c\omega SJ\Phi$ to $e^{iL_F}S\Gamma_c\omega SJ\Phi$ and $T_0 S\Gamma_c\omega SJ\Phi$ to $T_0 e^{iL_F} S\Gamma_c\omega SJ\Phi$. Since ω is in the dual of $T_0 e^{iL_F} S\Gamma_c\omega SJ\Phi$, this shows that $e^{-iL}\omega$ is in the dual of $T_0 S\Gamma_c\omega SJ\Phi$.

Corollary 43 (locality). *Elements of* $T_0 S\Gamma_c\omega SJ\Phi$ *with spacelike-separated supports commute when acting on the space of physical states of* $e^{-iL}\omega$.

Proof. By Theorem 34 the operators of the locality ideal act trivially on the space of physical states of ω. Since e^{-iL} preserves the locality ideal, the locality ideal also acts trivially on the space of physical states of $e^{-iL}\omega$. By Theorem 35 this implies that operators with spacelike separated supports commute on this space. □

This constructs the quantum field theory of a Lagrangian whose interaction part has compact support (and is infinitesimal). We now extend this to the case when the interaction part need not have compact support. We do this by using a cutoff function to give the Lagrangian compact support, and then we then try to show that the result is independent of the choice of cutoff function, provided it is 1 in a sufficiently large region. To do this we need to assume that spacetime is globally hyperbolic, and we also find that the result is not quite independent of the choice of cutoff.

If f is a smooth function on M then multiplication by f is a linear transformation of $\Gamma\omega SJ\Phi$ and therefore induces a homomorphism of $S\Gamma\omega SJ\Phi$, denoted by $A \to A^f$. If $A = e^{iL}$ is grouplike, then $A^f = e^{iLf}$. If f has compact support then so does A^f so that $A^f\omega$ is defined. We try to extend the definition of $A^f\omega$ to more general functions f in the hope that we can take f to be close to 1.

Lemma 44. *Suppose that f and g are compactly supported smooth functions on M and n is even. If $f = g$ on the past of $A_1 \dots A_n$ then (modulo the locality ideal)*

$$e^{-iL_F}A^f\omega(A_n \otimes \cdots \otimes A_1) = e^{-iL_F}A^g\omega(A_n \otimes \cdots \otimes A_1)$$

If $f = g$ on the future of $A_1 \dots A_n$ then

$$e^{-iL_F}A^f\omega(A_n \otimes \cdots \otimes A_1) = e^{-iL_F}A^g\omega(A^{g-f} \otimes 1 \otimes A_n \otimes \cdots \otimes A_1 \otimes 1 \otimes A^{g-f})$$

Proof. We work modulo the locality ideal. The first equality follows from

$$A^{-f}A_n \otimes \cdots \otimes A^{-f}A_1 = A^{-g}A_n \otimes \cdots \otimes A^{-g}A_1$$

which in turn follows from Theorem 34 by repeatedly inserting $A^{f-g} \otimes A^{f-g}$ (using the fact that n is even). The second equality follows in the same way from

$$\begin{aligned} &A^{-f} \otimes A^{-f} \otimes A^{-f}A_n \otimes \cdots \otimes A^{-f}A_1 \otimes A^{-f} \otimes A^{-f} \\ &= A^{-f} \otimes A^{-g} \otimes A^{-g}A_n \otimes \cdots \otimes A^{-g}A_1 \otimes A^{-g} \otimes A^{-f}. \end{aligned}$$ □

This lemma shows that the restriction of $A^f\omega$ to arguments with support in some fixed compact subset of M is almost independent of the choice of f provided that f is 1 on the convex hull of the argument: different choices of f are related by a locally inner automorphism of $T_0S\Gamma_c\omega SJ\Phi$, given by conjugation by elements of the form $1 \otimes A^h$. If the spacetime is globally hyperbolic in the sense that the convex hull of a compact set is contained in a compact set, then we can always find a suitable f that is 1 on the convex hull X of the argument, so we can construct the interacting quantum field theory. The result does not depend on the choice of cutoff f on the future of X, but does depend slightly on the choice of cutoff in the past of X. The choice of cutoff in the past corresponds to choices of the vacuum: roughly speaking, we turn off the interaction in the distant past, which gives different vacuums. More precisely, if we have two different cutoffs f and g then their vacuums, which are the images of $e^{i(L_F+fL_I)}$ and $e^{i(L_F+gL_I)}$ will differ by a factor of $e^{i(f-g)L_I}$. This does not change the observable physics, because all these choices of cutoffs give isomorphic quantum field theories. However it does cause difficulties in constructing a Lorentz invariant theory, because the choice of cutoff in the past is not Lorentz invariant, so the vacuums are also not Lorentz invariant, or in other words Lorentz invariance may be spontaneously broken. Presumably in theories with a mass gap one can take the limit as the cutoff in the past tends to time $-\infty$ and get a Lorentz invariant vacuum, but in theories with massless particles such as QED there is an obstruction to constructing a Lorentz invariant vacuum: Lorentz invariance might be spontaneously broken by infrared divergences. This is a well known problem, which is not worth worrying about too much, because the physical universe is not globally Lorentz invariant.

The time-ordered operator $T(A)$ of an element $A \in S\Gamma_c\omega SJ\Phi$ is defined to be $1 \otimes A$. This has the property that

$$T(A_n \dots A_1) = 1 \otimes A_n \dots A_1 = 1 \otimes A_n \otimes \cdots \otimes 1 \otimes A_1 = T(A_n) \dots T(A_1)$$

whenever the composite fields $A_i \in \Gamma_c\omega SJ\Phi$ are in order of increasing time of their supports. This formula is sometimes used as a "definition" of the time-ordered product $T(A_n \dots A_1)$, though this does not define it when some of the factors have overlapping supports, and in general the time-ordered product depends on the choice of a Feynman measure ω. The scattering matrix S of the quantum field theory is $S = T(e^{iL_I}) = 1 \otimes e^{iL_I}$; this is essentially the LSZ reduction formula of Lehmann, Symanzik, and Zimmermann [Lehmann et al. 1955].

We now show that if we change the Feynman measure, then we still get an isomorphic quantum field theory provided we make a suitable change in the Lagrangian. If we change ω to a different Feynman measure for the same cut local propagator, these will differ by a unique renormalization ρ; in other words the other Feynman measure will be $\rho\omega$. The quantum field theory $e^{-iL}\omega$ changes under this renormalization of ω by

$$\begin{aligned} e^{-iL}\omega(A_1 \otimes \cdots) = \omega(e^{iL}A_1 \otimes \dots) &= \rho(\omega)(\rho(e^{iL}A_1) \otimes \cdots) \\ &= \rho(e^{-iL})\rho(\omega)\big(\rho(e^{-iL})\rho(e^{iL}A_1) \otimes \cdots\big), \end{aligned}$$

so the quantum field theory stays the same under renormalization by ρ if we transform the Lagrangian by

$$iL \to \log(\rho(\exp(iL))),$$

which is a nonlinear transformation because renormalizations need not commute with products or exponentiation, and change the operators A_n by

$$A_n \to \rho(e^{-iL})\rho(e^{iL}A_n).$$

If A_n is a simple operator and ρ satisfies the condition of Example 25, then

$$\rho(e^{iL}A_n) = \rho(e^{iL})\rho(A_n) = \rho(e^{iL})A_n,$$

so in this special case A_n is unchanged, or in other words simple operators are not renormalized. The behavior of composite operators under renormalization can be quite complicated when expanded out in terms of fields. The usual Wightman distributions used to construct a quantum field theory use only simple operators, so the only effect of renormalization on Wightman distributions comes from the nonlinear transformation of the Lagrangian. This nonlinear transformation of Lagrangians is the usual action of renormalizations on Lagrangians used in physics texts to convert an infinite "bare" Lagrangian L to a finite physical one L_0; the bare and physical Lagrangians are related by $iL_0 = \log(\rho(\exp(iL))$, where ρ is an infinite renormalization taking an infinite Feynman measure, such as the one given by dimensional regularization, to a finite one.

The orbit of a Lagrangian under this nonlinear action of the ultraviolet group is in general infinite-dimensional. It can sometimes be cut down to a finite-dimensional space as follows. As in Example 26, we cut down to the group of renormalizations of mass dimension at most 0, which acts on the space of Lagrangians whose coupling constants all have mass dimension at least 0. If we also add the condition that the Lagrangian is Lorentz invariant, then we sometimes get finite-dimensional spaces of Lagrangians. The point is that the classical fields themselves tend to have positive mass dimension, so if the coupling constants all have nonnegative mass dimension then the fields appearing in any term of the Lagrangian have total mass at most d (canceling out the $-d$ coming from the density) which severely limits the possibilities. At one time the Lagrangians with all coupling constants of nonnegative mass dimension were called renormalizable Lagrangians, though now all Lagrangians are regarded as renormalizable in a more general sense where one allows an infinite number of terms in the Lagrangian.

7. Gauge invariance and anomalies

If a Lagrangian is invariant under some group, this does not imply that the quantum field theories we construct from it are also invariant, because as pointed out in [Fujikawa 1979] we also need to choose a Feynman measure and there may not be an invariant way of doing this. The obstructions to finding an invariant quantum field theory lie inside certain cohomology groups and are called anomalies. We show that if these anomalies vanish then we can construct invariant quantum field theories.

Suppose that a group G acts on $SJ\Phi$ and preserves the set of Feynman measures with given cut local propagator, and suppose that we have chosen one such Feynman measure ω. In practice we often start with an action of a Lie algebra or superalgebra, such as that generated by the BRST operator, which can be turned into a group action in the usual way by working over a ring with nilpotent elements. If $g \in G$ then $g\omega$ is another Feynman measure with the same propagator, so

$$\omega = \rho_g g \omega$$

for a unique renormalization ρ_g. This defines a nonabelian 1-cocycle: $\rho_{gh} = \rho_g g(\rho_h)$, where $g(\rho_h) = g\rho_h g^{-1}$. Since ω is invariant under $\rho_g g$, we find that

$$\omega(e^{iL}A_1) = \omega(\rho_g g(e^{iL}A_1)) = \omega\big(e^{iL}e^{-iL}\rho_g g(e^{iL}A_1)\big),$$

so that $e^{-L}\omega$ is invariant under the transformation that takes arguments A_1 to $e^{-iL}\rho_g g(e^{iL}A_1)$. This transformation fixes 1 if e^{iL} is fixed by $\rho_g g$. If in addition $\rho_g g(e^{iL}A_1) = \rho_g g(e^{iL})\rho_g g(A_1)$ (which is not automatic as ρ_g need not preserve products) then A_1 is taken to $\rho_g g(A_1)$ by this transformation.

This shows that we really want a Lagrangian L such that e^{iL} is invariant under the modified action $e^{iL} \to \rho_g g(e^{iL})$. This is not the same as asking for $\rho_g g(iL) = iL$ because ρ_g need not preserve products (although g usually does). In practice we usually have a Lagrangian L with L (and e^{iL}) invariant under G, and the problem is whether it can be modified to L' so that $e^{iL'}$ is invariant under the twisted action. The powers of L span a coalgebra all of whose elements are G-invariant. Conversely, given a coalgebra C all of whose elements are invariant under some group action, there is a canonical G-invariant grouplike element associated to this coalgebra with coefficients in the dual algebra of C. So a fundamental question is whether the maximal coalgebra in the space of G-invariant classical actions is isomorphic to the maximal coalgebra in the space of actions invariant under the twisted action of G.

The simplest case is when one can find a G-invariant Feynman measure, in which case the cocycle is trivial and the twisted action of G is the same as the untwisted action. In terms of the cocycle above, $\rho\omega$ is invariant for some renormalization ω if and only if $\rho_g = \rho^{-1}g(\rho)$ for all g (where $g(\rho) = g\rho g^{-1}$), in other words there is an invariant measure ω if and only if the cocycle is a coboundary. This case happens, for example, when spacetime M is Minkowski space and G is the Lorentz or Poincare group (or one of their double covers). Dimensional regularization in this case is automatically G-invariant, and so gives a G-invariant Feynman measure.

In the case of BRST operators, there need not be any G-invariant Feynman measure. In this case the following theorem shows that one can find suitable coalgebras provided that certain obstructions, called anomalies, all vanish. The renormalizations ρ_g need not preserve products in $S\Gamma\omega SJ\Phi$, but do preserve the coproduct and also fix all elements of $\Gamma\omega SJ\Phi$ if they are normalized as in Example 25. So we have an action of G on the space $V = \Gamma\omega SJ\Phi$, which lifts to two different actions of the coalgebra SV, the first $\sigma_1(g)$ preserving the product, and the second $\sigma_2(g) = \rho_g\sigma_1(g)$ given by twisting the first by the cocycle ρ_g.

Theorem 45. *Suppose that V is a real vector space acted on by a group G, and there are two extensions σ_1. σ_2 of this action to the coalgebra SV. If the cohomology group $H^1(G, V)$ vanishes then the maximal coalgebras in SV whose elements are fixed by these 2 actions of G are isomorphic under an isomorphism fixing the elements of V.*

Proof. We construct an isomorphism f from the maximal coalgebra in the space of σ_1-invariant elements to the maximal coalgebra in the space of σ_2-invariant elements by induction on the degree of elements. We start by taking f to be the identity map on elements of degree at most 1. We can assume that the 2 actions coincide on elements of degree less than n, and have to find an isomorphism f making them the same on elements of degree n, which we will do by adding elements of V to a basis of the elements of degree n. Suppose that a is an element of degree $n > 1$ contained in a coalgebra of G-invariant elements. We want to find $v \in V$ so that

$$\sigma_1(g)(a + v) = \sigma_2(g)(a) + v$$

or equivalently

$$\sigma_1(g)(v) - v = \sigma_2(g)(a) - a.$$

The right hand side, as a function of g, is a 1-coboundary of an element $a \in SV$, and therefore a 1-cocycle. We show that the right hand side is in V. We have

$$\Delta(a) = a \otimes 1 + 1 \otimes a + \sum_i b_i \otimes c_i$$

for some elements b_i and c_i of degrees less than n invariant under G (for both actions, which coincide on elements of degree less than n). Applying σ_2 we find that $\Delta(\sigma_2(g)a) = \sigma_2(g)a \otimes 1 + 1 \otimes \sigma_2(g)a + \sum_i b_i \otimes c_i$, so subtracting these two identities shows that $\sigma_2(g)(a) - a$ is a primitive element of SV and therefore in V. Therefore the right hand side, as a function of g, is a 1-cocycle with values in V. The solvability of the condition for v says exactly that this expression is the coboundary of some element $v \in V$. In other words the obstruction to finding a suitable v is exactly an element of the cohomology group $H^1(G, V)$, so as we assume this group vanishes we can always solve for v. $\square$

Example 46. We take V to be $\Gamma\omega SJ\Phi$, and G to be some group acting on V. Then the spaces of classical and quantum actions are coalgebras acted on by G, whose primitive elements can be identified with V. If $H^1(G, \Gamma\omega SJ\Phi)$ vanishes, then the maximal G-invariant coalgebra in the coalgebra of classical actions is isomorphic to the maximal G-invariant coalgebra in the coalgebra of quantum actions. So if L is a G-invariant classical Lagrangian, then e^L is a G-invariant classical action, so gives a G-invariant quantum action. One cannot get a G-invariant quantum action by exponentiating a G-invariant quantum Lagrangian because the space of quantum actions does not in general have a G-invariant product.

Example 47. Sometimes the group G only fixes classical Lagrangians up to boundary terms, in other words the Lagrangian is a G-invariant element of $\Gamma\omega SJ\Phi/D$. In this case one replaces the cohomology group $H^1(G, \Gamma\omega SJ\Phi)$ by

$$H^1(G, \Gamma\omega SJ\Phi/D).$$

The element e^{iL_F} lies in the completion of $S\Gamma\omega SJ\Phi$ and is fixed by the zeroth order part of the BRST operator. So the BRST operator acts on $e^{iL_F}S\Gamma\omega SJ\Phi$.

The groups $H^1(G, \Gamma\omega SJ\Phi)$ and $H^1(G, \Gamma\omega SJ\Phi/D)$ (and their variations for Poincare invariant Lagrangians) for the BRST operators of gauge theories have been calculated in many cases, at least for the case of Minkowski space (see [Barnich et al. 2000], for example) and are sometimes zero, in which case corresponding invariant quantum field theories exist.

References

[Abe 1980] E. Abe, *Hopf algebras*, Cambridge Tracts in Mathematics **74**, Cambridge University Press, Cambridge, 1980. MR 83a:16010 Zbl 0476.16008

[Barnich et al. 2000] G. Barnich, F. Brandt, and M. Henneaux, "Local BRST cohomology in gauge theories", *Phys. Rep.* **338**:5 (2000), 439–569. MR 2002d:81208 Zbl 1097.81571

[Bernstein 1972] I. N. Bernstein, "The analytic continuation of generalized functions with respect to a parameter", *Funkcional. Anal. i Priložen.* **6**:4 (1972), 26–40. In Russian; translated in *Funct. Anal. Appl.* **6** (1973), 273–285. MR 47 #9269 Zbl 0282.46038

[Borchers 1962] H.-J. Borchers, "On structure of the algebra of field operators", *Nuovo Cimento* (10) **24** (1962), 214–236. MR 25 #5713 Zbl 0129.42205

[Connes and Kreimer 2000] A. Connes and D. Kreimer, "Renormalization in quantum field theory and the Riemann–Hilbert problem, I: The Hopf algebra structure of graphs and the main theorem", *Comm. Math. Phys.* **210**:1 (2000), 249–273. MR 2002f:81070

[Etingof 1999] P. Etingof, "Note on dimensional regularization", pp. 597–607 in *Quantum fields and strings: a course for mathematicians*, vol. 1, edited by P. Deligne et al., Amer. Math. Soc., Providence, RI, 1999. MR 2001k:81177 Zbl 1137.81358

[Fujikawa 1979] K. Fujikawa, "Path-integral measure for gauge-invariant fermion theories", *Phys. Rev. Lett.* **42**:18 (1979), 1195–1198.

[Grothendieck 1967] A. Grothendieck, "Éléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas IV", *Inst. Hautes Études Sci. Publ. Math.* **32** (1967), 361. MR 39 #220

['t Hooft 1994] G. 't Hooft, *Under the spell of the gauge principle*, Advanced Series in Mathematical Physics **19**, World Scientific, River Edge, NJ, 1994. MR 96k:81165 Zbl 0881.53063

[Hörmander 1990] L. Hörmander, *The analysis of linear partial differential operators, I: Distribution theory and Fourier analysis*, 2nd ed., Grundlehren der Math. Wiss. **256**, Springer, Berlin, 1990. Reprinted in *Classics of Mathematics* series, Springer, 2003. MR 91m:35001a Zbl 0712.35001

[Kreimer 1998] D. Kreimer, "On the Hopf algebra structure of perturbative quantum field theories", *Adv. Theor. Math. Phys.* **2**:2 (1998), 303–334. MR 99e:81156

[Lehmann et al. 1955] H. Lehmann, K. Symanzik, and W. Zimmermann, "Zur Formulierung quantisierter Feldtheorien", *Nuovo Cimento* (10) **1** (1955), 205–225. MR 17,219h Zbl 0066.44006

[Streater and Wightman 2000] R. F. Streater and A. S. Wightman, *PCT, spin and statistics, and all that*, Princeton Landmarks in Physics, Princeton University Press, Princeton, NJ, 2000. Corrected third printing of the 1978 edition. MR 2003f:81154 Zbl 1026.81027

Communicated by Bjorn Poonen
Received 2010-08-23 Revised 2011-02-18 Accepted 2011-04-24

reb@math.berkeley.edu *Department of Mathematics, University of California, Berkeley, CA 94720-3840, United States*

ALGEBRA AND NUMBER THEORY 5:5(2011)

Density of rational points on isotrivial rational elliptic surfaces

Anthony Várilly-Alvarado

For a large class of isotrivial rational elliptic surfaces (with section), we show that the set of rational points is dense for the Zariski topology, by carefully studying variations of root numbers among the fibers of these surfaces. We also prove that these surfaces satisfy a variant of weak-weak approximation. Our results are conditional on the finiteness of Tate–Shafarevich groups for elliptic curves over the field of rational numbers.

1. Introduction

1A. *Del Pezzo surfaces of degree* 1*: a sample result.* Let X be a smooth projective geometrically integral surface over a number field k. Fix an algebraic closure $\bar{k}$ of k and assume that X is geometrically rational, i.e., the base extension $X_{\bar{k}} := X \times_k \bar{k}$ is birational to the projective plane $\mathbb{P}^2_{\bar{k}}$. A well-known result of Iskovskikh [1979] guarantees that X is k-birational to either a rational conic bundle or a del Pezzo surface.

Del Pezzo surfaces that are not geometrically isomorphic to $\mathbb{P}^1_{\bar{k}} \times \mathbb{P}^1_{\bar{k}}$ are classified by their *degree* $d := K_X^2$, an integer in the range $1 \leq d \leq 9$. Segre and Manin have shown that if X is a del Pezzo surface with $d \geq 2$, and if X contains a k-point not lying on an explicitly computable locus, then $X(k)$ is dense in the Zariski topology; moreover, X is k-unirational in this case [Manin and Hazewinkel 1974, Theorem 29.4]. Surfaces X with $d = 1$ come furnished with a rational point (the base point of the anticanonical linear system). Hence the question: is $X(k)$ dense for the Zariski topology? One of our goals in this paper is to shed some light on this question, in the case when $k = \mathbb{Q}$.

Theorem 1.1. *Let A, B be nonzero integers, and let X be the del Pezzo surface of degree* 1 *over $\mathbb{Q}$ given by*

$$w^2 = z^3 + Ax^6 + By^6 \tag{1}$$

MSC2000: primary 11G35; secondary 14G05, 11G05.
Keywords: rational elliptic surfaces, del Pezzo surfaces, root numbers.

in $\mathbb{P}_{\mathbb{Q}}(1, 1, 2, 3)$. *Assume that Tate–Shafarevich groups of elliptic curves over* $\mathbb{Q}$ *with* j*-invariant* 0 *are finite. If* $3A/B$ *is not a rational square*, *or if* A *and* B *are relatively prime and* $9 \nmid AB$, *then the rational points of* X *are Zariski dense.*

See Section 2 for statements of our most general results.

Remarks 1.2. (i) Every del Pezzo surface of degree 1 is isomorphic to a *smooth* sextic hypersurface in $\mathbb{P}_k(1, 1, 2, 3)$; conversely, a smooth sextic hypersurface in this weighted projective space is a del Pezzo surface of degree 1 [Kollár 1996, Theorem III.3.5].

(ii) Using explicit rational base changes, it is shown in [Ulas 2007, Corollary 4.4] that the conclusion of Theorem 1.1 holds *unconditionally* in the case $A = 1$.

(iii) The restriction in (1) that A and B are integers is not severe. If A and B are rational numbers, one can clear denominators and rescale the variables to obtain an equation of the form (1).

(iv) Using the methods in [Várilly-Alvarado 2008], we may compute $\operatorname{Pic} X$ for the surfaces (1). If $\operatorname{rk}\operatorname{Pic} X = 1$, then X is $\mathbb{Q}$-minimal, and is thus a "genuine" del Pezzo surface of degree 1, i.e., X is not the blow-up of a higher degree surface at closed $\mathbb{Q}$-points. This is the case, for example, if $A = B = p^3$, where $p > 3$ is a prime number; see Theorem 1.1 of that reference.

Blowing up the base point of the anticanonical linear system of a del Pezzo surface of degree 1, we obtain a rational elliptic surface. These are the main objects of study in our paper. However, we state our results in Section 2 in terms of hypersurfaces in $\mathbb{P}_{\mathbb{Q}}(1, 1, 2, 3)$ to emphasize the connection with del Pezzo surfaces of degree 1.

1B. ***Rational elliptic surfaces.*** Let k be a number field, and let $(\mathscr{E}, \rho, \sigma)$ be an *elliptic surface with base* $\mathbb{P}^1_k$, i.e., a smooth surface $\mathscr{E}$ together with a morphism $\rho \colon \mathscr{E} \to \mathbb{P}^1_k$ that has a section $\sigma : \mathbb{P}^1_k \to \mathscr{E}$, such that ρ is a relatively minimal elliptic fibration and has at least one (geometric) singular fiber. We often write $\mathscr{E}$ instead of $(\mathscr{E}, \rho, \sigma)$, the morphisms ρ and σ being understood. Suppose that $\mathscr{E} \times_k \bar{k}$ is birational to $\mathbb{P}^2_{\bar{k}}$ (in which case we say that $\mathscr{E}$ is *rational*). Then the generic fiber of ρ is an elliptic curve $E/k(T)$ that can given by a Weierstrass equation of the form

$$Y^2 = X^3 + a(T)X + b(T), \quad a(T), b(T) \in k[T], \tag{2}$$

where

$$\deg a(T) \le 4, \quad \deg b(T) \le 6 \quad \text{and} \quad \Delta := 4a(T)^3 + 27b(T)^2 \notin k.$$

Conversely, any elliptic curve $E/k(T)$ of the form (2) uniquely extends to a rational elliptic surface with base $\mathbb{P}^1_k$ (the Kodaira–Néron model of E).

We associate to $\mathscr{E}$ a sextic hypersurface X in the weighted projective space $\mathbb{P}_k(1,1,2,3)$ as follows. Let $k[x,y,z,w]$ be the graded ring where the variables x, y, z, w have weights $1, 1, 2, 3$, respectively. Set

$$\mathbb{P}_k(1,1,2,3) := \operatorname{Proj} k[x,y,z,w]$$

and let X be the sextic hypersurface

$$w^2 = z^3 + G(x,y)z + F(x,y), \tag{3}$$

where

$$G(x,y) = y^4 a(x/y) \quad \text{and} \quad F(x,y) = y^6 b(x/y).$$

The schemes X and $\mathscr{E}$ are birational: X can be obtained from $\mathscr{E}$ by contracting the image of the section σ as well as the components of the singular fibers of ρ that do not meet $\sigma(\mathbb{P}^1_k)$. In general, X will be a singular hypersurface.

We are interested in the qualitative distribution of the set $\mathscr{E}(k)$. In particular, we want to determine if the set $\mathscr{E}(k)$ (equivalently, the set $X(k)$) is dense for the Zariski topology. Our investigations rely heavily on the root numbers of the fibers of ρ, and for this reason we focus our attention on the case $k = \mathbb{Q}$.

To prove that $\mathscr{E}(\mathbb{Q})$ is Zariski dense, it suffices to show that for infinitely many $t \in \mathbb{P}^1(\mathbb{Q})$, the fiber $\mathscr{E}_t$ of ρ is an elliptic curve with positive Mordell–Weil rank. Assuming finiteness of Tate–Shafarevich groups, Nekovář, Dokchitser and Dokchitser have shown that the root number of an elliptic curve $E/\mathbb{Q}$ is $(-1)^{\operatorname{rank}(E)}$ (the parity conjecture; see [Nekovář 2001; Dokchitser and Dokchitser 2010]). We study the variation of root numbers among the smooth fibers of $\mathscr{E}$, hoping to find infinitely many fibers with negative root number.

Rohrlich [1993] pioneered the study of variations of root numbers on algebraic families of elliptic curves. Many authors followed suit; see, for example, [Manduchi 1995; Grant and Manduchi 1997; Grant and Manduchi 1998; Rizzo 2003; Conrad et al. 2005]. Some authors (see notably [Conrad et al. 2005, p. 686]) have observed that if the fibers of an elliptic surface lack "geometric variation," then often there are simple formulae that describe the root numbers of these fibers; see, for example [Rohrlich 1996, Corollary to Proposition 10]. For this reason, we restrict our attention to *isotrivial* rational elliptic surfaces, i.e., surfaces $\mathscr{E}$ as above for which the modular invariant $j(E)$ has no T-dependence. Such surfaces arise as families of (quadratic, cubic, quartic or sextic) twists of a fixed elliptic curve $E_0/\mathbb{Q}$:

(i) (quadratic twists) $Y^2 = X^3 + af(T)^2X + bf(T)^3$ with $a, b \in k$, $f(T) \in k[T]$ and $1 \le \deg f(T) \le 2$,

(ii) (cubic twists) $Y^2 = X^3 + f(T)^2$ with $f(T) \in k[T]$ and $1 \le \deg f(T) \le 3$,

(iii) (quartic twists) $Y^2 = X^3 + f(T)X$ with $f(T) \in k[T]$ and $1 \le \deg f(T) \le 4$,

(iv) (sextic twists) $Y^2 = X^3 + f(T)$ with $f(T) \in k[T] \setminus k[T]^2$ and $1 \le \deg f(T) \le 6$.

We use Rohrlich's formulae for local root numbers, together with those of Halberstadt [1998] and Rizzo [2003], to assemble root number formulae for quartic and sextic twists of elliptic curves over $\mathbb{Q}$ (see Propositions 4.8 and 4.4, respectively). We then combine our explicit formulae for root numbers with an adaptation of a sieve introduced by Gouvêa and Mazur [1991] and Greaves [1992]. The modified sieve allows us to search for infinitely many pairs of fibers on a surface that have *opposite* root numbers, which yields our density results (Theorems 2.1 and 2.3). For a similarly motivated idea, see [Manduchi 1995].

Outline of the paper. In Section 2, we state our density theorems (Theorems 2.1, 2.3 and 2.6), and we relate them to the literature, where many similar results can be found under the umbrella of Mazur's conjecture on the topology of rational points. In Section 3, we make precise the relation between isotrivial rational elliptic surfaces and del Pezzo surfaces of degree 1. In Section 4, we present formulae for the root numbers of elliptic curves $E_\alpha/\mathbb{Q}$ of the form $y^2 = x^3 + \alpha$ or $y^2 = x^3 + \alpha x$, where α is a nonzero integer. We use our formulae to give conditions on integers α and β under which E_α and E_β have opposite root numbers (Corollaries 4.5 and 4.9), a crucial input in the proof of our density results. In Section 5, we turn our attention to sieving, and present our modification of the squarefree sieve of Gouvêa, Mazur and Greaves. In Section 6, we use this sieve to locate infinite families of fibers on elliptic surfaces with *opposite* root number, and thus prove Theorems 2.1 and 2.3. In Section 7, we specialize to the case of "diagonal" del Pezzo surfaces of degree 1 over $\mathbb{Q}$. Finally, we prove Theorem 2.6 in Section 8.

2. Main results

Let $F(x, y) \in \mathbb{Z}[x, y]$ be a homogeneous binary form. We say that F *has a fixed prime divisor* if there is a prime number p such that $F(a, b) \in p\mathbb{Z}$ for all $a, b \in \mathbb{Z}$. Note that if the content of $F(x, y)$ is not divisible by p, then $F(x, y)$ mod p has at most deg $F(x, y)$ zeroes in $\mathbb{P}^1(\mathbb{F}_p)$. Hence, if p is a fixed prime divisor of $F(x, y)$, then $p + 1 \leq \deg F(x, y)$.

2A. *Sextic twists and del Pezzo surfaces of degree* 1.

Let

$$\rho : \mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$$

be an isotrivial rational elliptic surface whose associated sextic hypersurface

$$X \subseteq \mathbb{P}_{\mathbb{Q}}(1, 1, 2, 3)$$

is smooth (hence a del Pezzo surface of degree 1). We show in Section 3 that X must be isomorphic to a sextic of the form

$$w^2 = z^3 + F(x, y),$$

where $F(x, y)$ is a squarefree homogeneous form of degree 6. The generic fiber $E/\mathbb{Q}(T)$ of $\mathscr{E}$ is isomorphic to

$$Y^2 = X^3 + b(T), \quad \text{where } b(T) = F(T, 1) \text{ or } F(1, T),$$

and can be thought of family of sextic twists. We prove the following density result for this class of surfaces.

Theorem 2.1. *Let $F(x, y) \in \mathbb{Z}[x, y]$ be a homogeneous binary form of degree 6; assume that the coefficients of x^6 and y^6 are nonzero. Let X be the del Pezzo surface of degree 1 over $\mathbb{Q}$ given by*

$$w^2 = z^3 + F(x, y) \tag{4}$$

in $\mathbb{P}_{\mathbb{Q}}(1, 1, 2, 3)$. Let c be the content of F and write $F(x, y) = cF_1(x, y)$ for some $F_1(x, y) \in \mathbb{Z}[x, y]$. Suppose that F_1 has no fixed prime divisors and that $F_1 = \prod_i f_i$, where the $f_i \in \mathbb{Z}[x, y]$ are irreducible homogeneous forms. Assume further that

$$\mu_3 \not\subseteq \mathbb{Q}[t]/f_i(t, 1) \quad \text{for some } i, \tag{5}$$

where μ_3 is the group of third roots of unity. Finally, assume that Tate–Shafarevich groups of elliptic curves over $\mathbb{Q}$ with j-invariant 0 are finite. Then the rational points of X are dense for the Zariski topology.

Remark 2.2. The restriction that $F(x, y) \in \mathbb{Z}[x, y]$ in Theorem 2.1 is not severe; see Remark 1.2(i). Also, the assumption that the coefficients of x^6 and y^6 are nonzero is not a restriction: it can be achieved with a suitable linear transformation, without so changing the isomorphism class of X.

We use Theorem 2.1 to deduce Theorem 1.1, which addresses the question of Zariski density of rational points for "diagonal" del Pezzo surfaces of degree 1 over $\mathbb{Q}$. We believe that the extraneous-looking hypotheses in Theorem 1.1, such as "$3A/B$ is not a rational square" or "$9 \nmid AB$," are not necessary. Our method of proof, however, breaks down without them. For example, if $(A, B) = (27, 16)$ then *all* the nonsingular fibers of the corresponding elliptic surface $\rho\colon \mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ have *positive* root number, and thus (conjecturally) even rank. In this particular example one can even show that all but finitely many fibers have rank at least 2, whence Zariski density of rational points on X is still true. However, if, for example, $(A, B) = (243, 16)$, then again all associated root numbers are positive, but we are unable to show rational points on X are Zariski dense (see Example 7.1 and Remark 7.4).

2B. ***Quartic twists and (mildly singular) del Pezzo surfaces of degree* 1.** Let $\rho\colon \mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ be an isotrivial rational elliptic surface and suppose that its generic fiber is of the form

$$Y^2 = X^3 + a(T)X, \quad a(T) \in \mathbb{Q}[t],\ \deg a(T) \le 4,$$

which can be thought of as a family of quartic twists over $\mathbb{Q}$. The associated hypersurface $X \subseteq \mathbb{P}_{\mathbb{Q}}(1, 1, 2, 3)$, given by

$$w^2 = z^3 + G(x, y)z, \quad G(x, y) := y^4 a(x/y),$$

is not smooth (and hence not a del Pezzo surface of degree 1). However, X is not too far from being smooth: for example, when G is squarefree, its singular locus consists of four A_2-singularities ($w = z = G(x, y) = 0$). We prove the following density result for this class of surfaces.

Theorem 2.3. *Let $G[x, y] \in \mathbb{Z}[x, y]$ be a squarefree homogeneous binary form of degree* 4; *assume that the coefficients of x^4 and y^4 are nonzero. Let X be the hypersurface given by*

$$w^2 = z^3 + G(x, y)z \tag{6}$$

in $\mathbb{P}_{\mathbb{Q}}(1, 1, 2, 3)$. Let c be the content of G and write $G(x, y) = cG_1(x, y)$ for some $G_1(x, y) \in \mathbb{Z}[x, y]$. Suppose that G_1 has no fixed prime divisors and that $G_1 = \prod_i g_i$, where the $g_i \in \mathbb{Z}[x, y]$ are irreducible homogeneous forms. Assume further that

$$\mu_4 \not\subseteq \mathbb{Q}[t]/g_i(t, 1) \quad \text{for some } i, \tag{7}$$

where μ_4 is the group of fourth roots of unity. Finally, assume that Tate–Shafarevich groups of elliptic curves over $\mathbb{Q}$ with j-invariant 1728 *are finite. Then the rational points of X are dense for the Zariski topology.*

Remark 2.4. The assumption that the coefficients of x^4 and y^4 are nonzero is not a restriction: it can be achieved with a suitable linear transformation, without so changing the isomorphism class of X.

Remark 2.5. Ulas [2007; 2008] studied the question of Zariski density of rational points on certain del Pezzo surfaces of degree 1 over $\mathbb{Q}$ by looking at explicit rational base-changes of their associated elliptic surfaces. His results do not depend on arithmetic conjectures and are thus stronger than ours, whenever there is an overlap — compare our Theorem 2.1 with Theorems 2.1 and 2.2 of [Ulas 2007] and our Theorem 2.3 with Theorems 3.1 and 3.2 of the same reference.

2C. ***Toward weak-weak approximation.*** Write Ω_k for the set of places of a number field k, and let k_v be the completion of k at $v \in \Omega_k$. Recall that a geometrically integral variety X over k satisfies *weak-weak approximation* if there exists a finite set $T \subseteq \Omega_k$ such that for every other finite set $S \subseteq \Omega_k$ with $S \cap T = \varnothing$, the image of the embedding

$$X(k) \hookrightarrow \prod_{v \in S} X(k_v)$$

is dense for the product topology of the v-adic topologies. We say that X satisfies *weak approximation* if we can take $T = \varnothing$.

It is known that del Pezzo surfaces of low degree need not satisfy weak approximation; see [Colliot-Thélène et al. 1987, Example 15.5; Swinnerton-Dyer 1962; Kresch and Tschinkel 2008, Example 2, Várilly-Alvarado 2008, Theorem 1.1] for counterexamples in degrees 4, 3, 2 and 1, respectively. It is believed, however, that these surfaces satisfy weak-weak approximation. More generally, a conjecture of Colliot-Thélène predicts that unirational varietes satisfy weak-weak approximation (the conjecture implies a positive solution to the inverse Galois problem over number fields); see [Serre 2008, p. 30]. Following a suggestion of Colliot-Thélène, we use our modified squarefree sieve to show that the surfaces of Theorems 2.1 and 2.3 satisfy a "surrogate" property that would be easily implied by weak-weak approximation. For analogous results in this direction on certain elliptic surfaces without section, see [Colliot-Thélène et al. 1998a], and for more general fibrations over the projective line, see [Colliot-Thélène et al. 1998b].

Theorem 2.6. *Let $\rho\colon \mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ be an elliptic surface associated to one of the hypersurfaces considered in either Theorem 2.1 or 2.3. Let $\mathscr{R}$ be the set of points $x \in \mathbb{P}^1(\mathbb{Q})$ such that the fiber $\mathscr{E}_x = \rho^{-1}(x)$ is an elliptic curve of positive Mordell–Weil rank. Assume that Tate–Shafarevich groups of elliptic curves over $\mathbb{Q}$ with j-invariant 0 or 1728 are finite. Then there exists a finite set of primes P_0, containing the infinite prime, such that for every finite set of primes P with $P \cap P_0 = \varnothing$, the image of the embedding*

$$\mathscr{R} \hookrightarrow \prod_{p \in P} \mathbb{P}^1(\mathbb{Q}_p)$$

is dense for the product topology of the p-adic topologies.

Remark 2.7. The set P_0 in Theorem 2.6 is effectively computed in the proof of the theorem.

2D. *Mazur's conjecture and related work.* Mazur has made a series of conjectures on the topology of rational points on varieties, including the following.

Conjecture 2.8 [Mazur 1992, Conjecture 4]. Let $\mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ be an elliptic surface with base $\mathbb{P}^1_{\mathbb{Q}}$. Then one of the following two conditions hold:

(1) for all but finitely many $t \in \mathbb{P}^1(\mathbb{Q})$, the fiber $\mathscr{E}_t$ is an elliptic curve with Mordell–Weil rank equal to zero,

(2) the set of $t \in \mathbb{P}^1(\mathbb{Q})$ such that $\mathscr{E}_t$ is an elliptic curve with positive Mordell–Weil rank is dense in $\mathbb{P}^1(\mathbb{R})$.

Many authors have shown since that (2) holds for a range of elliptic surfaces. In particular, the set $\mathscr{E}(\mathbb{Q})$ is dense in the Zariski topology for these surfaces. For example, in [Rohrlich 1993, Theorem 3] Rohrlich shows, unconditionally and using elementary methods, that if $f(t) \in \mathbb{Q}[t]$ is a quadratic polynomial, then the Kodaira–Néron model $\mathscr{E}$ of the elliptic curve over $\mathbb{Q}(T)$ given by

$$Y^2 = X^3 + af(T)^2X + bf(T)^3 \quad a, b \in \mathbb{Q}$$

satisfies part (2) of Conjecture 2.8, provided that there exists $t \in \mathbb{Q}$ such that $f(t) \neq 0$ and that $\mathscr{E}_t$ has positive Mordell–Weil rank. Munshi has recently extended this result to rational elliptic surfaces over real number fields, provided there are at least two fibers of positive rank and one fiber with a 2-torsion point defined over the ground field [Munshi 2010, Theorem 2].

Kuwata and Wang have a similar result to Rohrlich's for quadratic twists by cubic polynomials [Kuwata and Wang 1993]. The resulting isotrivial elliptic surfaces, however, are not rational; they are $K3$ surfaces. Munshi [2007] examined Conjecture 2.8 for many kinds of isotrivial rational elliptic surfaces, including cubic twists, by studying "horizontal" elliptic or conic bundle structures on these surfaces. There is surprisingly little overlap between Munshi's and our investigations; in fact, our methods cannot yield density results for cubic twists (the squarefreeness of $F(x, y)$ in (4) is central to our sieving argument). We have conditionally addressed the question of Zariski density of rational points on some of the isotrivial cases left open in [Munshi 2007, §7].

Assuming the parity conjecture, Manduchi has shown that conclusion (2) of Conjecture 2.8 holds for large families of *nonisotrivial* elliptic surfaces with base $\mathbb{P}^1_{\mathbb{Q}}$; see [Manduchi 1995]. Over a general number field, and assuming the Birch–Swinnerton-Dyer conjecture, as well as a conjecture of Deligne and Gross, Grant and Manduchi have shown that rational points are *potentially dense* for nonisotrivial elliptic surfaces over a rational or elliptic base; see [Grant and Manduchi 1997; 1998]. Ulas [2007, Theorems 5.1 and 5.3] has obtained density results on extensive families of rational nonisotrivial elliptic surfaces by studying explicit rational base changes (see Remark 2.5 as well). Helfgott [2004] has also obtained density results for elliptic surfaces through his study of average root numbers in families. His results depend on classical arithmetical conjectures.

Elkies (private communication, 2009) has suggested that Conjecture 2.8 is false; he has a heuristic which indicates that certain families of quadratic twists by a polynomial of high degree should yield counterexamples.

Colliot-Thélène, Swinnerton-Dyer and Skorobogatov study in [Colliot-Thélène et al. 1998a] the vertical Brauer–Manin obstruction of a large class of elliptic surfaces *without section*. In particular, they show that the set of rational points of the elliptic surfaces they study is dense for the Zariski topology as soon as it is nonempty. Their results are conditional on the finiteness of Tate–Shafarevich groups and Schinzel's hypothesis (a wild generalization of the twin primes conjecture).

3. Isotrivial elliptic surfaces and del Pezzo surfaces of degree 1

Let k be a number field, and let $(\mathscr{E}, \rho, \sigma)$ be an isotrivial rational elliptic surface with base $\mathbb{P}^1_k$. The generic fiber $E/k(T)$ of $\mathscr{E}$ is isomorphic to a curve in the list (i)–(iv) on page 661. Suppose that the sextic hypersurface $X \subseteq \mathbb{P}_k(1, 1, 2, 3)$ associated to $\mathscr{E}$ is smooth (and hence a del Pezzo surface of degree 1). Then a straightforward (albeit tedious) application of the Jacobian criterion shows that $E/k(T)$ must be a family of sextic twists (iv), with $f(T)$ squarefree. Alternatively, we may argue as follows. Since $X_{\bar{k}}$ is isomorphic to $\mathbb{P}^2_{\bar{k}}$ blown-up at 9 distinct points in general position [Manin and Hazewinkel 1974], it follows from [Shioda 1990, Theorem 10.11] that the Mordell–Weil lattice of $E_{\bar{k}(T)}$ has rank 8. From the Shioda–Tate formula [Shioda 1990, Corollary 5.3], we deduce that $\mathscr{E}_{\bar{k}}$ has no reducible fibers, i.e., the singular fibers of $\rho_{\bar{k}} : \mathscr{E}_{\bar{k}} \to \mathbb{P}^1_{\bar{k}}$ must be of type I_0 or II, in Kodaira's notation. The isotriviality of $\mathscr{E}$ precludes singular fibers of type I_0 (because these fibers are semistable). Looking at Persson's classification [1990] of rational elliptic surfaces, we conclude that $\mathscr{E}_{\bar{k}}$ must have six singular fibers of type II. A quick application of Tate's algorithm to the Kodaira–Néron models of the possible generic fibers from the list leaves (iv) as the only possibility, under the additional hypothesis that $f(T)$ is squarefree. We have thus shown:

Proposition 3.1. *Let k be a number field and let $(\mathscr{E}, \rho, \sigma)$ be an isotrivial rational elliptic surface with base $\mathbb{P}^1_k$. Suppose that the sextic hypersurface $X \subseteq \mathbb{P}^1_k(1, 1, 2, 3)$ associated to $\mathscr{E}$ is smooth. Then X is isomorphic to a hypersurface of the form*

$$w^2 = z^3 + F(x, y),$$

where $F(x, y)$ is a squarefree homogeneous form. □

4. Root numbers and flipping

Let E be an elliptic curve over $\mathbb{Q}$. The *root number* $W(E)$ of E is defined as a product of local factors

$$W(E) = \prod_{p \leq \infty} W_p(E),$$

where p runs over the rational prime numbers and infinity, $W_p(E) \in \{\pm 1\}$ and $W_p(E) = +1$ for all but finitely many p. The *local root number* $W_p(E)$ of E at p is defined in terms of epsilon factors of Weil–Deligne representations of $\mathbb{Q}_p$; it is an invariant of the isomorphism class of the base extension $E_{\mathbb{Q}_p}$ of E. For a definition of these local factors see [Deligne 1973; Tate 1979]. If p is a prime of good reduction for E then $W_p(E) = +1$; furthermore, $W_\infty(E) = -1$ (see [Rohrlich 1993]). The computation of $W_p(E)$ for primes of bad reduction in terms of data associated to a Weierstrass model of E has been studied by various authors; see particularly [Rohrlich 1993; Halberstadt 1998; Rizzo 2003]. In this section, we build on their work to give formulae for the root numbers of elliptic curves over $\mathbb{Q}$ of the form

$$y^2 = x^3 + \alpha \quad \text{and} \quad y^2 = x^3 + \alpha x \quad (\alpha \neq 0).$$

Our formula for the root number of $y^2 = x^3 + \alpha$ has a flavor different from that found in [Liverance 1995]; in particular, it is visibly insensitive to primes $p \geq 5$ whose *square* does not divide α.

Conjecturally, the root number $W(E)$ of an elliptic curve is the sign in the functional equation for the L-series $L(E, s)$ of E:

$$(2\pi)^{-s}\Gamma(s)N^{s/2}L(E, s) = W(E)(2\pi)^{2-s}\Gamma(2-s)N^{(2-s)/2}L(E, 2-s),$$

where N is the conductor of E, and $\Gamma(s)$ is the usual gamma function. According to the Birch–Swinnerton-Dyer conjecture,

$$W(E) = (-1)^{\operatorname{rank}(E)}. \tag{8}$$

Equality (8) is itself known as the *parity conjecture*. By [Nekovář 2001] and [Dokchitser and Dokchitser 2010] the finiteness of Tate–Shafarevich groups is enough to prove the parity conjecture.

Notation. In addition to the notation introduced above, we use the following conventions. Throughout, for a prime $p \in \mathbb{Z}$ we denote the corresponding p-adic valuation by v_p. If a is a nonzero integer then $\left(\frac{a}{p}\right)$ will denote the usual Legendre symbol; if m is an odd positive integer, $\left(\frac{a}{m}\right)$ will denote the usual Jacobi symbol.

4A. ***The root number of $E_\alpha : y^2 = x^3 + \alpha$.*** Let α be a nonzero integer. We give a closed formula for the root number of the elliptic curve $E_\alpha/\mathbb{Q} : y^2 = x^3 + \alpha$, in terms of α. Throughout, we write $W(\alpha)$ for this root number and $W_p(\alpha)$ for the local root number of E_α at p. We begin by determining $W_2(\alpha)$ and $W_3(\alpha)$.

Lemma 4.1. *Let α be a nonzero integer. Define α_2 and α_3 by*

$$\alpha = 2^{v_2(\alpha)}\alpha_2 = 3^{v_3(\alpha)}\alpha_3.$$

Then

$$W_2(\alpha) = \begin{cases} -1 & \textit{if } v_2(\alpha) \equiv 0 \textit{ or } 2 \bmod 6 \\ & \textit{or if } v_2(\alpha) \equiv 1, 3, 4 \textit{ or } 5 \bmod 6 \textit{ and } \alpha_2 \equiv 3 \bmod 4, \\ +1 & \textit{otherwise} \end{cases}$$

and

$$W_3(\alpha) = \begin{cases} -1 & \textit{if } v_3(\alpha) \equiv 1 \textit{ or } 2 \bmod 6 \textit{ and } \alpha_3 \equiv 1 \bmod 3, \\ & \textit{or if } v_3(\alpha) \equiv 4 \textit{ or } 5 \bmod 6 \textit{ and } \alpha_3 \equiv 2 \bmod 3, \\ & \textit{or if } v_3(\alpha) \equiv 0 \bmod 6 \textit{ and } \alpha_3 \equiv 5 \textit{ or } 7 \bmod 9 \\ & \textit{or if } v_3(\alpha) \equiv 3 \bmod 6 \textit{ and } \alpha_3 \equiv 2 \textit{ or } 4 \bmod 9, \\ +1 & \textit{otherwise.} \end{cases}$$

Proof. According to [Rizzo 2003, §1.1], to determine the local root number at p of an elliptic curve given in Weierstrass form, we must find the smallest vector with nonnegative entries

$$(a, b, c) := (v_p(c_4), v_p(c_6), v_p(\Delta)) + k(4, 6, 12) \tag{9}$$

for $k \in \mathbb{Z}$, where c_4, c_6 and Δ are the usual quantities associated to a Weierstrass equation (see [Silverman 1992, Chapter III]). For the curves in question we have

$$c_4 = 0, \quad c_6 = -2^5 \cdot 3^3 \cdot \alpha, \quad \text{and} \quad \Delta = -2^4 \cdot 3^3 \cdot \alpha^2,$$

whence

$$(v_p(c_4), v_p(c_6), v_p(\Delta)) = (\infty, v_p(\alpha), 2v_p(\alpha)) + \begin{cases} (0, 5, 4) & \text{if } p = 2, \\ (0, 3, 3) & \text{if } p = 3, \end{cases}$$

Now it is a simple matter of using the tables in [Rizzo 2003, §1.1] to compute local root numbers. We illustrate the computation of $W_2(\alpha)$ in one example. Suppose that $v_2(\alpha) \equiv 4 \bmod 6$. Then $(a, b, c) = (\infty, 3, 0)$, and according to the entries under $(\geq 4, 3, 0)$ in Rizzo's Table III, we have $W_2(\alpha) = -1$ if and only if $c_6' := c_6/2^{v_2(c_6)} \equiv 3 \bmod 4$, i.e., if and only if $\alpha_2 \equiv 3 \bmod 4$. All other local root number computations are similar and we omit the details. □

Remark 4.2. We take the opportunity to note that the entry $(\geq 5, 6, 9)$ in Table II of [Rizzo 2003] has a typo. The "special condition" should read $c_6' \not\equiv \pm 4 \bmod 9$.

The elliptic curve E_α has potential good reduction at every nonarchimedean place. We will use the following proposition, due to Rohrlich, which gives a formula for the local root numbers of an elliptic curve at primes $p \geq 5$ of potential good reduction.

Proposition 4.3 [Rohrlich 1993, Proposition 2]. *Let $p \geq 5$ be a rational prime, and let $E/\mathbb{Q}_p$ be an elliptic curve with potential good reduction. Write $\Delta \in \mathbb{Q}_p^*$ for the discriminant of any generalized Weierstrass equation for E over $\mathbb{Q}_p$. Let*

$$e := \frac{12}{\gcd(v_p(\Delta), 12)}.$$

Then

$$W_p(E) = \begin{cases} 1 & \textit{if } e = 1, \\ \left(\frac{-1}{p}\right) & \textit{if } e = 2 \textit{ or } 6, \\ \left(\frac{-3}{p}\right) & \textit{if } e = 3, \\ \left(\frac{-2}{p}\right) & \textit{if } e = 4. \end{cases}$$ □

Proposition 4.4 (Root numbers for $y^2 = x^3 + \alpha$). *Let α be a nonzero integer, and let*

$$R(\alpha) = W_2(\alpha)\left(\frac{-1}{\alpha_2}\right)W_3(\alpha)(-1)^{v_3(\alpha)}. \tag{10}$$

Then

$$W(\alpha) = -R(\alpha) \prod_{\substack{p^2 \mid \alpha \\ p \geq 5}} \begin{cases} 1 & \textit{if } v_p(\alpha) \equiv 0, 1, 3, 5 \bmod 6, \\ \left(\frac{-3}{p}\right) & \textit{if } v_p(\alpha) \equiv 2, 4 \bmod 6. \end{cases} \tag{11}$$

Let β be another nonzero integer, and suppose that $\alpha \equiv \beta \bmod 2^{v_2(\alpha)+2} \cdot 3^{v_3(\alpha)+2}$. Then $R(\alpha) = R(\beta)$.

Proof. Since $\Delta(E_\alpha) = -2^4 3^3 \alpha^2$, applying Proposition 4.3 we obtain

$$W(\alpha) = -W_2(\alpha)W_3(\alpha) \prod_{\substack{p \mid \alpha \\ p \geq 5}} \begin{cases} 1 & \text{if } v_p(\alpha) \equiv 0 \bmod 6, \\ \left(\frac{-1}{p}\right) & \text{if } v_p(\alpha) \equiv 1, 3, 5 \bmod 6, \\ \left(\frac{-3}{p}\right) & \text{if } v_p(\alpha) \equiv 2, 4 \bmod 6. \end{cases} \tag{12}$$

Let r be the product of the primes $p \geq 5$ such that $v_p(\alpha) = 1$, let $b = \alpha/r$ and set

$$\alpha_2 := \frac{\alpha}{2^{v_2(\alpha)}}, \quad b_2 := \frac{b}{2^{v_2(b)}}.$$

Note that $r = \alpha_2/b_2 = \alpha/b$. We may rewrite (12) as

$$W(\alpha) = -W_2(\alpha)W_3(\alpha)\left(\frac{-1}{r}\right) \prod_{\substack{p \mid b \\ p \geq 5}} \begin{cases} 1 & \text{if } v_p(\alpha) \equiv 0 \bmod 6, \\ \left(\frac{-1}{p}\right) & \text{if } v_p(\alpha) \equiv 1, 3, 5 \bmod 6, \\ \left(\frac{-3}{p}\right) & \text{if } v_p(\alpha) \equiv 2, 4 \bmod 6. \end{cases} \tag{13}$$

On the other hand, we have

$$\left(\frac{-1}{r}\right) = \left(\frac{-1}{\alpha_2/b_2}\right) = \left(\frac{-1}{\alpha_2}\right) \cdot \left(\frac{-1}{b_2}\right) = \left(\frac{-1}{\alpha_2}\right) \cdot \left(\frac{-1}{3}\right)^{v_3(\alpha)} \cdot \prod_{\substack{p \mid b \\ p \geq 5}} \left(\frac{-1}{p}\right)^{v_p(\alpha)},$$

so we can write (13) as

$$W(\alpha)= -W_2(\alpha)\Big(\frac{-1}{\alpha_2}\Big)W_3(\alpha)(-1)^{v_3(\alpha)}\prod_{\substack{p\mid b\\ p\geq 5}}\begin{cases}\Big(\frac{-1}{p}\Big)^{v_p(\alpha)} & \text{if } v_p(\alpha)\equiv 0 \bmod 6,\\ \Big(\frac{-1}{p}\Big)^{1+v_p(\alpha)} & \text{if } v_p(\alpha)\equiv 1,3,5 \bmod 6,\\ \Big(\frac{-3}{p}\Big)\Big(\frac{-1}{p}\Big)^{v_p(\alpha)} & \text{if } v_p(\alpha)\equiv 2,4 \bmod 6.\end{cases}$$

This reduces to

$$W(\alpha)=-R(\alpha)\prod_{\substack{p^2\mid\alpha\\ p\geq 5}}\begin{cases}1 & \text{if } v_p(\alpha)\equiv 0,1,3,5 \bmod 6,\\ \Big(\frac{-3}{p}\Big) & \text{if } v_p(\alpha)\equiv 2,4 \bmod 6.\end{cases}$$

as desired, because, for $p\geq 5$, we have $p\mid b \iff p^2\mid\alpha$.

To prove the last claim of the proposition, note that if

$$\alpha\equiv\beta \bmod 2^{v_2(\alpha)+2}\cdot 3^{v_3(\alpha)+2}$$

then $v_2(\alpha)=v_2(\beta)$ and $v_3(\alpha)=v_3(\beta)$; thus we have

$$\frac{\alpha}{2^{v_2(\alpha)}}\equiv\frac{\beta}{2^{v_2(\beta)}} \bmod 4 \quad\text{and}\quad \frac{\alpha}{3^{v_3(\alpha)}}\equiv\frac{\beta}{3^{v_3(\beta)}} \bmod 9.$$

The claim now follows from Lemma 4.1 □

The following corollary describes conditions on two nonzero integers α and β which guarantee that the elliptic curves $y^2=x^3+\alpha$ and $y^2=x^3+\beta$ have *opposite* root numbers. This is one of the key inputs to the proof of Theorem 2.1. This corollary is similar in spirit to [Manduchi 1995, Corollary 2.1].

Corollary 4.5 (Flipping I). *Let α,β be nonzero integers such that*

(1) $\alpha\equiv\beta \bmod 2^{v_2(\alpha)+2}\cdot 3^{v_3(\alpha)+2}$,

(2) $\alpha=c\ell$, *where ℓ is squarefree and* $\gcd(c,\ell)=1$,

(3) $\beta=cq^{2+6k}\eta$, *where η is square free*, $\gcd(c,\eta)=\gcd(q,c\eta)=1$, $k\geq 0$, $q\geq 5$ *is prime and* $q\equiv 2 \bmod 3$.

Then $W(\alpha)=-W(\beta)$.

Proof. The first condition ensures that $R(\alpha)=R(\beta)$. Since ℓ is squarefree and $\gcd(c,\ell)=1$, the only primes greater than 3 contributing to $W(\alpha)$ are those whose square divides c. Similarly, since η is squarefree and $\gcd(c,\eta)=\gcd(q,\eta)=1$, the only primes greater than 3 contributing to $W(\beta)$ are those whose square divides c, and q. Since $\gcd(q, c) = 1$, $q \geq 5$ and $q \equiv 2 \bmod 3$, we have

$$W(\beta) = \left(\frac{-3}{q}\right) W(\alpha) = -W(\alpha) \qquad \square$$

Remark 4.6. To prove Zariski density of rational points on the elliptic surface $\mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ associated to a del Pezzo of degree 1 as in Theorem 2.1, it is enough to do the following. First, prove that there exist infinite sets $\mathscr{F}_1$ and $\mathscr{F}_2$ of coprime pairs of integers such that whenever $(m_1, n_1) \in \mathscr{F}_1$ and $(m_2, n_2) \in \mathscr{F}_2$ then

(1) $\alpha := F(m_1, n_1)$ and $\beta := F(m_2, n_2)$ are nonzero integers, and

(2) the integers α and β satisfy the hypotheses of Corollary 4.5.

Then, by Corollary 4.5, we know that either

$$W(F(m, n)) = -1 \text{ for all } (m, n) \in \mathscr{F}_1,$$

or

$$W(F(m, n)) = -1 \text{ for all } (m, n) \in \mathscr{F}_2.$$

Hence, there are infinitely many closed fibers of $\mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ with negative root number. Assuming the parity conjecture, this gives an infinite number of closed fibers with infinitely many points, and hence a Zariski dense set of rational points on $\mathscr{E}$.

4B. ***The root number of $E_\alpha : y^2 = x^3 + \alpha x$.*** Next, we give a closed formula for the root number of the elliptic curve $E_\alpha/\mathbb{Q} : y^2 = x^3 + \alpha x$, in terms of the nonzero integer α. The proofs mirror those of Section 4A, and thus we have omitted them. Throughout this section, we write $W(\alpha)$ for the root number of E_α and $W_p(\alpha)$ for the local root number at p of E_α.

Lemma 4.7. *Let α be a nonzero integer. Define α_2 and α_3 by $\alpha = 2^{v_2(\alpha)}\alpha_2 = 3^{v_3(\alpha)}\alpha_3$. Then*

$$W_2(\alpha) = \begin{cases} -1 & \text{if } v_2(\alpha) \equiv 1 \text{ or } 3 \bmod 4 \text{ and } \alpha_2 \equiv 1 \text{ or } 3 \bmod 8 \\ & \text{or if } v_2(\alpha) \equiv 0 \bmod 4 \text{ and } \alpha_2 \equiv 1, 5, 9, 11, 13 \text{ or } 15 \bmod 16 \\ & \text{or if } v_2(\alpha) \equiv 2 \bmod 4 \text{ and } \alpha_2 \equiv 1, 3, 5, 7, 11 \text{ or } 15 \bmod 16, \\ +1 & \text{otherwise;} \end{cases}$$

$$W_3(\alpha) = \begin{cases} -1 & \text{if } v_3(\alpha) \equiv 2 \bmod 4, \\ +1 & \text{otherwise.} \end{cases}$$

Proof. Proceed as in the proof of Lemma 4.1, using the quantities

$$c_4 = -2^4 \cdot 3 \cdot \alpha, \quad c_6 = 0, \quad \text{and} \quad \Delta = -2^6 \cdot \alpha^3. \qquad \square$$

Proposition 4.8 (Root numbers for $y^2 = x^3 + \alpha x$). *Let α be a nonzero integer, and let*

$$R(\alpha) = W_2(\alpha)\Big(\frac{-1}{\alpha_2}\Big) W_3(\alpha)(-1)^{v_3(\alpha)}. \tag{14}$$

Then

$$W(\alpha) = -R(\alpha) \prod_{\substack{p^2|\alpha \\ p\geq 5}} \begin{cases} \Big(\frac{-1}{p}\Big) & \text{if } v_p(\alpha) \equiv 2 \bmod 4, \\ \Big(\frac{2}{p}\Big) & \text{if } v_p(\alpha) \equiv 3 \bmod 4. \end{cases}$$

Let β be another nonzero free integer, and suppose that $\alpha \equiv \beta \bmod 2^{v_2(\alpha)+4} \cdot 3^{v_3(\alpha)}$. Then $R(\alpha) = R(\beta)$. □

The following corollary, which parallels Corollary 4.5, describes conditions on two nonzero integers α and β that guarantee that the elliptic curves $y^2 = x^3 + \alpha x$ and $y^2 = x^3 + \beta x$ have *opposite* root numbers. This is one of the key inputs to the proof of Theorem 2.3.

Corollary 4.9 (Flipping II). *Let α, β be nonzero integers such that*

(1) $\alpha \equiv \beta \bmod 2^{v_2(\alpha)+4} \cdot 3^{v_3(\alpha)}$,

(2) $\alpha = c\ell$, *where ℓ is squarefree and* $\gcd(c, \ell) = 1$,

(3) $\beta = cq^{2+4k}\eta$, *where η is square free*, $\gcd(c, \eta) = \gcd(q, c\eta) = 1$, $k \geq 0$, *$q \geq 5$ is prime and $q \equiv 3 \bmod 4$; or $\beta = cp^{3+4k}\eta$, where η is square free*, $\gcd(c, \eta) = \gcd(q, c\eta) = 1$, $k \geq 0$, *$q \geq 5$ is prime and $q \equiv 3$ or $5 \bmod 8$.*

Then $W(\alpha) = -W(\beta)$. □

Remark 4.10. To prove Zariski density of rational points on the elliptic surface $\mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ associated to a sextic hypersurface as in Theorem 2.3, it is enough to do the following. First, prove that there exist infinite sets $\mathscr{F}_1$ and $\mathscr{F}_2$ of coprime pairs of integers such that whenever $(m_1, n_1) \in \mathscr{F}_1$ and $(m_2, n_2) \in \mathscr{F}_2$ then

(1) $\alpha := G(m_1, n_1)$ and $\beta := G(m_2, n_2)$ are nonzero integers.

(2) The integers α and β satisfy the hypotheses of Corollary 4.9.

Then, arguing as in Remark 4.6 (using Corollary 4.9) we find infinitely many closed fibers of $\mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ with negative root number. This gives a Zariski dense set of rational point for $\mathscr{E}$, assuming the parity conjecture.

5. The modified square-free sieve

In this section we present a variation of a squarefree sieve by Gouvêa and Mazur [1991] and Greaves [1992]. It is the tool that allows us to identify families of fibers with negative root numbers on certain elliptic surfaces.

Let $F(m, n) \in \mathbb{Z}[m, n]$ be a binary homogeneous form of degree d, not divisible by the square of a nonunit in $\mathbb{Z}[m, n]$. Write $F = \prod_{i=1}^{t} f_i$, where the

$f_i(m,n) \in \mathbb{Z}[m,n]$ are irreducible, and assume that $\deg f_i \le 6$ for all i. Applying a unimodular transformation we may (and do) assume that the coefficients of m^d and n^d in $F(m,n)$ are nonzero. Call their respective coefficients a_d and a_0. Write $F(m,n) = a_d \prod (m - \theta_i n)$, where the θ_i are algebraic numbers and $1 \le i \le d$. Let

$$\Delta(F) = \left| a_0 a_d^{2d-1} \prod_{i \ne j} (\theta_i - \theta_j) \right|;$$

this is essentially the discriminant of the form F. It is nonzero if and only if F contains no square factors.

Fix a positive integer M, as well as a subset $\mathcal{S}$ of $(\mathbb{Z}/M\mathbb{Z})^2$. Our goal is to count pairs of integers (m,n) such that $(m \bmod M, n \bmod M) \in \mathcal{S}$ and $F(m,n)$ is not divisible by p^2 for any prime number p such that $p \nmid M$. This will allow us to give an asymptotic formula for the number of pairs of integers (m,n) with $0 \le m, n \le x$ such that

$$F(m,n) = \nu \cdot \ell,$$

where ν is a *fixed* integer and ℓ is a squarefree integer such that $\gcd(\nu, \ell) = 1$. The case $\nu = 1$ is handled in [Gouvêa and Mazur 1991] under the additional assumption that $\deg f_i \le 3$, and extended in [Greaves 1992] to the case $\deg f_i \le 6$. We build upon their work to prove an asymptotic formula when $\nu > 1$.

Remark 5.1. The role of the set $\mathcal{S}$ above is to "decouple" the congruence conditions on (m,n) from the sieving process. This artifact, suggested to us by Bjorn Poonen after an initial reading of the manuscript, cleans up the analytic proofs in the main-term estimate for our sieve.

We make use of the following (mild variation of an) arithmetic function studied by Gouvêa and Mazur: put $\rho(1) = 1$, and for $k \ge 2$ let

$$\rho(k) = \#\{(m,n) \in \mathbb{Z}^2 : 0 \le m, n \le k-1,\ F(m,n) \equiv 0 \bmod k\}.$$

By the Chinese remainder theorem, the function ρ is multiplicative; i.e., if k_1 and k_2 are relatively prime positive integers then $\rho(k_1 k_2) = \rho(k_1)\rho(k_2)$.

Lemma 5.2 [Gouvêa and Mazur 1991, Lemma 3(2)]. *For fixed F as above and squarefree ℓ, we have $\rho(\ell^2) = O(\ell^2 \cdot d_k(\ell))$ as $\ell \to \infty$, where $k = \deg(F) + 1$ and $d_k(\ell)$ denotes the number of ways in which ℓ can be expressed as a product of k factors. In particular, $\rho(p^2) = O(p^2)$ as $p \to \infty$.* □

We can now state the main result of this section.

Theorem 5.3. *Let $F(m,n) \in \mathbb{Z}[m,n]$ be a homogeneous binary form of degree d. Assume that no square of a nonunit in $\mathbb{Z}[m,n]$ divides $F(m,n)$, and that no irreducible factor of F has degree greater than 6. Fix a positive integer M, as well as a subset $\mathcal{S}$ of $(\mathbb{Z}/M\mathbb{Z})^2$. Let $N(x)$ be the number of pairs of integers (m,n) with*

$0 \le m, n \le x$ such that $(m \bmod M, n \bmod M) \in \mathcal{S}$ and $F(m,n)$ is not divisible by p^2 for any prime p such that $p \nmid M$. Then

$$N(x) = Cx^2 + O\left(\frac{x^2}{(\log x)^{1/3}}\right) \quad \text{as } x \to \infty,$$

where

$$C = \frac{|\mathcal{S}|}{M^2} \prod_{p \nmid M} \left(1 - \frac{\rho(p^2)}{p^4}\right).$$

Remark 5.4. By Lemma 5.2, $\rho(p^2) = O(p^2)$ as $p \to \infty$ for a fixed F, so the infinite product defining C converges.

Heuristically, the condition that $F(m,n)$ be squarefree outside a prescribed integer is well approximated by the condition that $F(m,n)$ not be divisible by the square of a prime that is "small relative to x." More precisely, let $\xi = \frac{1}{3}\log x$ and define the principal term

$$N'(x) = \big\{(m,n) \in \mathbb{Z}^2 : 0 \le m, n \le x,\ F(m,n) \not\equiv 0 \bmod p^2 \text{ for all } p \le \xi,\ p \nmid M \\ \text{and } (m \bmod M, n \bmod M) \in \mathcal{S}\big\}.$$

Let $F = \prod_{i=1}^{t} f_i$ be a factorization of F into irreducible binary forms. Define the partial i-th error term $E_i(x)$ by

$$E_0(x) = \#\big\{(m,n) \in \mathbb{Z}^2 : 0 \le m, n \le x,\ \ p \mid m \text{ and } p \mid n \text{ for some } p > \xi\big\},$$
$$E_i(x) = \#\big\{(m,n) \in \mathbb{Z}^2 : 0 \le m, n \le x,\ \ p^2 \mid f_i(m,n) \text{ for some } p > \xi\big\}.$$

The proof of [Gouvêa and Mazur 1991, Proposition 2], essentially unchanged, shows that $E(x) := \sum_{i=0}^{t} E_i(x)$ gives an upper bound for the error term of our approximation, as follows.

Proposition 5.5. *If $\xi > \max\{\Delta(F), M\}$ then*

$$N'(x) - E(x) \le N(x) \le N'(x). \qquad \square$$

The proposition implies that

$$N(x) = N'(x) + O(E(x)),$$

which is why we think of ξ as giving us the notion of "small prime relative to x." The choice of $\frac{1}{3}\log x$ is somewhat flexible (see [Gouvêa and Mazur 1991, §4]); what is important is that when ℓ is a squarefree integer divisible only by primes *smaller* than ξ then

$$\ell \le \prod_{p<\xi} p = \exp\left(\sum_{p<\xi} \log p\right) \le e^{2\xi} = x^{2/3}, \tag{15}$$

where the last inequality follows from the estimate

$$\sum_{p<\xi} \log p \le \sum_{p<\xi} \log \xi = \pi(\xi) \log \xi < 2\xi,$$

with $\pi(x) = \#\{p \text{ prime} : p < x\}$; see [Stopple 2003, p. 105].

Greaves [1992] showed that

$$E(x) = O\left(\frac{x^2}{(\log x)^{1/3}}\right) \quad \text{as } x \to \infty.$$

His proof requires the hypothesis that no irreducible factor of F have degree greater than 6, which explains the presence of this hypothesis in Theorem 5.3. Thus Theorem 5.3 follows from the next lemma.

Lemma 5.6. *With C as in Theorem 5.3, we have*

$$N'(x) = Cx^2 + O\left(\frac{x^2}{\log x}\right) \quad \text{as } x \to \infty.$$

Proof. Let ℓ be a squarefree integer divisible only by primes smaller than ξ, and such that $\gcd(\ell, M) = 1$. Let

$$N_\ell(M, \mathcal{S}; x)$$

be the number of pairs of integers (m, n) such that

$$0 \le m, n \le x, \quad (m \bmod M, n \bmod M) \in \mathcal{S}, \quad \text{and} \quad F(m, n) \equiv 0 \bmod \ell^2.$$

For a fixed congruence class modulo ℓ^2 of solutions of $F(m_0, n_0) \equiv 0 \bmod \ell^2$, satisfying $(m_0 \bmod M, n_0 \bmod M) \in \mathcal{S}$, we count the number of representatives in the box $0 \le m, n \le x$, and obtain

$$N_\ell(M, \mathcal{S}; x) = \frac{x^2 \cdot |\mathcal{S}|}{M^2} \cdot \frac{\rho(\ell^2)}{\ell^4} + O\left(x \cdot \frac{\rho(\ell^2)}{\ell^2}\right),$$

where the implied constant depends on F, M and $\mathcal{S}$, but not on ℓ or x. By the inclusion-exclusion principle we have

$$N'(x) = \sum_{\ell} \mu(\ell) N_\ell(M, \mathcal{S}; x),$$

where μ denotes the usual Möbius function and the sum runs over squarefree integers that are divisible only by primes smaller than ξ and that are relatively prime to M. Thus, by (15),

$$N'(x) = \frac{x^2 \cdot |\mathcal{S}|}{M^2} \sum_{\ell} \mu(\ell) \frac{\rho(\ell^2)}{\ell^4} + O\biggl(x \sum_{\ell \le x^{2/3}} \frac{\rho(\ell^2)}{\ell^2}\biggr)$$
$$= \frac{x^2 \cdot |\mathcal{S}|}{M^2} \prod_{p<\xi,\ p \nmid M} \biggl(1 - \frac{\rho(p^2)}{p^4}\biggr) + O\biggl(x \sum_{\ell \le x^{2/3}} \frac{\rho(\ell^2)}{\ell^2}\biggr).$$

Assume that x is large enough so that $\xi > M$. Then, by Lemma 5.2, we have

$$\prod_{p \ge \xi} \biggl(1 - \frac{\rho(p^2)}{p^4}\biggr) = \prod_{p \ge \xi} \biggl(1 - O\Bigl(\frac{1}{p^2}\Bigr)\biggr) = 1 - \sum_{p \ge \xi} O\Bigl(\frac{1}{p^2}\Bigr)$$
$$= 1 - O\biggl(\int_{t \ge \xi} \frac{1}{t^2}\, dt\biggr) = 1 - O\Bigl(\frac{1}{\xi}\Bigr).$$

Hence

$$N'(x) = \frac{x^2 \cdot |\mathcal{S}|}{M^2} \prod_{p \nmid M} \biggl(1 - \frac{\rho(p^2)}{p^4}\biggr) + O\Bigl(\frac{x^2}{\xi}\Bigr) + O\biggl(x \sum_{\ell \le x^{2/3}} \frac{\rho(\ell^2)}{\ell^2}\biggr).$$

By Lemma 5.2, we have

$$O\biggl(x \sum_{\ell \le x^{2/3}} \frac{\rho(\ell^2)}{\ell^2}\biggr) = O\biggl(x \sum_{\ell \le x^{2/3}} d_k(\ell)\biggr) = O(x \cdot x^{2/3} \log^{k-1} x),$$

with $k = \deg F + 1$, where we have used the well-known fact that

$$\sum_{n \le x} d_k(n) = O(x \log^{k-1} x);$$

see, for example, [Iwaniec and Kowalski 2004, (1.80)]. Since $\xi = \frac{1}{3} \log x$, it follows that

$$N'(x) = \frac{x^2 \cdot |\mathcal{S}|}{M^2} \prod_{p \nmid M} \biggl(1 - \frac{\rho(p^2)}{p^4}\biggr) + O\Bigl(\frac{x^2}{\xi}\Bigr) + O(x \cdot x^{2/3} \log^{k-1} x)$$
$$= \frac{x^2 \cdot |\mathcal{S}|}{M^2} \prod_{p \nmid M} \biggl(1 - \frac{\rho(p^2)}{p^4}\biggr) + O\Bigl(\frac{x^2}{\log x}\Bigr),$$

which concludes the proof. □

5A. ***Making sure that C does not vanish.*** In this section we explore the possibility that the constant C for the principal term of $N(x)$ is zero. This will depend on the particular binary form $F(m,n)$, the integer M and the set $\mathcal{S}$. For any prime

$p \nmid M$, let

$$C_p = \left(1 - \frac{\rho(p^2)}{p^4}\right),$$

so that

$$C = \frac{|\mathcal{S}|}{M^2} \prod_{p \nmid M} C_p.$$

For $p \nmid M$ we know that $\rho(p^2) = O(p^2)$ (see Lemma 5.2); hence C vanishes if and only if either $\mathcal{S} = \varnothing$, or one of the factors C_p vanishes.

Lemma 5.7. *With notation as above, if $p \nmid M$ and $p \geq \deg F$, then $C_p \neq 0$.*

Proof. If $p \nmid M$ then $C_p = 0$ if and only if $\rho(p^2) = p^4$, which happens if and only if all pairs of integers (m, n) modulo $\mathbb{Z}/p^2\mathbb{Z}$ are solutions to $F(m, n) \equiv 0 \bmod p^2$. But then *all* pairs of integers (m, n) give solutions to the given congruence equation. This can happen only if $p < \deg(F)$; see the beginning of Section 2. □

5B. ***An application of the modified sieve.***

Corollary 5.8 (Pseudosquarefree sieve). *Let $F(m, n) \in \mathbb{Z}[m, n]$ be a homogeneous binary form of degree d. Assume that no square of a nonunit in $\mathbb{Z}[m, n]$ divides $F(m, n)$, and that no irreducible factor of F has degree greater than 6. Fix*

- *a sequence $S = (p_1, \ldots, p_r)$ of distinct prime numbers and*
- *a sequence $T = (t_1, \ldots, t_r)$ of nonnegative integers.*

Let M be an integer divisible by $p_1^{t_1+1} \cdots p_r^{t_r+1}$ and by p^2 for all primes $p < \deg F$. Suppose that there exist integers a, b such that

$$F(a, b) \not\equiv 0 \bmod p^2 \quad \textit{whenever } p \mid M \textit{ and } p \neq p_i \textit{ for any } i, \tag{16}$$

and such that

$$v_{p_i}(F(a, b)) = t_i \quad \textit{for all } i. \tag{17}$$

Then there are infinitely many pairs of integers (m, n) such that

$$m \equiv a \bmod M, \quad n \equiv b \bmod M, \tag{18}$$

and

$$F(m, n) = p_1^{t_1} \cdots p_r^{t_r} \cdot \ell,$$

where ℓ is squarefree and $v_{p_i}(\ell) = 0$ for all i.

Proof. Let $\mathcal{S} = \{(a, b)\}$. By Theorem 5.3, there are infinitely many pairs of integers (m, n) such that

$$m \equiv a \bmod M, \quad n \equiv b \bmod M, \quad F(m, n) \not\equiv 0 \bmod p^2 \quad \text{whenever } p \nmid M.$$

(Note that $|\mathscr{S}| = 1$ and $C \neq 0$ by Lemma 5.7.) Condition (16) then guarantees that $F(m, n)$ is not divisible by the square of any prime outside the sequence S. We also have

$$m \equiv a \bmod p_i^{t_i+1}, \quad n \equiv b \bmod p_i^{t_i+1}, \quad \text{for all } i,$$

because $p_i^{t_i+1} \mid M$ for all i, and hence

$$F(m, n) = F(a, b) \bmod p_i^{t_i+1} \quad \text{for all } i.$$

Using condition (17), we conclude that

$$v_{p_i}(F(m, n)) = t_i. \qquad \square$$

6. Proof of Theorems 2.1 and 2.3

For a finite extension L/k of number fields, we let $S(L/k)$ denote the set of unramified prime ideals of k that have a degree 1 prime over k in L. Given two sets A and B, we write $A \doteq B$ if A and B differ by finitely many elements, and we write $A \sqsubseteq B$ if $x \in A \implies x \in B$ with finitely many exceptions.

Proposition 6.1 (Bauer; see [Neukirch 1999, p. 548]). *Let k be a number field, N/k a Galois extension of k and M/k an arbitrary finite extension of k. Then*

$$S(M/k) \sqsubseteq S(N/k) \iff M \supseteq N.$$

Lemma 6.2. *Let $f(t) \in \mathbb{Z}[t]$ be an irreducible nonconstant polynomial, and let $N = \mathbb{Q}[t]/f(t)$. Let μ_3 denote the group of third roots of unity, and suppose that $\mathbb{Q}(\mu_3) \not\subseteq N$. Then there are infinitely many rational primes p such that $p \equiv 2 \pmod 3$ and such that there exists a degree-1 prime $\mathfrak{p} \subseteq N$ lying over p.*

Proof. Since $\mathbb{F}_p^\times$ contains an element of order 3 if and only if $3|(p-1)$, it follows that

$$S(\mathbb{Q}(\mu_3)/\mathbb{Q}) \doteq \{p \in \mathbb{Z} : p \text{ prime and } p \equiv 1 \bmod 3\}.$$

Suppose that the following implication holds (with possibly finitely many exceptions):

$$p \in \mathbb{Z} \text{ has a degree 1 prime in } N \implies p \equiv 1 \bmod 3.$$

Then

$$S(N/\mathbb{Q}) \sqsubseteq S(\mathbb{Q}(\mu_3)/\mathbb{Q}).$$

It follows from Proposition 6.1 that $\mathbb{Q}(\mu_3) \subseteq N$, a contradiction. $\square$

A similar argument proves the following entirely analogous lemma.

Lemma 6.3. *Let $g(t) \in \mathbb{Z}[t]$ be an irreducible nonconstant polynomial, and let $N = \mathbb{Q}[t]/g(t)$. Let μ_4 denote the group of fourth roots of unity, and suppose that $\mathbb{Q}(\mu_4) \not\subseteq N$. Then there are infinitely rational primes p such that $p \equiv 3 \pmod 4$ and such that there exists a degree-1 prime $\mathfrak{p} \subseteq N$ lying over p.* $\square$

Proof of Theorem 2.1. Since the surface in $\mathbb{P}_{\mathbb{Q}}(1,1,2,3)$ given by an equation of the form (4) is smooth (by the definition of a del Pezzo surface), it follows that F_1 is a squarefree binary form of degree 6 (see Section 3). Blowing up the anticanonical point $[0:0:1:1]$ of X we obtain an elliptic surface $\rho\colon \mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ whose fiber above $[m:n] \in \mathbb{P}^1(\mathbb{Q})$ is isomorphic to a curve in $\mathbb{P}^2_{\mathbb{Q}}$ whose affine equation is given by

$$y^2 = x^3 + F(m,n). \tag{19}$$

This is an elliptic curve for almost all $[m:n]$.

Write $c = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, where the p_i are distinct primes. Let $S = (p_1, \dots, p_r)$, $T = (0, \dots, 0)$ and let

$$M = (2 \cdot 3 \cdot 5)^3 \cdot (p_1 \cdots p_r).$$

Since $F_1(m,n)$ has no fixed prime divisors, we know that for each prime $p \mid M$ with $p \neq p_i$ for all i there exist congruence classes a_p, b_p modulo p^2 such that

$$F_1(a_p, b_p) \not\equiv 0 \bmod p^2.$$

Similarly, for a prime p_i in the sequence S there exist congruence classes a_{p_i}, b_{p_i} modulo p_i such that

$$F_1(a_{p_i}, b_{p_i}) \not\equiv 0 \bmod p_i;$$

in other words, $v_{p_i}(F_1(a_{p_i}, b_{p_i})) = 0$. By the Chinese remainder theorem there exist congruence classes a, b modulo M such that

$$(a,b) \equiv \begin{cases} (a_p, b_p) \bmod p^2 & \text{for all primes } p \text{ such that } p \mid M,\ p \neq p_i \text{ for any } i, \\ (a_{p_i}, b_{p_i}) \bmod p_i & \text{for all primes } p_i \text{ in the sequence } S. \end{cases} \tag{20}$$

By Corollary 5.8, applied to F_1, S, T, M, a and b as above, there is an infinite set $\mathscr{F}_1$ of pairs $(m,n) \in \mathbb{Z}^2$ such that

$$F_1(m,n) = \ell,$$

where ℓ is a squarefree integer with $\gcd(c, \ell) = 1$, by our choice of S and T. Note that the elements m, n of each pair must be coprime since $F_1(m,n)$ is squarefree. Furthermore, the congruence class of ℓ modulo $2^3 \cdot 3^3$ is fixed (by our choice of M) and nonzero (because ℓ is squarefree). Thus, for $(m,n) \in \mathscr{F}_1$ we have

$$F(m,n) = c\ell \qquad \gcd(c, \ell) = 1,$$

and the congruence class of $c\ell/2^{v_2(c\ell)}3^{v_3(c\ell)}$ modulo $2^2 \cdot 3^2$ is fixed and nonzero.

By Lemma 6.2, applied to a number field $N := \mathbb{Q}[t]/f_i(t,1)$ such that (5) holds, there is a rational prime $q \equiv 2 \bmod 3$ and a degree 1 prime $\mathfrak{q}$ in N lying over q. In fact, we may choose q so that $q > 5$, $\gcd(q, c) = 1$, and so that it does not divide the discriminant of $f_i(t,1)$.

We apply Corollary 5.8 again to $F_1(m, n)$. This time we let $S = (p_1, \ldots, p_r, q)$ and $T = (0, \ldots, 0, 2+6k)$, where k is a large positive integer[1]. Let

$$M = (2 \cdot 3 \cdot 5)^3 \cdot (p_1 \cdots p_r) \cdot q^{3+6k}.$$

We claim that there exist integers m_q, n_q such that

$$v_q(F_1(m_q, n_q)) = 2 + 6k.$$

Indeed, since q has a prime $\mathfrak{q}$ of degree 1 in N and it does not divide the discriminant of $f_i(t, 1)$, the equation

$$f_i(t, 1) = 0$$

has a simple root in $\mathbb{F}_q$. By Hensel's lemma, this solution lifts to a root in $\mathbb{Q}_q$. Hence $F_1(t, 1) = 0$ has a root in $\mathbb{Q}_q$. Approximating this solution by a rational number $r_q = m_q/n_q$ we can control $v_q(F_1(r_q, 1))$ modulo 6; i.e., there exists a pair $(m_q, n_q) \in \mathbb{Z}^2$ of coprime integers such that $v_q(F_1(m_q, n_q)) = 2 + 6k$ for some (possibly very large) positive integer k. By the Chinese remainder theorem, there exists a pair of integers (a, b) simultaneously satisfying (20) and

$$a \equiv m_q \bmod q^{3+6k}, \quad \text{and } b \equiv n_q \bmod q^{3+6k}. \tag{21}$$

By Corollary 5.8, applied to F_1, S, T, M, a and b as above, there is an infinite set $\mathcal{F}_2$ of pairs $(m, n) \in \mathbb{Z}^2$ such that

$$F_1(m, n) = q^{2+6k}\eta,$$

for some squarefree integer η with $\gcd(c, q\eta) = \gcd(q, \eta) = 1$, by our choice of S and T. Suppose that $(m, n) \in \mathcal{F}_2$. Then

$$F(m, n) = cq^{2+6k}\eta \quad \gcd(c, \eta) = \gcd(q, c\eta) = 1.$$

Furthermore, we claim that $\gcd(m, n) = 1$. To see this, note that since η is squarefree and F_1 is homogeneous of degree 6, then $\gcd(m, n)$ is some power of q; by (18), (21), and because $\gcd(m_q, n_q) = 1$, this power of q must be 1. As before, the congruence class of $cq^{2+6k}\eta/2^{v_2(c\eta)}3^{v_3(c\eta)} \bmod 2^2 \cdot 3^2$ is fixed, nonzero, and equal to that of $F_1(m, n)$ for $(m, n) \in \mathcal{F}_1$ (by our choice of a and b).

Whenever (19) is smooth, we write $W(F(m, n))$ for its root number. By Corollary 4.5, if $(m_1, n_1) \in \mathcal{F}_1$ and $(m_2, n_2) \in \mathcal{F}_2$ then

$$W(F(m_1, n_1)) = -W(F(m_2, n_2)).$$

Zariski density of rational points on X now follows by arguing as in Remark 4.6. □

[1] We will pick k large enough to ensure that $C \neq 0$ upon application of the pseudosquarefree sieve.

The proof of Theorem 2.3 is similar; we give enough details so that the interested reader can reconstruct it from the proof of Theorem 2.1.

Proof of Theorem 2.3. Blowing up the singular locus of X, as well as the base-point $[0:0:1:1]$ of $|-K_X|$, we obtain an elliptic surface $\rho\colon \mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ whose fiber above $[m:n] \in \mathbb{P}^1(\mathbb{Q})$ is isomorphic to a curve in $\mathbb{P}^2_{\mathbb{Q}}$ whose affine equation is given by

$$y^2 = x^3 + G(m,n)x, \tag{22}$$

which is an elliptic curve for almost all $[m:n]$.

We apply Corollary 5.8 twice, as in the proof of Theorem 2.1. First, we apply it to $G_1(m,n)$ by taking $S = (p_1, \dots, p_r)$, $T = (0, \dots, 0)$, where $c = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, and the p_i are distinct primes. We use

$$M = (2^2 \cdot 3)^3 \cdot (p_1 \cdots p_r).$$

This way we obtain an infinite set $\mathscr{F}_1$ of coprime pairs of integers (m,n) such that

$$G(m,n) = c\ell \quad \text{with} \ \ \gcd(c,\ell) = 1,$$

and the congruence class of $c\ell/2^{v_2(c\ell)}3^{v_3(c\ell)}$ modulo $2^4 \cdot 3^2$ is fixed and nonzero.

By Lemma 6.3, applied to a number field $N := \mathbb{Q}[t]/g_i(t,1)$ such that (7) holds, there is a rational prime $q \equiv 3 \bmod 4$ and a degree 1 prime $\mathfrak{q}$ in N lying over q. In fact, we may choose q so that $q > 5$, $\gcd(q,c) = 1$, and so that it does not divide the discriminant of $g_i(t,1)$.

We apply Corollary 5.8 again to $G_1(m,n)$ with $S = (p_1, \dots, p_r, q)$ and $T = (0, \dots, 0, 2 + 4k)$, where k is a large positive integer, and

$$M = (2^2 \cdot 3)^3 \cdot (p_1 \cdots p_r) \cdot q^{3+4k}$$

Using Hensel's lemma as in the proof of Theorem 2.1, we obtain a different infinite set $\mathscr{F}_2$ of coprime pairs integers (m,n) such that

$$G(m,n) = cq^{2+4k}\eta \quad \text{with} \ \ \gcd(c,\eta) = \gcd(q,c\eta) = 1,$$

where η is a squarefree integer. As before, the congruence class of

$$cq^{2+4k}\eta/2^{v_2(c\eta)}3^{v_3(c\eta)}$$

modulo $2^4 \cdot 3^2$ is fixed, nonzero, and equal to that of $G_1(m,n)$ for $(m,n) \in \mathscr{F}_1$ (by our choice of a and b).

Whenever (22) is smooth, we write $W(G(m,n))$ for its root number. By Corollary 4.9, if $(m_1,n_1) \in \mathscr{F}_1$ and $(m_2,n_2) \in \mathscr{F}_2$ then

$$W(G(m_1,n_1)) = -W(G(m_2,n_2)).$$

Zariski density of rational points on X now follows by arguing as in Remark 4.10. $\square$

7. Diagonal del Pezzo surfaces of degree 1

We begin this section with two examples of del Pezzo surfaces of degree 1 that show how the sieving technique used in the proof of Theorems 2.1 and 2.3 can fail. In one case, however, we can show that rational points are Zariski dense, by exhibiting explicit nontorsion sections of the associated elliptic surfaces.

Example 7.1. Consider the del Pezzo surface of degree 1 given by

$$w^2 = z^3 + 27x^6 + 16y^6$$

in $\mathbb{P}_{\mathbb{Q}}(1, 1, 2, 3)$. Let $\rho\colon \mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ be its associated elliptic fibration. The elliptic curve $E_{m,n}$ above the point $[m : n] \in \mathbb{P}^1(\mathbb{Q})$ is given by

$$E_{m,n}\colon \quad y^2 = x^3 + 27m^6 + 16n^6.$$

We claim that $W(E_{m,n}) = +1$ *for all* $[m : n] \in \mathbb{P}^1(\mathbb{Q})$. We may assume that $\gcd(m, n) = 1$. Let $\alpha = 27m^6 + 16n^6$, and suppose that $p \geq 5$ divides α (in particular, $p \nmid m$). Then

$$-3 \equiv (4n^3/3m^3)^2 \bmod p,$$

and thus $\left(\frac{-3}{p}\right) = 1$; hence the product over $p^2 \mid \alpha$ in (11) is equal to 1. In the notation of Proposition 4.4, it remains to see that $R(\alpha) = -1$. Since $\gcd(m, n) = 1$, we have $v_2(\alpha) = 4$ or 0, according to whether $2 \mid m$ or not. In either case, using Lemma 4.1, we see that

$$W_2(\alpha) \cdot \left(\frac{-1}{\alpha_2}\right) = 1 \quad \text{for all } \alpha.$$

Similarly, $v_3(\alpha) = 0$ or 3 according to whether $3 \nmid n$ or not. By Lemma 4.1 it also follows that

$$W_3(\alpha) \cdot (-1)^{v_3(\alpha)} = -1 \quad \text{for all } \alpha,$$

and hence $R(\alpha) = -1$, as desired.

The flipping technique of Corollary 4.5 thus cannot possibly work! Furthermore, assuming the parity conjecture, it follows that $E_{m,n}$ has even Mordell–Weil rank for all $[m : n] \in \mathbb{P}^1(\mathbb{Q})$. In fact, we claim that *all but finitely many* fibers have even rank ≥ 2. To see this note the family contains the points

$$(-3m^2, 4n^3) \quad \text{and} \quad \left(\frac{9m^4}{4n^2}, \frac{27m^6}{8n^3} + 4n^3\right).$$

We can check that these points are independent on the fiber above $[m : n] = [1 : 1]$, and thus they are independent as points on the generic fiber of $\mathscr{E}$. Then Silverman's specialization theorem [1994, Theorem 11.4] shows that the points are independent for all but finitely many pairs (m, n). Hence, rational points are Zariski dense on the original del Pezzo surface[2].

Example 7.2. Consider the del Pezzo surface of degree 1 given by

$$w^2 = z^3 + 6(27x^6 + y^6)$$

in $\mathbb{P}_{\mathbb{Q}}(1, 1, 2, 3)$. The elliptic curve $E_{m,n}$ above a point $[m : n] \subseteq \mathbb{P}^1(\mathbb{Q})$ of the associated elliptic surface $\mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ is given by

$$E_{m,n}: \quad y^2 = x^3 + 6(27m^6 + n^6).$$

As in Example 7.1 we can show that $W(E_{m,n}) = +1$ for all $[m : n] \in \mathbb{P}^1(\mathbb{Q})$. However, we cannot find readily available sections; Zariski density of rational points on this surface remains an open question.

The key point behind both of examples above is that condition (5) on the form $F_1(m, n)$ fails. The following lemma gives a necessary condition for the failure of (5) to occur, and suggests how to find the above examples.

Lemma 7.3. *Let* $F_1(m, n) = Am^6 + Bn^6 \in \mathbb{Z}[m, n]$, *and assume that* $\gcd(A, B) = 1$. *Write* $F_1 = \prod_i f_i$, *where the* $f_i \in \mathbb{Z}[m, n]$ *are irreducible homogeneous forms. Let* μ_3 *denote the group of third roots of unity. Then*

$$\mu_3 \subseteq \mathbb{Q}[t]/f_i(t, 1) \text{ for all } i \implies 3A/B \text{ is a rational square.} \tag{23}$$

Proof. The proof is an exercise in Galois theory. We will prove the case where F_1 is irreducible to illustrate the method. Choose a sixth root ξ of $-B/A$ and an isomorphism $\mathbb{Q}[t]/(At^6 + B) \xrightarrow{\sim} \mathbb{Q}(\xi)$. Suppose that $\mathbb{Q}(\mu_3) \subseteq \mathbb{Q}(\xi)$, so that $\mathbb{Q}(\xi)/\mathbb{Q}$ is a Galois extension of degree 6. Its unique quadratic subextension is $\mathbb{Q}(\mu_3) = \mathbb{Q}(\sqrt{-3})$, hence

$$\xi^3 = a + b\sqrt{-3} \quad \text{for some } a, b \in \mathbb{Q}.$$

Squaring both sides of the above equation and rearranging we obtain

$$-B/A - a^2 + 3b^2 = 2ab\sqrt{-3}$$

so that $ab = 0$. Since $\xi^3 \notin \mathbb{Q}$, it follows that $a = 0$ and $B/A = 3b^2$. □

If $3A/B$ is a rational square, it is often the case that not all fibers of the associated elliptic surface have positive root number: the 2-adic and 3-adic part of $Am^6 + Bn^6$ may vary enough to guarantee the existence of infinitely many fibers with root number -1. This idea, together with Theorem 2.1, are the necessary ingredients in the proof of Theorem 1.1.

[2]In fact, this surface is not minimal. The two nontorsion sections of $\mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ correspond to exceptional curves on X that are defined over $\mathbb{Q}$. Contracting these curves gives a del Pezzo surface of degree 3 with a rational point. This surface is unirational by the Segre–Manin Theorem.

Proof of Theorem 1.1. Let $F(x, y) = Ax^6 + By^6$ and put $c = \gcd(A, B)$. Write $F_1(x, y) = A_1x^6 + B_1y^6$, where $cA_1 = A$ and $cB_1 = B$. One easily checks that F_1 has no fixed prime factors. Write $F_1 = \prod_i f_i$, where the $f_i \in \mathbb{Z}[x, y]$ are irreducible homogeneous forms. If $3A/B$ is not a rational square then it follows from Lemma 7.3 that

$$\mu_3 \not\subseteq \mathbb{Q}[t]/f_i(t, 1) \quad \text{for some } i,$$

so by Theorem 2.1, $X(\mathbb{Q})$ is Zariski dense in X.

If, on the other hand, $3A/B$ is a rational square, then by assumption $c = 1$ and $9 \nmid AB$. After possibly interchanging A and B, we may write $A = 3a^2$ and $B = b^2$ for some relatively prime $a, b \in \mathbb{Z}$ not divisible by 3. A smooth fiber above $[m : n] \in \mathbb{P}^1(\mathbb{Q})$ of the elliptic surface $\mathscr{E} \to \mathbb{P}^1_{\mathbb{Q}}$ associated to X is the plane curve

$$E_\alpha : \quad y^2 = x^3 + \alpha,$$

where $\alpha = 3a^2m^6 + b^2n^6$. Arguing as in Example 7.1 we see that the product over $p^2 \mid \alpha$ in (11) is equal to 1.

To conclude the proof, it suffices to show that there are infinitely many pairs (m, n) of relatively prime integers such that $R(\alpha) = 1$ (see Proposition 4.4 for the definition of $R(\alpha)$). To construct such pairs (m, n), first suppose that $3 \mid n$ (whence $3 \nmid m$). Then $v_3(\alpha) = 1$ and $\alpha_3 \equiv 1 \bmod 3$, so by Lemma 4.1

$$W_3(\alpha) \cdot (-1)^{v_3(\alpha)} = (-1) \cdot (-1) = 1.$$

Next, we compute the product

$$w_2 := W_2(\alpha)\Big(\frac{-1}{\alpha_2}\Big).$$

We proceed by analyzing two cases, according to the 2-adic valuation of b, which we may assume is either 0, 1 or 2. We use Lemma 4.1 to compute the local root number at 2:

(1) $v_2(b) = 0$: choose n even. Then, regardless of the value of $v_2(a)$ (which we may also assume is 0, 1 or 2), we obtain $v_2(\alpha)$ even and $\alpha_2 \equiv 3 \bmod 4$, whence $w_2 = 1$.

(2) $v_2(b) = 1$ or 2: choose m odd, so that $v_2(\alpha) = 0$ and $\alpha_2 \equiv 3 \bmod 4$, whence $w_2 = 1$.

In any case, there are infinitely many pairs $(m, n) \in \mathbb{Z}^2$ with $R(3a^2m^6 + b^2n^6) = 1$, as desired. □

Remark 7.4. If $3A/B$ is a rational square, and either $\gcd(A, B) \neq 1$ or $9 \mid AB$, then it can happen that all the elliptic curves that are fibers of the rational surface associated to X have root number $+1$ (see Examples 7.1 and 7.2). Even when $9 \mid AB$ there are examples of surfaces, such as

$$w^2 = z^3 + 3^5 x^6 + 2^4 y^6,$$

where we were not able to find nontorsion sections.

8. Proof of Theorem 2.6

We carry out the details for the case of a surface X as in Theorem 2.1, the other case being similar. The fiber of ρ above $[m : n] \in \mathbb{P}^1(\mathbb{Q})$ is isomorphic to the plane curve

$$y^2 = x^3 + F(m, n) \tag{24}$$

which is an elliptic curve for almost all $[m : n]$. As in Theorem 2.1, we write c for the content of F and $F_1(m, n) := (1/c)F(m, n)$. By Lemma 6.2, applied to a number field $N := \mathbb{Q}[t]/f_i(t, 1)$ such that (5) holds, there is a rational prime $q \equiv 2 \bmod 3$ and a prime $\mathfrak{q}$ in N lying over q of degree 1 over $\mathbb{Q}$. We may assume that $q > 5$, $\gcd(c, q) = 1$, and that q does not divide the discriminant of $f_i(t, 1)$. Write $c = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, where the p_i are distinct primes. Let $P_0 = \{2, 3, 5, p_1, \cdots, p_r, q, \infty\}$.

Fix a finite set of distinct primes $P = \{q_1 \ldots, q_s\}$ such that $P \cap P_0 = \varnothing$, as well as a point $[m_p : n_p] \in \mathbb{P}^1(\mathbb{Q}_p)$ for each $p \in P$. We may assume that $m_p, n_p \in \mathbb{Z}_p$, and without loss of generality[3] we will further assume that $n_p \in \mathbb{Z}_p^\times$ for every $p \in P$. Let $\epsilon > 0$ be given and choose an integer N large so that

$$1/p^N < \epsilon \quad \text{and} \quad v_p(F_1(m_p, n_p)) < N \quad \text{for every } p \in P. \tag{25}$$

Let

$$\begin{aligned} S &= (p_1, \ldots, p_r, q_1, \ldots, q_s), \\ T &= \bigl(0, \ldots, 0, v_{q_1}(F_1(m_{q_1}, n_{q_1})), \ldots, v_{q_s}(F_1(m_{q_s}, n_{q_s}))\bigr), \end{aligned}$$

and let

$$M = (2 \cdot 3 \cdot 5)^3 \cdot (p_1 \cdots p_r) \cdot (q_1 \cdots q_s)^N.$$

Since $F_1(m, n)$ has no fixed prime factors, for any prime $p \mid M$ such that $p \neq p_i$ for all i and $p \notin P$, there exist congruence classes a_p, b_p modulo p^2 such that

$$F_1(a_p, b_p) \not\equiv 0 \bmod p^2.$$

Similarly, for a prime p_i with $1 \leq i \leq r$, there exist congruence classes a_{p_i}, b_{p_i} modulo p_i such that

$$F_1(a_{p_i}, b_{p_i}) \not\equiv 0 \bmod p_i.$$

[3]In fact, we may only really assume that either $m_p \in \mathbb{Z}_p^\times$ or $n_p \in \mathbb{Z}_p^\times$. We can interchange the roles of m_p and n_p in any one step of the proof without much difficulty, so the assumption that $n_p \in \mathbb{Z}_p^\times$ is an artifact to clean up the details of the proof.

By the Chinese remainder theorem there exist congruence classes a, b modulo M such that

$$(a,b) \equiv \begin{cases} (a_p, b_p) \bmod p^2 & \text{for primes } p \text{ such that } p \mid M,\ p \notin P, \text{ and} \\ & \qquad\qquad\qquad\qquad\qquad p \neq p_i \text{ for all } i, \\ (a_{p_i}, b_{p_i}) \bmod p_i & \text{for primes } p_i \text{ with } 1 \le i \le r, \\ (m_p, n_p) \bmod p^N & \text{for primes } p \in P. \end{cases} \tag{26}$$

By construction,

$$F_1(a,b) \equiv F_1(m_p, n_p) \bmod p^N \quad \text{for all } p \in P.$$

It follows from (25) that

$$v_p(F_1(a,b)) = v_p(F_1(m_p, n_p)) \quad \text{for all } p \in P.$$

By Corollary 5.8, applied to F_1, S, T, M, a, b as above, there is an infinite set $\mathscr{F}_1$ of pairs $(m,n) \in \mathbb{Z}^2$ such that

$$F_1(m,n) = \ell,$$

where ℓ is a squarefree integer with $\gcd(c,\ell) = 1$, by our choice of S and T. Furthermore, the congruence class of ℓ modulo $2^3 \cdot 3^3$ is fixed (by our choice of M) and nonzero (because ℓ is squarefree). Thus, for $(m,n) \in \mathscr{F}_1$ we have

$$F(m,n) = c\ell \quad \gcd(c,\ell) = 1,$$

and the congruence class of $c\ell/2^{v_2(c\ell)}3^{v_3(c\ell)}$ modulo $2^2 \cdot 3^2$ is fixed and nonzero.

We apply Corollary 5.8 again to $F_1(m,n)$. This time we let

$$\begin{aligned} S &= (p_1, \ldots, p_r, q_1, \ldots, q_s, q), \\ T &= \big(0, \ldots, 0, v_{q_1}(F_1(m_{q_1}, n_{q_1})), \ldots, v_{q_s}(F_1(m_{q_s}, n_{q_s})), 2+6k\big), \end{aligned}$$

where k is a large positive integer (large enough to ensure that $C \neq 0$ upon application of the sieve), and we let

$$M = (2 \cdot 3 \cdot 5)^3 \cdot (p_1 \cdots p_r) \cdot (q_1 \cdots q_s)^N \cdot q^{3+6k}.$$

Arguing as in the proof of Theorem 2.1, using Hensel's lemma and Lemma 6.2, we can show that there exist integers a_q, b_q such that

$$v_q(F_1(a_q, b_q)) = 2 + 6k$$

for some large positive integer k. By the Chinese remainder theorem, there exist congruence classes a, b modulo M such that (26) holds, and in addition

$$a \equiv a_q \bmod q^{3+6k} \quad \text{and} \quad b \equiv b_q \bmod q^{3+6k}.$$

By Corollary 5.8 there is an infinite set $\mathcal{F}_2$ of pairs $(m,n) \in \mathbb{Z}^2$ such that

$$F_1(m,n) = q^{2+6k}\eta, \tag{27}$$

where η is a squarefree integer such that $\gcd(c,\eta) = \gcd(q, c\eta) = 1$ (by the choice of S and T). In summary, for $(m,n) \in \mathcal{F}_2$, we have

$$F(m,n) = cq^{2+6k}\eta \quad \text{with} \quad \gcd(c,\eta) = \gcd(q,c\eta) = 1,$$

and the congruence class of $cq^{2+6k}\eta/2^{v_2(c\eta)} \bmod 2^2 \cdot 3^2$ is fixed, nonzero, and equal to that of $F_1(m,n)$ for $(m,n) \in \mathcal{F}_1$.

Whenever (24) is smooth, we write $W(F(m,n))$ for its root number. By Corollary 4.5, if $(m_1,n_1) \in \mathcal{F}_1$ and $(m_2,n_2) \in \mathcal{F}_2$, then

$$W(F(m_1,n_1)) = -W(F(m_2,n_2)).$$

Hence, there exists a pair $(m_0,n_0) \in \mathcal{F}_1 \cup \mathcal{F}_2$ such that $W(F(m_0,n_0)) = -1$. By the assumption that Tate–Shafarevich groups are finite we conclude that the fiber of ρ above $[m_0 : n_0]$ has positive Mordell–Weil rank, i.e., $[m_0 : n_0] \in \mathcal{R}$. By construction, $n_0 \neq 0$, and

$$m_0 \equiv m_p \bmod p^N, \quad \text{and} \quad n_0 \equiv n_p \bmod p^N \quad \text{for all } p \in P.$$

Hence

$$\left| \frac{m_p}{n_p} - \frac{m_0}{n_0} \right|_p = |m_p n_0 - m_0 n_p|_p \leq \frac{1}{p^N} < \epsilon \quad \text{for all } p \in P,$$

and $[m_0 : n_0]$ is arbitrarily close to $[m_p : n_p]$ for all $p \in P$. This concludes the proof of the theorem. □

Acknowledgements

I thank my thesis advisor, Bjorn Poonen, for numerous helpful conversations and for suggestions following a careful reading of the manuscript. I thank David Zywina for useful conversations on sieves. I thank Brendan Hassett and Cecilia Salgado for helpful conversations on elliptic surfaces. I thank Jean-Louis Colliot-Thélène for suggesting that I use the methods in this paper to prove Theorem 2.6. Finally, I thank the anonymous referee for several useful suggestions.

References

[Colliot-Thélène et al. 1987] J.-L. Colliot-Thélène, J.-J. Sansuc, and P. Swinnerton-Dyer, "Intersections of two quadrics and Châtelet surfaces, II", *J. Reine Angew. Math.* **374** (1987), 72–168. MR 88m:11045b Zbl 0622.14030

[Colliot-Thélène et al. 1998a] J.-L. Colliot-Thélène, A. N. Skorobogatov, and P. Swinnerton-Dyer, "Hasse principle for pencils of curves of genus one whose Jacobians have rational 2-division points", *Invent. Math.* **134**:3 (1998), 579–650. MR 99k:11095 Zbl 0924.14011

[Colliot-Thélène et al. 1998b] J.-L. Colliot-Thélène, A. N. Skorobogatov, and P. Swinnerton-Dyer, "Rational points and zero-cycles on fibred varieties: Schinzel's hypothesis and Salberger's device", *J. Reine Angew. Math.* **495** (1998), 1–28. MR 99i:14027 Zbl 0883.11029

[Conrad et al. 2005] B. Conrad, K. Conrad, and H. Helfgott, "Root numbers and ranks in positive characteristic", *Adv. Math.* **198**:2 (2005), 684–731. MR 2006m:11080 Zbl 1113.11033

[Deligne 1973] P. Deligne, "Les constantes des équations fonctionnelles des fonctions L", pp. 501–597 in *Modular functions of one variable, II* (Antwerp, 1972), edited by P. Deligne and W. Kuyk, Lecture Notes in Math. **349**, Springer, Berlin, 1973. MR 50 #2128 Zbl 0271.14011

[Dokchitser and Dokchitser 2010] T. Dokchitser and V. Dokchitser, "On the Birch–Swinnerton-Dyer quotients modulo squares", *Ann. of Math.* (2) **172**:1 (2010), 567–596. MR 2011h:11069 Zbl 1223.11079

[Gouvêa and Mazur 1991] F. Gouvêa and B. Mazur, "The square-free sieve and the rank of elliptic curves", *J. Amer. Math. Soc.* **4**:1 (1991), 1–23. MR 92b:11039 Zbl 0725.11027

[Grant and Manduchi 1997] G. R. Grant and E. Manduchi, "Root numbers and algebraic points on elliptic surfaces with base $\mathbf{P}^1$", *Duke Math. J.* **89**:3 (1997), 413–422. MR 99a:14028 Zbl 0907.14012

[Grant and Manduchi 1998] G. R. Grant and E. Manduchi, "Root numbers and algebraic points on elliptic surfaces with elliptic base", *Duke Math. J.* **93**:3 (1998), 479–486. MR 99m:11071 Zbl 1029.11028

[Greaves 1992] G. Greaves, "Power-free values of binary forms", *Quart. J. Math. Oxford Ser.* (2) **43**:169 (1992), 45–65. MR 92m:11098 Zbl 0768.11034

[Halberstadt 1998] E. Halberstadt, "Signes locaux des courbes elliptiques en 2 et 3", *C. R. Acad. Sci. Paris Sér. I Math.* **326**:9 (1998), 1047–1052. MR 99k:11082 Zbl 0933.11030

[Helfgott 2004] H. A. Helfgott, "On the behaviour of root numbers in families of elliptic curves", preprint, 2004. arXiv math/0408141

[Iskovskih 1979] V. A. Iskovskih, "Minimal models of rational surfaces over arbitrary fields", *Izv. Akad. Nauk SSSR Ser. Mat.* **43**:1 (1979), 19–43, 237. In Russian; translated in *Math. USSR Izv.* **14**:1 (1980), 17–39. MR 80m:14021

[Iwaniec and Kowalski 2004] H. Iwaniec and E. Kowalski, *Analytic number theory*, AMS Colloquium Publications **53**, American Mathematical Society, Providence, RI, 2004. MR 2005h:11005 Zbl 1059.11001

[Kollár 1996] J. Kollár, *Rational curves on algebraic varieties*, Ergebnisse der Math. **32**, Springer, Berlin, 1996. MR 98c:14001 Zbl 0877.14012

[Kresch and Tschinkel 2008] A. Kresch and Y. Tschinkel, "Two examples of Brauer–Manin obstruction to integral points", *Bull. Lond. Math. Soc.* **40**:6 (2008), 995–1001. MR 2009k:14039 Zbl 1161.14019

[Kuwata and Wang 1993] M. Kuwata and L. Wang, "Topology of rational points on isotrivial elliptic surfaces", *Internat. Math. Res. Notices* **1993**:4 (1993), 113–123. MR 94a:11079 Zbl 0804.14008

[Liverance 1995] E. Liverance, "A formula for the root number of a family of elliptic curves", *J. Number Theory* **51**:2 (1995), 288–305. MR 96e:11086 Zbl 0831.14012

[Manduchi 1995] E. Manduchi, "Root numbers of fibers of elliptic surfaces", *Compositio Math.* **99**:1 (1995), 33–58. MR 96j:11076 Zbl 0878.14028

[Manin and Hazewinkel 1974] Y. I. Manin and M. Hazewinkel, *Cubic forms: algebra, geometry, arithmetic*, North-Holland, Amsterdam, 1974. MR 57 #343 Zbl 0582.14010

[Mazur 1992] B. Mazur, "The topology of rational points", *Experiment. Math.* **1**:1 (1992), 35–45. MR 93j:14020 Zbl 0784.14012

[Munshi 2007] R. Munshi, "Density of positive rank fibers in elliptic fibrations", *J. Number Theory* **125**:1 (2007), 254–266. MR 2008d:11051 Zbl 1126.11031

[Munshi 2010] R. Munshi, "Density of positive rank fibers in elliptic fibrations, II", *Int. J. Number Theory* **6**:1 (2010), 15–23. MR 2011m:11118 Zbl 05687122

[Nekovář 2001] J. Nekovář, "On the parity of ranks of Selmer groups, II", *C. R. Acad. Sci. Paris Sér. I Math.* **332**:2 (2001), 99–104. MR 2002e:11060 Zbl 1090.11037

[Neukirch 1999] J. Neukirch, *Algebraic number theory*, Grundlehren der Math. Wiss. **322**, Springer, Berlin, 1999. MR 2000m:11104 Zbl 0956.11021

[Persson 1990] U. Persson, "Configurations of Kodaira fibers on rational elliptic surfaces", *Math. Z.* **205**:1 (1990), 1–47. MR 91f:14035 Zbl 0722.14021

[Rizzo 2003] O. G. Rizzo, "Average root numbers for a nonconstant family of elliptic curves", *Compositio Math.* **136**:1 (2003), 1–23. MR 2005b:11074 Zbl 1021.11020

[Rohrlich 1993] D. E. Rohrlich, "Variation of the root number in families of elliptic curves", *Compositio Math.* **87**:2 (1993), 119–151. MR 94d:11045 Zbl 0791.11026

[Rohrlich 1996] D. E. Rohrlich, "Galois theory, elliptic curves, and root numbers", *Compositio Math.* **100**:3 (1996), 311–349. MR 97m:11075 Zbl 0860.11033

[Serre 2008] J.-P. Serre, *Topics in Galois theory*, 2nd ed., Research Notes in Mathematics **1**, A K Peters, Wellesley, MA, 2008. MR 2008i:12010 Zbl 1128.12001

[Shioda 1990] T. Shioda, "On the Mordell–Weil lattices", *Comment. Math. Univ. St. Paul.* **39**:2 (1990), 211–240. MR 91m:14056 Zbl 0725.14017

[Silverman 1992] J. H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics **106**, Springer, New York, 1992. MR 95m:11054 Zbl 0585.14026

[Silverman 1994] J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics **151**, Springer, New York, 1994. MR 96b:11074 Zbl 0911.14015

[Stopple 2003] J. Stopple, *A primer of analytic number theory: from Pythagoras to Riemann*, Cambridge University Press, Cambridge, 2003. MR 2004m:11143 Zbl 1029.11001

[Swinnerton-Dyer 1962] H. P. F. Swinnerton-Dyer, "Two special cubic surfaces", *Mathematika* **9** (1962), 54–56. MR 25 #3413 Zbl 0103.38302

[Tate 1979] J. Tate, "Number theoretic background", pp. 3–26 in *Automorphic forms, representations and L-functions* (Corvallis, OR., 1977), vol. 2, edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, Amer. Math. Soc., Providence, 1979. MR 80m:12009 Zbl 0422.12007

[Ulas 2007] M. Ulas, "Rational points on certain elliptic surfaces", *Acta Arith.* **129**:2 (2007), 167–185. MR 2008h:11063 Zbl 1142.11017

[Ulas 2008] M. Ulas, "Rational points on certain del Pezzo surfaces of degree one", *Glasg. Math. J.* **50**:3 (2008), 557–564. MR 2009g:11034 Zbl 1223.14041

[Várilly-Alvarado 2008] A. Várilly-Alvarado, "Weak approximation on del Pezzo surfaces of degree 1", *Adv. Math.* **219**:6 (2008), 2123–2145. MR 2009j:14045 Zbl 1156.14017

Communicated by Hendrik W. Lenstra
Received 2010-09-22 Revised 2010-09-26 Accepted 2010-10-24

varilly@rice.edu Department of Mathematics, Rice University, MS 136, Houston, TX 77005, United States
http://math.rice.edu/~av15/

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in ANT are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@mathscipub.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 5 No. 5 2011