ON IMAGINARY QUADRATIC NUMBER FIELDS WITH 2-CLASS GROUP OF RANK 4 AND INFINITE 2-CLASS FIELD TOWER

Elliot Benjamin

Let $k$ be an imaginary quadratic number field with $C_{k,2}$, the 2-Sylow subgroup of its ideal class group $C_k$, of rank 4. We show that $k$ has infinite 2-class field tower for particular families of fields $k$, according to the 4-rank of $C_k$, the Kronecker symbols of the primes dividing the discriminant $\Delta_k$ of $k$, and the number of negative prime discriminants dividing $\Delta_k$. In particular we show that if the 4-rank of $C_k$ is greater than or equal to 2 and exactly one negative prime discriminant divides $\Delta_k$, then $k$ has infinite 2-class field tower.

Introduction.

Let $k$ denote an algebraic number field and $C_{k,2}$ denote its 2-class group, i.e., the 2-Sylow subgroup of the ideal class group $C_k$ (in the wide sense) of $k$; denote by $k_1$ the Hilbert 2-class field of $k$. Let $k_n$ (for $n$ a nonnegative integer) be defined inductively as $k_0 = k$ and $k_{n+1} = (k_n)_1$. Then $k_0 \subseteq k_1 \subseteq k_2 \subseteq \ldots \subseteq k_n \subseteq \ldots$ is called the 2-class field tower of $k$. If $n$ is the minimal integer such that $k_n = k_{n+1}$, then $n$ is called the length of the tower. If no such $n$ exists, then the tower is said to be of infinite length.

In 1964, Golod and Shafarevich (cf. [4]) established for the first time the existence of infinite $p$-class field towers, for $p$ prime. In the case $p = 2$, their criterion (as refined by Gaschütz and Vinberg [10]) can be stated in the following way, where $E_{k,2}$ denotes the unit group of $k$ mod its squares, $E_k/E_{k,2}^2$: If rank $C_{k,2} \geq 2 + 2(\text{rank } E_{k,2} + 1)^{1/2}$ then $k$ has infinite 2-class field tower. We shall refer to the above inequality as the Golod-Shafarevich inequality. We immediately see that for $k$ imaginary with rank $C_{k,2} \geq 5$, or $k$ real with rank $C_{k,2} \geq 6$, the Golod-Shafarevich inequality is satisfied and $k$ thereby has infinite 2-class field tower. It is well-known that for $k$ imaginary with rank $C_{k,2} = 2$ or 3, the 2-class field tower of $k$ may be finite or infinite, and that if rank $C_{k,2} = 1$ then the 2-class field tower of $k$ is finite and has length 1 (cf. [3], [6], [14], [17], [20]). It has been conjectured that for $k$ imaginary with rank $C_{k,2} = 4$, $k$ has infinite 2-class field tower (cf. [17], [18]).
A partial result in this direction, as proved by Hajir, is that if \( k \) is an imaginary quadratic number field such that \( C_{k,2} \) contains a subgroup isomorphic to \( \mathbb{Z}_4 \times \mathbb{Z}_4 \times \mathbb{Z}_4 \), then \( k \) has infinite 2-class field tower (cf. [6], [7]). We extend this result to particular fields \( k \) with rank \( C_{k,2} = 4 \) and 4-rank of \( C_k \) less than 3. Denoting the discriminant of \( k \) by \( \Delta_k \), our fields \( k \) are classified according to the 4-rank of \( C_k \), the Kronecker symbols \((p/q)\) of the primes dividing \( \Delta_k \), and the number of negative prime discriminants dividing \( \Delta_k \). We demonstrate that if the 4-rank of \( C_k \) is greater than or equal to 2 and exactly one negative prime discriminant divides \( \Delta_k \), then \( k \) has infinite 2-class field tower.

Preliminaries.

Our initial results are directly related to the following inequality (cf. [6], [7]):

**Proposition 1.** Let \( F \) be a totally real field of degree \( n \), \( E \) a totally complex quadratic extension of \( F \), and \( t \) the number of prime ideals of \( F \) which ramify in \( E \). If \( t \geq 3 + 2\sqrt{n} + 1 \) then the 2-class field tower of \( E \) is infinite.

We will also need to utilize the well-known ambiguous class number formula, where for a cyclic extension \( K/F \) an ambiguous ideal class is an ideal class of \( K \) that remains invariant under the action of \( \text{Gal}(K/F) \). We denote the subgroup of ambiguous ideal classes by \( \text{Am}(K/F) \) and its Sylow 2-subgroup by \( \text{Am}_2(K/F) \). We state the following two propositions: (cf. [12], [15]).

**Proposition 2.** Let \( K/F \) be a cyclic extension of prime degree \( p \). Then \( \left| \text{Am}(K/F) \right| = h(F) \cdot p^{t-1}/(E : H) \) where \( t \) is the number of (finite or infinite) primes of \( F \) which are ramified in \( K/F \), \( E = E_F \) is the unit group of \( F \), \( H = E \cap N_{K/F}K^x \) is the subgroup of units which are norms of elements of \( K^x \), and \( K^x \) is the multiplicative group of \( K \).

**Proposition 3.** Let \( K/F \) be a quadratic extension of an algebraic number field where \( h(F) \) is odd. Then \( \left| \text{Am}_2(K/F) \right| = 2^e \) where \( e \) is the 2-rank of \( C_K \).

Results.

We begin by obtaining some conditions on the Kronecker symbols of the primes dividing \( \Delta_k \), directly related to Proposition 1, to insure that an imaginary quadratic number field \( k \) with rank \( C_{k,2} = 4 \) has infinite 2-class field tower.

**Lemma 1.** Let \( k \) be an imaginary quadratic number field such that rank \( C_{k,2} = 4 \). If for some prime \( p_j \equiv 1 \mod 4 \), or \( p_j = 2 \) in which case we further assume that 8 is a fundamental discriminant dividing \( \Delta_k \), we have
\( \left( \frac{p_j}{p_k} \right) = \left( \frac{p_j}{p_l} \right) = \left( \frac{p_k}{p_m} \right) = 1, \) \( p_j, p_k, p_l, p_m \) distinct primes, \( p_j p_k p_l p_m | \Delta_k, \) then \( k \) has infinite 2-class field tower.

**Proof.** We proceed in a similar way to Hajir in [6]. Let \( F = Q(\sqrt{p_j}), \) \( E = k(\sqrt{p_l}) \). We see that \( E \) is a CM field with maximal field subfield \( F \) such that either 7 or 8 primes ramify from \( F \) to \( E \). Since \( 7 \geq 3 + 2\sqrt{2} + 1 \), it follows from Proposition 1 that \( E \) has infinite 2-class field tower. Since \( E \) is an unramified quadratic extension of \( k, k \) has infinite 2-class field tower as well. \( \square \)

We utilize the following notational convenience: If \( d_j \) is a negative prime discriminant we let \( p_j \) denote the prime dividing \( d_j \) if \( d_j \neq -4 \), and \( p_j = 1 \) if \( d_j = -4 \).

**Lemma 2.** Let \( k \) be an imaginary quadratic number field such that rank \( C_k, 2 = n, n \geq 1 \). Let \( L = Q(\sqrt{d_j}) \) and \( F = k(\sqrt{d_j}) \), where \( d_j \) is a negative prime discriminant, \( p_j | \Delta_k \). Then exactly 2n prime ideals in \( L \) are ramified in \( F \) if and only if \( \left( \frac{-p_j}{p_i} \right) = 1, i \neq j, \) for all primes \( p_i | \Delta_k, 1 \leq i \leq n + 1 \).

**Proof.** Assume \( \left( \frac{-p_j}{p_i} \right) = 1, i \neq j, \) for all primes \( p_i | \Delta_k, 1 \leq i \leq n + 1 \). It follows that there are exactly \( n \) primes \( p_i \) dividing \( \Delta_k \) that split in \( L = Q(\sqrt{d_j}) \). Since \( F = k(\sqrt{d_j}) \) is an unramified quadratic extension of \( k \), these \( n \) primes \( p_i \) each have ramification index 2 in \( F \). We therefore can conclude that each of these primes \( p_i \) must ramify from \( L \) to \( F \). There are no other prime ideals that ramify from \( L \) to \( F \), since if there were a prime ideal \( P_m \) in \( L \) that ramifies in \( F \) such that \( P_m \cap Q = p_m \neq p_i, 1 \leq i \leq n + 1 \), it would imply that \( p_m \) ramifies in \( F \). But \( p_m \) does not divide \( \Delta_k \) unless \( p_m = p_j \), and \( F \) is an unramified quadratic extension of \( k \). Since \( p_j \) has ramification index 2 in \( F \), we therefore conclude that exactly 2n prime ideals in \( L \) are ramified in \( F \). The converse is proved in a similar way and is left to the reader. \( \square \)

We note that in our proof of Lemma 1 we were able to utilize the full strength of Proposition 1 by requiring only 7 primes to ramify from \( F \) to \( E \), whereas Hajir, in his original proof that if the 4-rank of \( C_k \) is greater than or equal to 3 then \( k \) has infinite 2-class field tower (cf. [6]), assumed that \( p_i \equiv 1 \mod 4, \left( \frac{p_i}{p_k} \right) = 1, j \neq i, \) and therefore 8 primes ramified from \( F \) to \( E \).

We illustrate Hajir’s method of proof of the above result in the case where a negative prime discriminant \( d_j \) divides \( \Delta_k, \left( \frac{-p_j}{p_i} \right) = 1, j \neq i, \) as follows (cf. [7]):

**Lemma 3.** Let \( k \) be an imaginary quadratic number field such that rank \( C_k, 2 = 4 \). Assume there exists a negative prime discriminant \( d_j, d_j | \Delta_k, \)
such that \( \left( -\frac{p_j}{p_i} \right) = 1, \) \( j \neq i, \) for all primes \( p_i | \Delta_k, \) \( 1 \leq i \leq 5. \) Then \( k \) has infinite 2-class field tower.

**Proof.** Let \( L = \mathbb{Q}(\sqrt{-p_j}) \) and \( F = k(\sqrt{-p_j}). \) By Lemma 2 we see that exactly 8 prime ideals in \( L \) are ramified in \( F. \) By Proposition 2 and Proposition 3, it follows that rank \( C_{F,2} \geq 6. \) Since rank \( E_{F,2} = 2, \) we obtain the Golod-Shafarevich inequality: rank \( C_{F,2} \geq 6 \geq 2 + 2\sqrt{2} + 1 \) and therefore \( F \) has infinite 2-class field tower. Since \( F \) is an unramified quadratic extension of \( k, \) \( k \) has infinite 2-class field tower as well. \( \square \)

We state the following corollaries of Lemma 3:

**Corollary 1.** Let \( k \) be an imaginary quadratic number field such that rank \( C_{k,2} = 4, \) exactly one negative prime discriminant divides \( \Delta_k, \) and \( \Delta_k \equiv 4 \mod 8. \) Then \( k \) has infinite 2-class field tower.

**Proof.** Since \( k \) has exactly one negative prime discriminant and \( \Delta_k \equiv 4 \mod 8, \) all the odd primes dividing \( \Delta_k \) are congruent to 1 mod 4. We therefore have \( \left( -\frac{1}{p_i} \right) = 1 \) for all odd primes \( p_i | \Delta_k \) and our result follows immediately from Lemma 3. \( \square \)

**Corollary 2.** Let \( k \) be an imaginary quadratic number field, rank \( C_{k,2} = 4, \) such that five negative prime discriminants divide \( \Delta_k. \) Then the following fields \( k \) have infinite 2-class field tower, where \( q_i, \) \( 1 \leq i \leq 5, \) is a prime congruent to 3 mod 4:

\[
Q(\sqrt{-q_1q_2q_3q_4q_5}), \quad \left( -\frac{q_i}{q_j} \right) = \left( -\frac{q_i}{q_k} \right) = \left( -\frac{q_i}{q_l} \right) = \left( -\frac{q_i}{q_m} \right) = 1, \\
\{i, j, k, l, m\} = \{1, 2, 3, 4, 5\}
\]

\[
Q(\sqrt{-q_1q_2q_3q_4}), \quad \left( -\frac{q_i}{q_j} \right) = \left( -\frac{q_i}{q_k} \right) = \left( -\frac{q_i}{q_l} \right) = \left( -\frac{q_i}{2} \right) = 1, \\
\{i, j, k, l\} = \{1, 2, 3, 4\}.
\]

**Proof.** It is immediate by applying Lemma 3 to each field \( k \) that we have infinite 2-class field tower. We note that these are all the possible fields satisfying the assumptions of our corollary for which we are able to apply Lemma 3. \( \square \)

For the cases when exactly one negative prime discriminant divides \( \Delta_k \) where \( \Delta_k \not\equiv 4 \mod 8, \) and exactly three negative prime discriminants divide \( \Delta_k, \) we utilize the following lemma:

**Lemma 4.** Let \( k \) be an imaginary quadratic number field such that rank \( C_{k,2} = 4, \) at least two of the prime discriminants dividing \( \Delta_k \) are positive, and \( \left( \frac{p_1}{p_3} \right) = \left( \frac{p_2}{p_3} \right) = 1 \) where \( p_1 \) and \( p_2 \) are distinct primes dividing positive
prime discriminants dividing $\Delta_k$, and $p_3$ is a prime dividing a positive or negative prime discriminant dividing $\Delta_k$, $p_1 \neq p_3 \neq p_2$. Then $k$ has infinite 2-class field tower.

Proof. By the assumptions of our lemma, we can write $k = \mathbb{Q}(\sqrt{-p_1p_2p_3p_4p_5})$ or $k = \mathbb{Q}(\sqrt{-p_1p_2p_3p_4})$ where $p_1 \equiv p_2 \equiv 1 \mod{4}$, or $p_1 = 2$ and $p_2 \equiv 1 \mod{4}$. By Martinet (cf. [17], Proposition 5) we see immediately that $k$ has infinite 2-class field tower. □

We now let $k = \mathbb{Q}(\sqrt{-p_1p_2p_3p_4p_5})$ where $\Delta_k \not\equiv 4 \mod{8}$ and exactly one negative prime discriminant divides $\Delta_k$. We define a Kronecker symbol configuration of $k$ to be a complete list of Kronecker symbols $(\frac{p_i}{p_j})$, $i \leq j$, $l \leq i \leq 5$, $l \leq j \leq 5$. We denote a Kronecker symbol configuration by listing all the Kronecker symbols $(\frac{p_i}{p_j})$ as above with $(\frac{p_i}{p_j}) = 1$ (respectively $-1$), where the remaining Kronecker symbols $(\frac{p_i}{p_j})$, $i < j$, are assumed to be $-1$ (respectively 1).

In Table 1 we utilize the Rédei & Reichardt conditions [19] to list all possible Kronecker symbol configurations (without loss of generality) according to the 4-rank of $C_k$.

<table>
<thead>
<tr>
<th>4-rank of $C_k$</th>
<th>possible Kronecker symbol configurations</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>all Kronecker symbols equal 1</td>
</tr>
<tr>
<td>3</td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_4}\right) = \left(\frac{p_1}{p_5}\right) = 1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = -1$</td>
</tr>
<tr>
<td>2</td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_4}\right) = \left(\frac{p_2}{p_3}\right) = 1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_4}\right) = \left(\frac{p_1}{p_5}\right) = \left(\frac{p_2}{p_5}\right) = 1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_3}\right) = -1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_3}\right) = -1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_4}\right) = \left(\frac{p_2}{p_3}\right) = -1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_4}\right) = \left(\frac{p_2}{p_3}\right) = \left(\frac{p_1}{p_5}\right) = \left(\frac{p_2}{p_5}\right) = 1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_4}\right) = \left(\frac{p_1}{p_3}\right) = \left(\frac{p_1}{p_5}\right) = \left(\frac{p_2}{p_3}\right) = 1$</td>
</tr>
</tbody>
</table>
We are now able to state the following theorem:

**Theorem 1.** Let $k$ be an imaginary quadratic number field with rank $C_{k,2} = 4$, $\Delta_k \not\equiv 4 \mod 8$, and exactly one negative prime discriminant dividing $\Delta_k$. Then the following fields $k$ have infinite 2-class field tower, where $\{i, j, k, l, m\} = \{1, 2, 3, 4, 5\}$:

(A) 4-rank of $C_k$ equal to 2, 3 or 4

<table>
<thead>
<tr>
<th>4-rank of $C_k$</th>
<th>possible Kronecker symbol configurations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_4}\right) = \left(\frac{p_1}{p_4}\right) = \left(\frac{p_2}{p_3}\right) = 1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_4}\right) = \left(\frac{p_2}{p_3}\right) = -1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_4}\right) = \left(\frac{p_2}{p_3}\right) = -1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_4}\right) = \left(\frac{p_2}{p_3}\right) = -1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_4}\right) = 1$</td>
</tr>
<tr>
<td>0</td>
<td>all Kronecker symbols equal $-1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = 1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_3}\right) = 1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_3}\right) = \left(\frac{p_2}{p_4}\right) = 1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_3}\right) = \left(\frac{p_2}{p_4}\right) = \left(\frac{p_3}{p_5}\right) = 1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_3}\right) = \left(\frac{p_2}{p_4}\right) = \left(\frac{p_3}{p_5}\right) = 1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_3}\right) = \left(\frac{p_2}{p_4}\right) = \left(\frac{p_3}{p_5}\right) = 1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_3}\right) = \left(\frac{p_2}{p_4}\right) = \left(\frac{p_3}{p_5}\right) = 1$</td>
</tr>
<tr>
<td></td>
<td>$\left(\frac{p_1}{p_2}\right) = \left(\frac{p_1}{p_3}\right) = \left(\frac{p_2}{p_4}\right) = \left(\frac{p_3}{p_5}\right) = 1$</td>
</tr>
</tbody>
</table>
(B) 4-rank of $C_k$ equal to 1 and Kronecker symbol configuration of $k$ not $(\frac{p_i}{p_j}) = (\frac{p_j}{p_k}) = 1$ where either $p_j$ or $p_k$ is the prime dividing the negative prime discriminant dividing $d_k$.

(C) 4-rank of $C_k$ equal to 0 (i.e., $C_k,2$ elementary) and Kronecker symbol configuration of $k$ not one of the following types:

- all Kronecker symbols equal −1
- $(\frac{p_j}{p_i}) = 1$
- $(\frac{p_i}{p_j}) = (\frac{p_k}{p_l}) = 1$
- $(\frac{p_j}{p_i}) = (\frac{p_k}{p_l}) = (\frac{p_i}{p_m}) = 1$ where either $p_j$ or $p_m$ is the prime dividing the negative prime discriminant dividing $d_k$.

Proof. For case (A) with 4-rank of $C_k$ equal to 3 or 4 the result has been established by Hajir (cf. [6], [7]). For case (A) with 4-rank of $C_k$ equal to 2, and cases (B) and (C), we apply Lemma 4 to our fields listed in Table 1 to establish our result.

From Table 1 we see that there are 32 possible Kronecker symbol configurations when exactly one negative prime discriminant divides $\Delta_k$, $\Delta_k \not\equiv 4 \mod 8$. From Theorem 1 we find that for 27 of these Kronecker symbol configurations, $k$ has infinite 2-class field tower. The unknown cases can be summarized by means of the 4-rank of $C_k$ as follows:

<table>
<thead>
<tr>
<th>4-rank of $C_k$</th>
<th>number of Kronecker symbol configurations where 2-class field tower of $k$ may be finite</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Remark 1. For the case when exactly three negative prime discriminants divide $\Delta_k$, one can again utilize the Rédei & Reichardt conditions and Lemma 4 to obtain fields with infinite 2-class field tower. We note that in this case the Kronecker symbol configuration $(\frac{p_j}{p_i}) = (\frac{p_k}{p_l}) = (\frac{p_i}{p_m}) = 1$ may not satisfy the requirements of Lemma 4; however, Lemma 1 may be used when $p_1 \not\equiv 1 \mod 4$, or when $p_1 = 2$ if 8 is a fundamental discriminant dividing $\Delta_k$. For the case when five negative prime discriminants divide $\Delta_k$, one can utilize the Rédei & Reichardt conditions and Corollary 2 to obtain
fields with infinite 2-class field tower. In a follow-up paper we will demonstrate that our techniques allow us to conclude that $k$ has infinite 2-class field tower for imaginary quadratic number fields $k$ when rank $C_{k,2} = 4$, $C_k$ has 4-rank equal to 2, and either five negative prime discriminants divide $\Delta_k$ or $\Delta_k \not\equiv 4 \mod 8$ (cf. [1]).

Examples.

From Lemma 1, Corollary 2, and Theorem 1 we immediately obtain that the following fields $k$ have infinite 2-class field tower. We list our fields according to the 4-rank of $C_k$.

$C_{k,2}$ elementary: $k = \mathbb{Q}(\sqrt{-61}, 620) = \mathbb{Q}(\sqrt{-3 \cdot 5 \cdot 13 \cdot 79})$

$k = \mathbb{Q}(\sqrt{-120}, 180) = \mathbb{Q}(\sqrt{-5 \cdot 13 \cdot 17 \cdot 29})$

$k = \mathbb{Q}(\sqrt{-122}, 655) = \mathbb{Q}(\sqrt{-3 \cdot 5 \cdot 13 \cdot 17 \cdot 37})$

$k = \mathbb{Q}(\sqrt{-212}, 135) = \mathbb{Q}(\sqrt{-5 \cdot 7 \cdot 11 \cdot 19 \cdot 29})$

$k = \mathbb{Q}(\sqrt{-256}, 360) = \mathbb{Q}(\sqrt{-2 \cdot 5 \cdot 13 \cdot 17 \cdot 29})$

$k = \mathbb{Q}(\sqrt{-430}, 360) = \mathbb{Q}(\sqrt{-2 \cdot 5 \cdot 7 \cdot 29 \cdot 53})$

$k = \mathbb{Q}(\sqrt{-440}, 115) = \mathbb{Q}(\sqrt{-3 \cdot 5 \cdot 13 \cdot 37 \cdot 61})$

$k = \mathbb{Q}(\sqrt{-850}, 135) = \mathbb{Q}(\sqrt{-5 \cdot 11 \cdot 13 \cdot 29 \cdot 41})$

$k = \mathbb{Q}(\sqrt{-2,035}, 240) = \mathbb{Q}(\sqrt{-2 \cdot 5 \cdot 17 \cdot 41 \cdot 73})$

$k = \mathbb{Q}(\sqrt{-5,863}, 655) = \mathbb{Q}(\sqrt{-5 \cdot 7 \cdot 29 \cdot 53 \cdot 109})$

$4 - \text{rank of } C_k = 1$: $k = \mathbb{Q}(\sqrt{-184}, 008) = \mathbb{Q}(\sqrt{-2 \cdot 3 \cdot 11 \cdot 17 \cdot 41})$

$k = \mathbb{Q}(\sqrt{-531}, 867) = \mathbb{Q}(\sqrt{-3 \cdot 7 \cdot 19 \cdot 31 \cdot 43})$

$k = \mathbb{Q}(\sqrt{-2,657}, 415) = \mathbb{Q}(\sqrt{-3 \cdot 5 \cdot 29 \cdot 41 \cdot 149})$

$k = \mathbb{Q}(\sqrt{-6,425}, 679) = \mathbb{Q}(\sqrt{-3 \cdot 13 \cdot 37 \cdot 61 \cdot 73})$

$4 - \text{rank of } C_k = 2$: $k = \mathbb{Q}(\sqrt{-3,989}, 095) = \mathbb{Q}(\sqrt{-5 \cdot 11 \cdot 29 \cdot 41 \cdot 61})$
Remark 2. Since the fields satisfying the conditions of Lemmas 1, 3, and 4 possess an unramified quadratic extension which satisfy the Golod-Shafarevich inequality, (cf. [19] in regard to Lemmas 1 and 4) it follows from Theorem 6 of Hajir (cf. [5]) that the rank of the 2-class groups of these fields tend to infinity.

In conclusion, we see that the conjecture concerning the 2-class field tower of \( k \) being infinite holds in a number of particular fields \( k \) when the 4-rank of \( C_k \) is equal to 0, 1, or 2, and always holds when the 4-rank of \( C_k \) is greater than or equal to 3. Our techniques allow us to obtain families of fields \( k \) with 4-rank of \( C_k \) equal to 0, 1, or 2 and \( k \) having infinite 2-class field tower, as well as the rank of the 2-class groups of the fields in the tower of \( k \) tending to infinity. However, the complete resolution of the conjecture concerning all imaginary quadratic number fields \( k \) with rank \( C_{k,2} = 4 \) is still a very open question.

Acknowledgement. I would like to express my gratitude to Chip Snyder for a number of helpful suggestions throughout this paper.

References


Received December 9, 1999 and revised April 19, 2000.

Department of Mathematics
Unity College
Unity, Maine 04988-9502
E-mail address: benjamin@mint.net
ON NON-ORIENTABLE RIEMANN SURFACES WITH $2p$ OR $2p + 2$ AUTOMORPHISMS

Emilio Bujalance, Grzegorz Gromadski, and Peter Turbek

It is known that the maximal order of a cyclic group of automorphisms admitted by a Klein surface or real algebraic curve of algebraic genus $p$ is $2p$ or $2(p + 1)$, depending on whether $p$ is odd or even. In this paper, we classify the automorphism groups of all non-orientable Klein surfaces, without boundary, which admit an automorphism group of order $2p$, or $2(p + 1)$. We determine that the automorphism groups are cyclic precisely when the surfaces are hyperelliptic. Defining equations for all but one family of these Klein surfaces are given.

There are certain properties that can be shown to exist for at least one Riemann surface of each genus $g$ or Klein surface of each algebraic genus $p$. For example, for each $g \geq 2$, there exists a Riemann surface of genus $g$ which possesses $8g + 8$ automorphisms [1], [7]. Similarly, for each $p \geq 2$, there exist orientable and non-orientable Klein surfaces of algebraic genus $p$ which possess $4(p + 1)$ and $4p$ automorphisms respectively [10]. In each case these bounds are sharp, since there are infinitely many $g$ and $p$ for which there are no Riemann or Klein surfaces which possess more automorphisms. Similarly, for each $p \geq 2$, there exists a Klein surface which admits a cyclic group of automorphisms of order $2(p + 1)$ if $p$ is even, or $2p$, if $p$ is odd [3], [9]. Although Klein surfaces which possess such large cyclic automorphism groups are so numerous, in this paper we prove an interesting converse. Recall that a non-orientable Klein surface without boundary is called a non-orientable Riemann surface. We show that any non-orientable Riemann surface which admits a group $G$ of automorphisms of order $2(p + 1)$ or $2p$ has the property that $G$ is either cyclic or an extension of a cyclic group by $\mathbb{Z}_2$. We further determine that $G$ is cyclic if and only if the non-orientable Riemann surface is hyperelliptic.

Recall that the category of Klein surfaces is equivalent to the category of real algebraic curves. Each Klein surface can be realized as an algebraic curve, defined by real equations, upon which complex conjugation acts. Non-orientable Riemann surfaces correspond to algebraic curves whose real locus is empty. We determine the defining equations for the Riemann double
covers for all but one family of surfaces found in this paper. We explicitly determine the symmetry which yields each Klein surface and give explicit equations for the automorphism group of order $2p$ or $2(p + 1)$. This work, in conjunction with the examination of bordered Klein surfaces in [6], provides an extensive analysis of all Klein surfaces which admit automorphism groups of order $2(p + 1)$ or $2p$.

1. Preliminaries.

Let $U$ denote the upper half plane and let $W$ be a compact Klein surface of algebraic genus $p \geq 2$. Then $W$ can be realized as $U/\Gamma$ for some non-euclidean crystallographic (NEC) group $\Gamma$. In addition, $\Gamma$ can be chosen to be a surface group, meaning that it has no nonidentity orientation preserving elements of finite order. If $W$ admits a group of automorphisms $G$, then there exists an NEC group $\Lambda$ such that $G \cong \Lambda/\Gamma$, and $W/G$ and $U/\Lambda$ are equivalent Klein surfaces. Important properties of $\Lambda$ are contained in its signature

$$ (g; \pm; [m_1, m_2, \ldots, m_r]; \{(n_{11}, \ldots, n_{1s_1}), \ldots, (n_{k1}, \ldots, n_{ks_k})\}). $$

The above signature indicates that $U/\Lambda$ has topological genus $g$ and $k$ boundary components. Each $m_i$ is called a proper period, each $n_{ij}$ is called a link period, and each term $(n_{11}, \ldots, n_{is_i})$ is called a period cycle of $\Lambda$. It is well-known that $\Lambda$, with signature (1), can be generated by one of the following two sets of elements:

$$ \{x_u, c_{ij}, e_v, a_w, b_w \mid 1 \leq u \leq r, 1 \leq i \leq k, 0 \leq j \leq s_i, 1 \leq v \leq k, 1 \leq w \leq g \}, $$

(2) $$ \{x_u, c_{ij}, e_v, d_w \mid 1 \leq u \leq r, 1 \leq i \leq k, 0 \leq j \leq s_i, 1 \leq v \leq k, 1 \leq w \leq g \}. $$

The first set generates $\Lambda$ if the signature of $\Lambda$ has a plus sign, and the second set generates $\Lambda$ if the signature has a minus sign. In either case, the generators satisfy the relations

$$ \{x_u^{m_u} = 1, c_{i,s_i} = e_i^{-1}c_{i,0}e_i, c_{i,j}^2 = (c_{i,j-1}c_{ij})^{n_{ij}} = 1 \mid 1 \leq u \leq r, 1 \leq i \leq k, 1 \leq j \leq s_i, 1 \leq w \leq g \}, $$

and the additional relation $x_1 \ldots x_re_1 \ldots e_k[a_1, b_1] \ldots [a_g, b_g] = 1$ if there is a plus sign in the signature, and $x_1 \ldots x_re_1 \ldots e_kd_1^2 \ldots d_g^2 = 1$ if there is a minus sign in the signature.

Since $\Gamma$ is a surface group, it has a signature of the form

$$ (g'; \pm; [-]; \{(-), \ldots, (-)\}). $$

Let $k'$ denote the number of empty period cycles in (4).
The Riemann-Hurwitz formula yields that

\[ \frac{\alpha g' + k' - 2}{|G|} = \alpha g + k - 2 + \sum_{j=1}^{r} \left( 1 - 1/m_{ij} \right) + \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{s_i} \left( 1 - 1/n_{ij} \right). \]  

(5)

The algebraic genus \( p \) of a Klein surface is defined to be the genus of its Riemann double cover. If a surface has signature (1), then \( p = \alpha g + k - 1 \), where \( \alpha \) equals 2 or 1, depending on whether there is a + or − in the signature.

A non-orientable Riemann surface is a non-orientable Klein surface without boundary. Note that if the surface group \( \Gamma \) in (4) corresponds to a non-orientable Riemann surface of algebraic genus \( p \), then \( \Gamma \) has the signature

\[ (p + 1; -; [-]; \{-\}). \]  

(6)

It is well-known that an NEC group with signature (1) exists if and only if the right hand side of (5) is positive. Therefore, a way to construct Klein surfaces which admit a given automorphism group \( G \) is to define a group homomorphism \( \theta \) from an NEC group \( \Lambda \), with signature (1), onto a finite group \( G \). If \( \Gamma = \ker(\theta) \), then the Klein surface \( U/\Gamma \) admits \( G \cong \Lambda/\Gamma \) as a group of automorphisms.

Throughout the paper we will use the following presentation for a group \( G \) of order \( 2n \), where \( n \) is odd, which possesses a cyclic group of order \( n \):

\[ \langle A, B \mid A^n = B^2 = 1, BAB = A^\alpha \rangle. \]  

(7)

In (7), \( \alpha \) and \( n \) are relatively prime and \( \alpha^2 \equiv 1 \mod n \). Note that if \( \alpha = -1 \), then \( G \) is the dihedral group \( D_n \), and if \( \alpha = 1 \), then \( G \) is cyclic. However we will use the presentation

\[ \langle a \mid a^{2n} = 1 \rangle \]  

(8)

for the cyclic group of order \( 2n \). In addition, we denote the greatest common divisor of the integers \( a \) and \( b \) by \( \gcd(a, b) \).

We state without proof an elementary result concerning groups of order \( 2n \) where \( n \) is odd.

**Proposition 1.1.** Let \( G \) be a group of order \( 2n \), where \( n \) is odd. Then \( G \) contains a unique normal subgroup \( H \) of order \( n \) which contains all the elements of \( G \) of odd order.

We now define notation to be used for the rest of the paper. Define \( \gamma = p \) if \( p \) is odd or \( p + 1 \) if \( p \) is even. Let \( W \) denote a non-orientable Riemann surface of algebraic genus \( p \) which admits a group \( G \) of automorphisms of order \( 2\gamma \) and let its corresponding NEC group \( \Gamma \) have signature (6). Let \( \Lambda \) denote an NEC group such that \( \Gamma \vartriangleleft \Lambda \), \( G \cong \Lambda/\Gamma \), and \( X/G = U/\Lambda \). We assume \( \Lambda \) has signature (1) and generators and relations (2) and (3). Let
\( \theta : \Lambda \to G \) denote the canonical map, and let \( H \) denote the unique subgroup of \( G \) of order \( \gamma \).

We state two Propositions which allow us to determine the possible signatures for \( \Lambda \).

**Proposition 1.2.** Let \( W, \Lambda, G, \) and \( H \) be defined as above. Let \( \Lambda \) have signature (1). Then no proper period \( m_i \) in the signature for \( \Lambda \) is divisible by 4, and each link period \( n_{ij} \) in the signature of \( \Lambda \) must be odd.

**Proof.** Let \( \theta : \Lambda \to G \) be the canonical epimorphism. Since \( \Gamma = \ker(\theta) \), and \( \Gamma \) is a surface group, it cannot contain elliptic elements. Since 4 doesn’t divide \( |G| \), no proper period is divisible by 4. We now show that no link period can be even. Assume \( n_{ij} \) is an even link period and let \( c_{ij} \) be the elements of order two such that \( c_{ij} \) has even order \( n_{ij} \). Since \( W \) has no boundary, \( c_{ij} \) is not in \( \ker(\theta) \), and \( \theta(c_{ij}) \) has order \( n_{ij} \). Therefore \( G \) contains a dihedral group of order \( 2n_{ij} \), which implies that 4 divides \( |G| \), a contradiction. Thus each \( n_{ij} \) cannot be even.

**Proposition 1.3.** Let \( W, \Lambda, \Gamma, G, \) and \( H \) be defined as above. Assume all of the proper periods, if they exist, in the signature of \( \Lambda \) are odd. Then \( \Lambda \) cannot have one of the following signatures:

i. \( (1; -; m_1, m_2, \ldots, m_r); \{ - \} \),
ii. \( (0; +; m_1, m_2, \ldots, m_r); \{ (-) \} \),
iii. \( (0; +; m_1, \ldots, m_r); \{ (n_1, n_2, \ldots, n_s) \} \),
iv. \( (0; +; [-]; \{ (n_1, \ldots, n_s) \} \).

**Proof.** In each case, let \( \theta : \Lambda \to G \) be the canonical map. We say a proper generator of \( \Lambda \) is a generator in (2) which is not in \( \Gamma \). A proper word of \( \Lambda \) is the product of proper generators of \( \Lambda \). Recall that \( \ker(\theta) \) is non-orientable if and only if a glide reflection (one of the \( d_{ij} \) in (2)) or a non-orientable proper word belongs to \( \ker(\theta) \). We will show that the above signatures imply that \( \ker(\theta) \) is orientable. Considering (2), in Cases (i)-(iv) generators for \( \Lambda \) are:

i. \( \{ x_1, \ldots, x_r, d_1 \} \),
ii. \( \{ x_1, \ldots, x_r, e_1, c_0 \} \),
iii. \( \{ x_1, \ldots, x_r, e_1, c_0, \ldots, c_s \} \),
iv. \( \{ c_0, \ldots, c_s \} \).

Since each \( x_i \) has odd order \( \theta(x_i) \in H \). In Case (i), this implies that \( \theta(d_1) \notin H \), otherwise \( \theta \) is not onto. In Cases (ii)-(iii), this implies that \( \theta(e_1) \in H \), since \( x_1 \ldots x_r e_1 = 1 \). Note that for each \( j, c_j \notin \Gamma \), since \( W \) has no boundary, therefore \( \theta(c_j) \notin H \). For each case, let \( w \) be a proper, non-orientable word in \( \Lambda \) and let \( w_1 = \theta(w) \). For each case, we consider \( w_1 \) in \( G/H \). For Case (i) we obtain that \( w_1 H = \theta(d_1) H \), where \( q \) is the number of appearances of \( d_1 \) in \( w \). Since \( w \) is non-orientable, \( q \) is odd, therefore \( w_1 \notin H \), therefore \( w \notin \ker(\theta) \). In Cases (ii)-(iv), \( w_1 H = \theta(c_0) H \),
where \( q \) is the number of appearances of \( c_0, \ldots, c_s \) in \( w \). Again, since \( w \) is non-orientable, \( q \) is odd, and therefore \( w \notin \ker(\theta) \). Therefore, no proper, non-orientable word in \( \Lambda \) is an element of \( \Gamma = \ker(\theta) \). This contradicts that \( W \) is non-orientable.

We now determine possible signatures for \( \Lambda \). Let \( \Lambda \) have the signature (1) and let \( \Gamma \) have the signature (6). From (5), we obtain

\[
\frac{1}{2} > \frac{p-1}{2\gamma} = \alpha g + k - 2 + \sum_{i=1}^{r} (1 - 1/m_i) + \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{s_i} (1 - 1/n_{ij}).
\]

Therefore \( \alpha g + k \leq 2 \) and we have the following cases:

(a) \( g = 1, \alpha = 1, k = 1 \),
(b) \( g = 1, \alpha = 1, k = 0 \),
(c) \( g = 0, k = 2 \),
(d) \( g = 0, k = 1 \),
(e) \( g = 2, \alpha = 1, k = 0 \).

We consider each case in conjunction with (9). In Case a), if \( r > 0 \), or if \( r = 0 \) and \( s \geq 2 \), then the right hand side of (9) is not less than 1/2. Therefore, for case a), \( r = 0 \) and \( s_1 = 1 \). In Case b), \( r \) must be two. In Case c), \( r \) must be 0, however, it is not possible for both \( s_1 \) and \( s_2 \) to be greater than 0. In Case d), \( r \leq 2 \). If \( r = 2 \), then \( s_1 = 0 \) or 1. If \( r = 1 \), then \( s_1 = 1 \) or 2. Proposition 1.3 yields that \( r \neq 0 \). Finally, in Case e), the right hand side of (9) is zero or greater than or equal to 1/2. Therefore we arrive at the list of relevant signatures for which the right hand side of (9) is less than 1/2:

(a) \( (1; -; [-]; \{(n)\}) \),
(b) \( (1; -; [m_1, m_2]; \{-\}) \),
(c) \( (0; +; [-]; \{(-), (n)\}) \),
(d1) \( (0; +; [m_1, m_2]; \{(n)\}) \),
(d2) \( (0; +; [m_1, m_2]; \{(-)\}) \),
(d3) \( (0; +; [m]; \{(n_1, n_2)\}) \),
(d4) \( (0; +; [m]; \{(n)\}) \).

2. Main theorems.

We now determine which signatures for \( \Lambda \) yield non-orientable Riemann surfaces which admit \( 2\gamma \) automorphisms. We also determine all of the possible epimorphisms from \( \Lambda \) onto a group of order \( 2\gamma \).

**Theorem 2.1.** Let \( W \) be a non-orientable Riemann surface of even algebraic genus \( p \geq 2 \) which admits a group of automorphisms \( G \) of order \( 2p + 2 \).
Then $G \cong \mathbb{Z}_{2p+2}$ and $W$ is hyperelliptic, or $G$ is a non-cyclic group with presentation (7).

Proof. We continue to use the notation established above. Note that $\Lambda$ cannot have signatures a) or c) listed above since in these cases (9) implies that $n = (p + 1)/2$, which is impossible since $p$ is even.

We now consider Case b). In this case, $\Lambda$ has signature $(1; +; [m_1, m_2]; \{-\})$ and so (9) yields that $1/2 + 1/(p + 1) = 1/m_1 + 1/m_2$. If $m_1 = 2$, then $m_2 = p + 1$. If $m_1$ and $m_2$ are both greater than two, then (9) yields that $2(m_1 + m_2)/(p + 1) = m_1m_2/(p + 3)$, therefore one of $m_1$ or $m_2$ is even. However, if $m_1$ is even, then, since no proper period is divisible by 4, $m_1 \geq 6$. However, $m_1 \geq 6$ and $m_2 \geq 3$ contradicts (9). Therefore, Case b) yields the signature

$$\text{(10) } (1; +; [2, p + 1]; \{-\}).$$

We now determine the possible signatures for Cases d1)–d4). Case d1) is impossible. To see this, assume $m_1 = m_2 = 2$. Then (9), yields that $n = (p + 1)/2$ which is impossible, since $p$ is even. If $m_1 > 2$, then (9) yields that $1/2 > -1/m_1 - 1/m_2 + 3/2 - 1/(2n) \geq -1/3 - 1/2 + 3/2 - 1/6 = 1/2$, a contradiction.

The argument in Case d2) is analogous to Case b) above, and easily yields the signature

$$\text{(11) } (0; +; [2, p + 1]; \{-\}).$$

In Case d3), (9) yields

$$\text{(12) } \frac{1}{2} > \frac{p - 1}{2(p + 1)} = 1 - \frac{1}{m} - \frac{1}{2n_1} - \frac{1}{2n_2}.$$ 

We will show that $m = 2$. If $m > 2$, Propositions 1.2 and 1.3 yield that $m \geq 6$. However, since $n_1$ and $n_2$ are odd, this contradicts (12). Therefore $m = 2$. From (12) we obtain $2n_1n_2 = (n_1 + n_2)(p + 1)$. Let $n_1 \leq n_2$ and assume that $n_2 < p + 1$. Then $2n_1n_2 > n_2(n_1 + n_2)$, so $n_1 > n_2$ which is a contradiction. Thus, since each link period must be odd, $n_2 = p + 1$. But then $1/(p + 1) = 1/n_1$ which give $n_1 = p + 1$ also. Hence Case d3) only yields the signature

$$\text{(13) } (0; +; [2]; \{(p + 1, p + 1)\}).$$

We now examine Case d4). Equation (9) yields that

$$\text{(14) } 2mn = (p + 1)(2n + m).$$

If $m = 2(p + 1)$ then $n = p + 1$ and so $\Lambda$ has signature $(0; +; [2(p + 1)]; \{(p + 1)\})$. However this signature cannot occur, since it implies that the element $x \in \Lambda$ corresponding to the proper period has order $2(p + 1)$ in $G$, which implies that $G$ is cyclic. On the other hand, this signature also implies that $G$ contains two elements of order two, whose product is $p + 1$, which
is a contradiction. Therefore \( m \leq (p + 1) \), however, this contradicts (14). Therefore \( \Lambda \) cannot have a signature of the form \( d4 \).

Therefore the possible signatures for \( \Lambda \) are

i. \((1; -; [2, p + 1]; \{-\})\),
ii. \((0; +; [2, p + 1]; \{-\})\),
iii. \((0; +; [2]; \{(p + 1, p + 1)\})\).

Associated to the above signatures are the following presentations for \( \Lambda \):

i'. \( \langle d, x_1, x_2 \mid x_1 x_2 d^2 = 1, x_1^2 = x_2^{p+1} = 1 \rangle \),
ii'. \( \langle e, x_1, x_2, c \mid x_1 x_2 e = 1, x_1^2 = x_2^{p+1} = 1, c^2 = 1, ec = ce \rangle \),
iii'. \( \langle e, x, c_0, c_1, c_2 \mid xe = 1, x^2 = 1, c_0^1 = c_1^2 = c_2^2 = 1, ec_2 e^{-1} = c_0, (c_0 c_1)^{p+1} = (c_1 c_2)^{p+1} = 1 \rangle \).

We will show that there is an epimorphism \( \theta : \Lambda \to G \) only if \( \Lambda \) has one of the signatures ii) or iii) above.

Assume that \( \Lambda \) has signature i) with the associated presentation listed above. Let \( H \) be the normal subgroup of \( G \) of order \( p + 1 \). Clearly \( \theta(x_2) \in H \), and \( \theta(d^2) \in H \). Since \( x_1 \) has order 2, \( \theta(x_1) \notin H \). Therefore \( \theta(x_1 x_2 d^2) \notin H \). This contradicts that \( x_1 x_2 d^2 = 1 \).

We now determine the epimorphisms if \( \Lambda \) has signature ii) with its associated presentation above. Since \( \Lambda \) has elements of orders two and \( p + 1 \), and \( G \) has order \( 2(p + 1) \), we see that \( G \) must be a group with presentation (7). Let \( \theta : \Lambda \to G \) be an epimorphism. We may assume that \( \theta(x_1) = B \) and \( \theta(x_2) = A \), therefore \( \theta(e) = (BA)^{-1} \). Then \( \theta(c) = A^k B \), for some integer \( k \), and \( \theta(c^2) = A^k BA^k B = A^k A^k = A^{k+1} \). Therefore

\[
\left(k(\alpha + 1) \equiv 0 \mod p + 1 \right. \tag{15}
\]

To satisfy the defining relations of \( \Lambda \), we must have \( \theta(ec) = \theta(ce) \), or equivalently, \( BA^{k+1} B = A^k BBA \). This yields that

\[
\left(k + 1)(\alpha - 1) \equiv 0 \mod p + 1 \right. \tag{16}
\]

which combined with Equation (15), yields that \( k \equiv 2^{-1}(\alpha - 1) \mod p + 1 \). Since \( \alpha^2 \equiv 1 \mod p + 1 \), this value of \( k \) satisfies both Equations (15) and (16). Therefore \( \theta(c) = A^{(\alpha-1)/2} B \) or \( A^{(\alpha+p)/2} B \) depending on whether \( \alpha \) is odd or even respectively. We have shown a group \( G \) is an epimorphic image of \( \Lambda \) if and only if it has presentation (7). In addition, each epimorphism \( \theta \) must have the above form.

We now consider signature iii) with its associated presentation for \( \Lambda \) above. It is clear that \( G \) must be dihedral since it contains two elements of order two whose product has order \( p + 1 \). Let \( G \) have presentation (7) with \( \alpha = -1 \). We now determine the epimorphisms \( \theta : \Lambda \to G \). Clearly we may assume that \( \theta(c_0) = B \) and \( \theta(c_1) = AB \). Then \( \theta(c_2) \) must be of the form \( A^k B \), where \( k - 1 \) and \( p + 1 \) are relatively prime. Since \( \theta(x) \notin H \), \( \theta(e) \notin H \). Thus \( \theta(e) = A^k B \), and the relation \( ec_2 e^{-1} = c_0 \) implies that \( 2j \equiv k \mod p + 1 \).
Therefore, given \( k \), it follows that \( j = k/2 \) if \( k \) is even, or \( (k + p + 1)/2 \) if \( k \) is odd. We have shown that for each \( k \) with \( \gcd(k - 1, p + 1) = 1 \), there exists an epimorphism onto \( G \) with \( \theta(c_2) = A^kB \) and \( \theta(x) = \theta(e) = A^{k/2}B \) or \( A^{(k+p+1)/2}B \) depending on whether \( k \) is even or odd respectively.

**Theorem 2.2.** Let \( W \) be a non-orientable Riemann surface of odd algebraic genus \( p \geq 2 \) which admits a group of automorphisms \( G \) of order \( 2p \). Then either \( G \cong \mathbb{Z}_{2p} \) and \( W \) is hyperelliptic, or \( G \cong D_p \).

**Proof.** We again employ Equation (9), with \( \gamma = p \), in conjunction with signatures a) through d4).

If \( \Lambda \) has signature a) or c), then (9) easily yields that \( n = p \). Therefore \( \Lambda \) has signature \((1; -; [-]; \{(p)\})\) or \((0; +; [-]; \{(-), (p)\})\). If \( \Lambda \) has signature b), then (9) yields that

\[
1/2 > \frac{p-1}{2p} = 1 - 1/m_1 - 1/m_2.
\]

Assume \( m_1 \leq m_2 \). Then (17) easily yields that \( m_1 \leq 3 \). If \( m_1 = 2 \), then (17) yields that \( m_2 = 2p \), thus the signature of \( \Lambda \) is \((1; -; [2p]; \{-\})\). If \( m_1 = 3 \), then (17) yields that \( m_2 = 6 - 18/(p+3) \), therefore \( m_2 \) is 3, 4, or 5. However \( m_2 = 4 \) contradicts Proposition 1.2, while \( m_2 = 3 \) or 5 contradicts Proposition 1.3.

For Case d1), (9) yields \( 1/2 > (p-1)/(2p) = 3/2 - 1/m_1 - 1/m_2 - 1/(2n) \). This implies that neither \( m_1 \) nor \( m_2 \) can be greater than 2. This yields the signature \((0; +; [2, 2p]; \{(p)\})\) for \( \Lambda \).

The calculations for Case d2) are exactly the same as for Case b) and yield the signatures \((0; +; [2, 2p]; \{(-)\}), (0; +; [3, 3]; \{(-)\}), (0; +; [3, 5]; \{(-)\})\). The last two signatures contradict Proposition 1.3.

For Case d3), (9) yields

\[
1/2 > (p-1)/(2p) = 1 - 1/m - 1/(2n_1) - 1/(2n_2).
\]

Assume \( n_1 \leq n_2 \). Clearly \( m \leq 6 \), since \( n_1 \) and \( n_2 \) are odd. From Proposition 1.2, \( m \neq 4 \), and from Proposition 1.3, \( m \neq 3 \) or 5. However, if \( m = 2 \), then (18) yields that \( p(n_1 + n_2) = n_1n_2 \) which contradicts that \( n_1 \) and \( n_2 \) are both odd. Therefore \( \Lambda \) cannot have a signature of the form d3).

For Case d4), (9) yields that \( mn = p(2n + m) \), which implies that \( m > 2p \), which contradicts that \( m \) divides \( 2p \).

We now summarize the possible signatures for \( \Lambda \) found above. For each signature, we will determine the possible epimorphisms of \( \Lambda \) onto a group of order \( 2p \).

i. \((1; -; [-]; \{(p)\})\),
ii. \((0; +; [-]; \{(p), (-)\})\),
iii. \((1; -; [2, 2p]; \{-\})\),
iv. \((0; +; [2, 2]; \{(p)\})\),
v. \( (0; +; [2, 2p]; \{-\}). \)

Associated to each of the above signatures are the following presentations for \( \Lambda \).

i. \( \langle e, d, c_0, c_1 \mid ed^2 = 1, e_0^2 = c_1^2 = 1, ec_1e^{-1} = c_0, (c_0c_1)^p = 1 \rangle \),

ii. \( \langle e_1, e_2, c_{1,0}, c_{1,1}, c_{2,0} \mid e_1 e_2 = 1, c_{1,0}^2 = c_{1,1}^2 = c_{2,0}^2 = 1, e_1 e_2 c_1 e_1^{-1} = c_{1,0}, e_2 c_2 e_0 = c_{2,0} e_2, (c_1, 0 c_{1,1})^{p+1} = 1 \rangle \),

iii. \( \langle d, x_1, x_2 \mid x_1 x_2 d^2 = 1, x_1^2 = x_2^2p = 1 \rangle \),

iv. \( \langle e, x_1, x_2, c_0, c_1 \mid x_1 x_2 e = 1, x_1^2 = x_2^2 = 1, c_0^2 = c_1^2 = (c_0 c_1)^p = 1, ec = ec \rangle \),

v. \( \langle e, x_1, x_2, c \mid x_1 x_2 e = 1, x_1^2 = x_2^{2p} = 1, c^2 = 1, ec = ec \rangle \).

Assume \( \Lambda \) has signature i) with associated presentation i) above. Then \( G \cong D_p \), since it contains two elements of order two whose product has order \( p \). Let \( G \) have presentation (7) with \( \alpha = -1 \), and let \( H = \langle A \rangle \). We may assume that \( \theta(c_0) = B \), and \( \theta(c_1) = AB \). Since \( \theta(d^2) \) must have odd order, \( \theta(d^2) \in H \), therefore, since \( ed^2 = 1 \), \( \theta(e) \in H \), say \( \theta(e) = A^k \). From \( \theta(ec_1 e^{-1}) = \theta(c_0) \), we deduce that \( A^{2k+1} = 1 \), so \( \theta(e) = A^{(p-1)/2} \). If \( \theta(d) \notin H \), then \( \theta(d^2) = 1 \), contradicting \( ed^2 = 1 \). Therefore \( \theta(d) = A^j \), for some integer \( j \). To satisfy the relation \( ed^2 = 1 \), we must have \( 2j \equiv (p+1)/2 \) mod \( p \) therefore, \( \theta(d) = A^{(p+1)/4} \) or \( A^{(3p+1)/4} \), depending, respectively, on whether \( (p+1)/2 \) is even or odd.

Assume \( \Lambda \) has signature ii) with associated presentation ii) above. Then clearly \( G \) is dihedral. Let \( G \) have presentation (7) with \( \alpha = -1 \) and let \( H = \langle A \rangle \). We may assume that \( \theta(c_{1,0}) = B \), \( \theta(c_{1,1}) = AB \), and \( \theta(c_{2,0}) = A^k B \) for some integer \( k \). Note that \( \theta(e_1) \neq 1 \), therefore \( \theta(e_2) \neq 1 \). The only nonidentity element of \( G \) which commutes with \( A^k B \) is itself, therefore \( \theta(e_2) = A^k B \). This yields that \( \theta(e_1) = A^k B \) also. This, in conjunction with \( \theta(e_1 c_{1,1} e_1^{-1}) = \theta(c_{1,0}) \) yields that \( k \) must be \( (p+1)/2 \). Therefore, \( \theta(c_{2,0}) = \theta(e_1) = \theta(e_2) = A^{(p+1)/2} B \).

Assume \( \Lambda \) has signature iii) with associated presentation iii) above. In this case \( G \cong Z_{2p} \) with presentation (8). We may assume that \( \theta(x_1) = a^p \) and \( \theta(x_2) = a \). But then \( \theta(d) = a^{(p-1)/2} \). Observe that the inverse image of \( \langle a^p \rangle \) has signature \( (1; -; [2, \ldots, 2]; \{-\}) \). So \( W \) is hyperelliptic, by [5].

Assume \( \Lambda \) has signature iv) with associated presentation iv) above. In this case, \( G \) is again dihedral. Assume \( G \) has the presentation (7) with \( \alpha = -1 \) and let \( H = \langle A \rangle \). We may assume that \( \theta(c_0) = B \) and \( \theta(c_1) = AB \). Since \( \theta(x_1) \) and \( \theta(x_2) \) have order 2, they do not lie in \( H \), and since \( x_1 x_2 e = 1 \), \( \theta(e) \in H \). Thus \( \theta(e) = A^k \), where \( k \) is chosen so that \( \theta(ec_1 e^{-1}) = c_0 \). This yields that \( k = (p-1)/2 \). Therefore \( \theta(x_1) = A^s B \) and \( \theta(x_2) = A^t B \) where \( s \) and \( t \) are chosen so that \( A^s B A^t B = A^{(p+1)/2} \). Therefore \( s \) and \( t \) can be any integers such that \( 2(s-t) = p + 1 \).
Assume $\Lambda$ has signature $v$ with associated presentation $v$ above. Then $G$ is cyclic with presentation (8). Therefore $\theta(x_1) = a^p, \theta(x_2) = a, \theta(e) = a^{p-1}$ and $\theta(e) = a^p$. In this case, the inverse image of $\langle a^p \rangle$ has signature $(0; +; [2, \ldots, 2]; \{(\cdot)\})$, therefore $W$ is hyperelliptic.

If $m$ is odd, a group with presentation (7) has a nontrivial center if and only if it is cyclic. Therefore we have the following corollary.

**Corollary 2.3.** Let $W$ be a non-orientable Riemann surface of algebraic genus $p \geq 2$ which admits a group of automorphisms $G$ of order $2p + 2$, if $p$ is even, or $2p$, if $p$ is odd. Then $W$ is hyperelliptic if and only if $G$ is cyclic.

### 3. Defining equations.

We now determine the defining equations for the Klein surfaces whose signatures were determined in the previous section. We do this by determining the Riemann double cover of each Klein surface and the conformal automorphism group of order $2\gamma$. We then determine the symmetry of the Riemann double cover which yields the Klein surface as its quotient.

We continue to use the notation established earlier. Let $W$ denote the Klein surface with a group $G$ of $2\gamma$ automorphisms whose defining equations we seek. Let $X = W/G$. Since $G$ contains a normal subgroup $H$ of index two, we let $Z$ denote the Klein surface $W/H$. The double covers of $W, Z, \text{ and } X$, will be denoted by $\hat{W}, \hat{Z}, \text{ and } \hat{X}$ respectively. Let $\sigma$ denote the symmetry acting on $\hat{W}$, such that $\hat{W}/\langle \sigma \rangle = W$. The groups $G$ and $G/H$ induce automorphism groups of $\hat{W}$ and $\hat{Z}$ respectively. We shall use the same notation for the elements of $G$ and their induced actions on the Riemann double covers. Note that $Z$ must be non-orientable, since $W$ is non-orientable and the order of $H$ is odd. In addition, the symmetry $\sigma$, acting on $\hat{Z}$ must be fixed point free. To see this, assume it possesses a fixed point $z_0$. Above $z_0$, there are an odd number of points of $\hat{W}$, say $w_1, \ldots, w_k$. On $\hat{W}$, $\sigma$ is fixed point free, therefore $\sigma$ permutes $w_1, \ldots, w_k$. This contradicts that a permutation of order two cannot act without fixed points on an odd number of objects.

In the following sections we compute defining equations for $\hat{W}$, explicitly determine the symmetry $\sigma$ such that $W = \hat{W}/\langle \sigma \rangle$, and explicitly determine its automorphism group of order $2\gamma$. The only family of surfaces whose equations we do not compute is the one corresponding to the signature $(0; +; [2, 2]; \{(p)\})$ found in the proof of Theorem 2.2. The computations for this family are quite difficult, due to the fact that this family has real dimension two in Teichmüller space, while the other families listed in the proofs of Theorems 2.1 and 2.2 have real dimension one.

**Proposition 3.1.** The following basic results concerning defining equations will be freely used.
i. Assume \( w^n - r(x)/s(x) \) is a defining equation of \( \tilde{W} \), where \( r(x) \) and \( s(x) \) are relatively prime polynomials. Then there is a defining equation for \( \tilde{W} \) of the form \( w_1^n - f(x) \), where \( f(x) \) is a polynomial. One such equation is obtained by defining \( w_1 := s(x)w \) and \( f(x) = r(x)s^{n-1}(x) \).

ii. Assume \( \tilde{W} \) has a defining equation of the form \( w^{dn} - (x-a)^{dn}f(x) \), where \((m,n) = 1\), \((n,d) = 1\) and \( f(x) \) is a polynomial. Then \( x \) is a defining equation for \( \tilde{W} \) of the form \( w_1^n - (x-a)^{d}f_1(x) \). One such equation can be obtained in the following way. Since \((m,n) = 1\), there exist integers \( u \) and \( v \) such that \( mu = 1 - nv \). We can choose \( u \) so that \((d,u) = 1\). Thus \( w^{dn} = (x-a)^{du}f^u(x) = (x-a)^{d-v}f^u(x) \).

Thus define \( w_1 = (x-a)^v w^n \) and \( f_1(x) = f^u(x) \). Since \((u,dn) = 1\), note that \( w \) can be expressed in terms of \( w_1 \) and \( x \), thus \( C(w_1,x) = C(w,x) = C(\tilde{W}) \).

iii. Let \( w^n - (x-a)^m f(x) \) be a defining equation for \( \tilde{W} \), where \( (x-a) \) and \( f(x) \) are relatively prime polynomials. Assume that \( m = nu + v \), where \( u \geq 1 \). Then \( \tilde{W} \) has a defining equation of the form \( w_1^n - (x-a)^nu f(z) \), by defining \( w_1 := w/(x-a)^{u} \).

iv. Let \( w^n - r(x)/s(x) = 0 \) be a defining equation for \( \tilde{W} \), and assume \( r(x) \) and \( s(x) \) are polynomials with no factors in common. Assume \( (x-a)^n \) divides either \( r \) or \( s \) and \( (x-a)^{n+1} \) does not. Let \( d \) denote the greatest common factor of \( m \) and \( n \). Then there exists \( d \) points of \( \tilde{W} \) which lie over the point \( a \in \tilde{X} \) and the ramification index there is \( n/d \).

v. Assume \( C(\tilde{W}) \) is a cyclic Galois extension of degree \( n \) of \( C(X) \). Assume that \( C(\tilde{W}) = C(X)[w] \), and that the automorphism \( A \) is the identity on \( C(X) \) but \( A(w) = \epsilon w \), where \( \epsilon \) is a primitive \( n \)th unit of unity. Assume \( \lambda \) is a symmetry or an automorphism of \( C(\tilde{W}) \) of order 2, and assume \( \lambda A = A^\alpha \). Let \( \lambda(w) = a_0 + a_1 w + \cdots + a_{n-1} w^{n-1} \), where each \( a_i \in C(X) \). Then

\[
A \circ \lambda(w) = a_0 + a_1(\epsilon w) + \cdots + a_{n-1}(\epsilon w)^{n-1},
\]

\[
\lambda \circ A^\alpha = \epsilon^i (a_0 + a_1 w + \cdots + a_{n-1} w^{n-1}),
\]

where \( i = -1 \) if \( \lambda \) is a symmetry, but \( i = 1 \) otherwise. We obtain that each \( a_j = 0 \) except for \( a_0 \), if \( i = 1 \), and \( a_{n-i} \), if \( i = -1 \). In particular, we have the following cases. If the automorphism \( B \) has order two and \( BAB = A^{-1} \), then \( B(w) = h_1/w \), for some \( h_1 \in C(X) \). If \( \lambda \) is a symmetry which commutes with \( A \), then \( \lambda(w) = h_2/w \) for some \( h_2 \in C(X) \).

vi. Assume that the polynomial \( F(z,w) \) is a defining equation for a Riemann surface \( \tilde{W} \) and \((r,s)\) is a nonsingular solution of \( F \). If \( \tilde{W} \) admits a symmetry \( \sigma \), then there exists an induced symmetry \( \sigma \) on \( C(\tilde{W}) \) such that \( \sigma(i) = -i \). Note that \((r,s) \) is the unique point on \( F \) which satisfies
Then \((r, s)\) is a fixed point of \(\sigma\) if and only if \((r, s)\) is also a solution of \(\sigma(z) - \tau \) and \(\sigma(w) - \tau\). In particular, we have the following cases. Assume \((r, s)\) is a fixed point of \(\sigma\). If \(\sigma(z) = z\), then \(r\) is real. If \(\sigma(z) = -z\), then \(r\) is pure imaginary. If \(\sigma(z) = 1/z\), then \(r\) is a complex number with \(|r| = 1\). It is not possible for \((r, s)\) to be a fixed point of \(\sigma\) and \(\sigma(z) = -1/z\).

3.1. Defining equations for even \(p\).

3.1.1. The signature \((0; +; [2, p + 1]; \{-\})\). Assume that \(\Lambda\) from the proof of Theorem 2.1 has the above signature. Then its associated Fuchsian group \(\Lambda^+\) has signature \((0; +; [2, 2, p + 1, p + 1]; \{-\})\) and the four points of \(\hat{X} := U/\Lambda^+\) fixed by the elliptic elements of \(\Lambda^+\) lie above interior points of \(X := U/\Lambda\). Assume that \(G\) satisfies presentation (7). Note that \(\hat{Z} := \hat{W}/\langle A \rangle\) has genus zero, since two points of \(\hat{X}\) are ramified in it. From the remark at the beginning of this section, the induced action of \(\sigma\) on \(\hat{Z}\) is fixed point free.

Let us choose coordinates for \(C(x) = C(\hat{X})\) so that the induced action of \(\sigma\) on \(C(x)\) is conjugation. Assume that \(x = a \pm bi\) are the points of \(\hat{X}\) with ramification index two in \(\hat{Z}\), where \(a\) and \(b\) are real and \(b \neq 0\). By the real change of coordinates \(x \mapsto (2/b)(x - a)\), we may assume that \(x = \pm 2i\) are the points ramified in the covering of \(\hat{X}\) by \(\hat{Z}\). Let us choose coordinates for \(\hat{Z}\) so that \(z = i\) and \(z = -i\) lie over \(x = 2i\) and \(x = -2i\) respectively. In addition, let us choose coordinates so that \(z = \infty\) is one of the points lying over \(x = \infty\). Note that \(B\) is an automorphism of order two such that \(\hat{Z}/\langle B \rangle = \hat{X}\). Since \(z = i\) are the ramified points, they are fixed by \(B\). This uniquely identifies \(B\) as the map \(B(z) = -1/z\). Note that \(x_0 := z - 1/z\) is fixed by \(B\), therefore \(x_0 \in C(x)\). However, \(x_0 = x\), since each function agrees at the points \(z = i, z = -i,\) and \(z = \infty\). Note that \(z\) satisfies the minimal polynomial \(z^2 - zx - 1 = 0\). The roots of this are \(z\) and \(-1/z\). Since \(\sigma(x) = x\), we have that \(\sigma(z) = z\) or \(\sigma(z) = -1/z\). The first of these yields that \(\hat{Z}\) has fixed points under \(\sigma\), therefore \(\sigma(z) = -1/z\).

Let \(c + di\), with \(c\) and \(d\) real, be one of the points of \(\hat{Z}\) with ramification index \(p + 1\). Note that \(d \neq 0\), otherwise \(c + di\) lies over a point on the boundary of \(X\). Note that a real transformation of the form

\[
\begin{align*}
z \mapsto \beta z - 1 & \quad \frac{z}{z + \beta}
\end{align*}
\]

fixes both \(i\) and \(-i\). If \(c \neq 0\), define

\[
\beta = -\frac{\sqrt{c^4 + 2c^2(d^2 + 1) + (d^2 - 1)^2 + c^2 + d^2 - 1}}{2c}.
\]
Then $\beta$ is a real number, and one can check that the map (19) maps $c + di$ to a complex number with real part equal to 0. Thus we may assume that the ramified points are $z = ki, -ki, 1/(ki)$ and $-1/(ki)$. The above change of coordinates still yields that $B(z) = -1/z$ and $\sigma(z) = -1/z$.

Since $C(W)$ is a cyclic extension of $C(z)$, Proposition 3.1 yields that there is a defining equation of $\hat{W}$ of the form

$$w^{p+1} - (z - ki)(z + ki)v(z - 1/ki)^r(z + 1/ki)^s = 0$$

where each of $v, r, s$ are between 0 and $p + 1$ and relatively prime to $p + 1$. In addition, since $\infty$ is not ramified, $p + 1$ divides $1 + v + r + s$. Let $A$ be the map which is the identity on $C(\hat{Z})$ and which maps $w$ to $\epsilon w$. Since $\sigma$ and $A$ commute, from Proposition 3.1 we have that $\sigma(w) = h_1(z)/w$ for some $h_1(z) \in C(z)$. Applying this to (21) yields that

$$h_1(z)^{p+1} = w^{p+1}(-1/2 + ki)(-1/2 - ki)v(-1/2 + 1/2)^r(-1/2 - 1/2)^s.$$

Thus $h_1^{p+1} z^{1 + v + r + s} = (22)$

$$(z - ki)^{1 + v - r - s}(z + ki)^{1 + r}(z - 1/ki)1 + r(z + 1/ki)^{s + v},$$

therefore $p + 1$ divides both $1 + r$ and $v + s$, so $r = p$ and $s = p + 1 - v$. This implies that $(-1)^{v+s}(ki)^{1+v-r-s} = (-1)^{p+1}z^{2p-2p}$ and $h_1 = \eta_1(z - ki)(z + ki)(z - 1/ki)(z + 1/ki)/z^2$, where $\eta_1$ is a $p + 1$st root of $k^{2v-2p}(-1)^{v+1}$. Since $\sigma$ has order two and fixes $(z - ki)(z + ki)(z - 1/ki)(z + 1/ki)/z^2$, we deduce that $\eta_1$ must be real. A defining equation for $\hat{W}$ is

$$w^{p+1} = f(z) := (z - ki)(z + ki)v(z - 1/ki)^p(z + 1/ki)^{p+1-v} = 0.$$

Recall that $AB = BA^\alpha$ and that $B(z) = -1/z$. We may assume that $0 \leq \alpha < p + 1$. From (v) of Proposition 3.1 we deduce that $B(w) = h_2(z)^{w^\alpha}$ for some $h_2(z) \in C(z)$. Applying this to (23) above yields

$$(h_2(z))^{p+1} = f(z)^{-\alpha}(-1/2 - ki)(-1/2 + ki)v(-1/2 - 1/2)^p(-1/2 + 1/2)^{p+1-v}.$$

Thus $h_2(z)^{p+1}$ equals

$$h_2(z) = \frac{(z - ik)^{p+1-v-\alpha}(z + ik)^{p-v\alpha}(z - 1/ik)^{v-p\alpha}(z + 1/ik)^{1-(p+1-v)\alpha}}{(-1)^{v+1}k^{2p-2v}z^{2p+2}}.$$

Considering the power of $z - ik$, this implies that $v = p + 1 - \alpha$. Since $\alpha^2 \equiv 1 \mod p + 1$, (24) is a $(p + 1)$st power. If $k_2$ is chosen so that $(p + 1)k_2 = 1 - \alpha^2$, then

$$h_2(z) = \frac{e^j\eta_1}{z^2}(z + ik)^{1-\alpha-k^2}(z - 1/ik)^{1-\alpha}(z + 1/ik)^{k_2}$$

where $j \in \mathbb{Z}$ and $\epsilon$ is a primitive $(p + 1)$st root of unity. A tedious calculation shows that $B^2(w) = \epsilon^{j(\alpha+1)}w$, so $j(\alpha+1) \equiv 0 \mod p + 1$. By redefining $B$ as
$A^{-j}B$, we may assume $j = 0$. With this definition of $h_2(z)$, $B(w) = h_2(z)w^\alpha$ and $BAB = A^\alpha$.

We now check that $B$ and $\sigma$ commute. Recall that $\sigma(w) = h_1/w$, where $h_1 = \eta_1(z - ik)(z + ik)(z - 1/ik)(z + 1/ik)$. However, since $v = p + 1 - \alpha$, we see that $\eta_1 = \eta_2$. Since $B\sigma(z) = \sigma B(z)$, it is sufficient to show that $B\sigma(w) = \sigma B(w)$. However $B\sigma(w) = B(h_1(z)/w) = B(h_1(z))/(h_2(z)w^\alpha)$. On the other hand, $\sigma B(w) = \sigma(h_2(z)w^\alpha) = (h_2)h_1^{-\alpha}/w^\alpha$. A tedious calculation shows that $B(h_1) = h_2\sigma(h_2)h_1^\alpha$. Therefore, $B\sigma(w) = \sigma B(w)$, thus $B$ and $\sigma$ commute.

We have shown that a defining equation for $\hat{W}$ is

$$w^{p+1} - (z - ik)(z + ik)^{p+1-\alpha}(z - 1/ik)^{p}(z + 1/ik)^\alpha = 0.$$  

In addition to $A$, it possesses the following symmetries and automorphisms:

$$\sigma(z) = z, \quad \sigma(w) = \eta_1(z^2 + k^2)(z^2 + 1/k^2)/(z^2 w),$$

$$B(z) = -1/z, \quad B(w) = h_2(z)w^\alpha,$$

where $h_2(z)$ is defined in (25) with $j = 0$, and $\eta_1$ is the real $p + 1$st root of $(-1)^{\alpha}k^{2-2\alpha}$. Since $\sigma$ is fixed point free on $C(z)$, $\sigma$ is fixed point free on $\hat{W}$, therefore the Klein surface $W = \hat{W}/\langle \sigma \rangle$ has no boundary.

### 3.1.2. The signature $(0; +; [2]; \{(p + 1, p + 1)\})$.

Assume $\Lambda$ from the proof of Theorem 2.1 has the above signature. In this case $G \cong D_{p+1}$ and two boundary points of $X$ and one interior point of $X$ are ramified in $W$. Note that $X$ has genus 0 and assume coordinates are chosen so that the action of $\sigma$ on $X$ is conjugation. By a real change of coordinates, we may assume $\infty$ is one of the fixed points of $X$ which has ramification index $p + 1$ in $\hat{W}$. Note that two points of the form $a \pm bi$ of $X$ are ramified with index 2 in $\hat{W}$, where $a$ and $b$ are real and $b \neq 0$. By making the change of coordinates $x \mapsto 2(x-a)/b$, we may assume that the points $\pm 2i$ are ramified with ramification index 2 in $\hat{W}$. Let $\tilde{Z}$ be the orbit space of $\hat{W}$ under the action of $\langle A \rangle$. Exactly as in Section 3.1.1 we may assume that $C(\tilde{Z}) = C(z)$, where $x = z - 1/z, B(z) = -1/z$ and $\sigma(z) = -1/z$. Note that $z = 0$ and $z = \infty$ are the two points lying over $x = \infty$. There is another point of $\bar{X}$ which has ramification index $p + 1$ in $\hat{W}$, and let $z = k$ be one of the points of $\bar{Z}$ which lies over it. Then the point must be $x = k - 1/k$ and the two points lying over it are $z = k$ and $z = -1/k$. It is easy to see, since $k - 1/k$ is real, that $k$ is real.

To obtain the equation of $W$, we note that the $C(\hat{W})$ is a field extension of $C(z)$ of index $p + 1$ in which $z = 0, k, -1/k$, and $\infty$ are all ramified. Therefore a defining equation of $W$ is of the form

$$w^{p+1} - z(z - k)^u(z + 1/k)^v.$$
where \( u \) and \( v \) are between 0 and \( p+1 \), relatively prime to \( p+1 \), and \( 1+u+v \) is also relatively prime to \( p+1 \). Let \( A(w) = \epsilon w \), where \( \epsilon \) is a primitive \( p+1 \)st root of unity. We must show that the maps \( B(z) = -1/z \) and \( \sigma(z) = -1/z \) lift to \( \hat{W} \) with the properties that \( BA = A^{-1}B \) and \( \sigma \) commutes with both \( A \) and \( B \).

From Proposition 3.1, we have that \( \sigma(w) = f/w \) and \( B(w) = g/w \) for some \( f \) and \( g \in C(z) \). Note that both \( B \) and \( \sigma \) map \( z-k \mapsto -1/z - k = (-k/z)(z + 1/k) \) and \( z + 1/k \mapsto -1/z + 1/k = (1/zk)(z - k) \). Therefore letting \( h = f \) or \( g \) depending on whether we are considering \( \sigma \) or \( B \), we obtain

\[
0 = h^{p+1} - w^{p+1}(-1/z)(-k/z)^u(1/zk)^v(z + 1/k)^u(z - k)^v
= h^{p+1} - (-1)^{1+u}k^{u+v}(z - k)^{u+v}(z + 1/k)^{u+v}/z^{u+v}.
\]

Therefore \( v = p+1 - u \). Let \( \lambda \) be the real \( p+1 \)st root of \( (-1)^{1+u}k^{2u-p-1} \), then \( f = \epsilon_1 \lambda(z - k)(z + 1/k)/z \) and \( g = \epsilon_2 \lambda(z - k)(z + 1/k)/z \), where \( \epsilon_1 \) and \( \epsilon_2 \) are \( p+1 \)st roots of unity. Since \( \sigma \) is a symmetry, \( w = \sigma^2(w) \) yields that \( \epsilon_1 = 1 \).

By redefining \( B \) as \( A^j B \) for an appropriate integer \( j \), we may assume that \( \epsilon_2 = 1 \). Thus a defining equation for \( \hat{W} \) is \( w^{p+1} - z(z - k)^u(z + 1/k)^{p+1-u} \), where \( u \) is relatively prime to \( p+1 \). In addition, \( \hat{W} \) possesses the following automorphisms and symmetries:

\[
A(w) = \epsilon w, A(z) = z, B(w) = \lambda(z - k)(z + 1/k)/(zw), B(z) = -1/z,
\]

\[
\sigma(w) = \lambda(z - k)(z + 1/k)/(zw), \sigma(z) = -1/z, \sigma(i) = -i,
\]

where \( \lambda \) is a real number such that \( \lambda^{p+1} = (-1)^{1+u}k^{2u-p-1} \).

3.2. Defining equations for odd \( p \).

3.2.1. The signatures \((1; -; \{\cdot\}; \{(p)\})\) and \((0; +; \{\cdot\}; \{(-), (p)\})\). Let \( \Lambda \) have one of the above signatures. Note that \( \hat{X} \) has genus one and recall that \( G \cong D_p \). The two signatures are distinguished by the number of fixed ovals of \( \sigma \) acting on \( \hat{X} \); the first yields one fixed oval and the latter yields two. If we define \( \hat{Z} = W/\langle A \rangle \), we see that \( \hat{Z} \) is an unramified cover of \( \hat{X} \), therefore \( \hat{Z} \) has genus one also. There is a distinguished boundary point of \( X \) which has ramification index \( p \) in \( W \). Lying above this point there is a unique point of \( \hat{X} \) and there are two points of \( \hat{Z} \). Let us choose coordinates for \( \hat{Z} \) so that one of these two points is the point at infinity of a defining equation of the form \( y^2 - f(z) = 0 \), where \( f(z) \) has three distinct zeros. Recall that \( B \) is an automorphism of order 2 acting on \( \hat{Z} \), such that \( \hat{X} = \hat{Z}/\langle B \rangle \). Since \( B \) is fixed point free, \( B \) is a translation of order two in the group structure of \( \hat{Z} \), therefore \( \infty \) is mapped to one of the roots of \( f(z) \).
We may make a change of coordinates to assume that $z = 0$ is the root of $f(z)$ to which $\infty$ is mapped under $B$. Thus $B(z) = c^2/z$ for some complex number $c^2$. By making the change of coordinates $z := z/c$, we may assume that $c = 1$.

Therefore $B(z) = 1/z$ is an automorphism of the defining equation
\begin{equation}
y^2 - z(z - a)(z - b),
\end{equation}
of $\hat{Z}$. Note that $B(y)$ must be of the form $r(z)y/s(z)$, where $r(z)$ and $s(z)$ are relatively prime polynomials. Therefore $z^3r^2y^2 - abs^2(z - 1/a)(z - 1/b) = 0$ and so $z^4r^2(z - a)(z - b) - abs^2(z - 1/a)(z - 1/b) = 0$. If $z - a$ does not divide $(z - 1/a)(z - 1/b)$, it must divide $s(z)$. But then $(z - a)$ divides $r(z)$, a contradiction. Therefore we may assume that $z - a$ divides $(z - 1/a)(z - 1/b)$. If $a = 1/a$, then $a = \pm 1$, and similarly, $b = 1/b$, so $b = \pm 1$. Therefore $y^2 - z^2$, and $B(y) = \pm iy/z^2$. However, in this case, the point corresponding to $z = 1, y = 0$ is fixed by $B$, thus $B$ is not fixed point free. Therefore, $a = 1/b$, and we may assume a defining equation for $\hat{Z}$ is
\begin{equation}
y^2 - z(z - a)(z - 1/a),
\end{equation}
with $B(y) = -y/z^2$.

Before we compute the symmetry acting on $\hat{Z}$, we first determine the orbit space of $X = \hat{Z}/\langle B \rangle$. This will help us determine the action of $\sigma$ on $\hat{Z}$ that will produce one or two fixed ovals in $\hat{X}$. Note that both $t := z + 1/z$ and $u := y - y/z^2$ are fixed by $B$. It is easy to verify that they satisfy the defining equation
\begin{equation}
u^2 - (t - 2)(t + 2)(t - (a^2 + 1)/a)
\end{equation}
and that $C(X) = C(t, u)$.

Recall that $\sigma$ commutes with $B$ and is fixed point free on $\hat{Z}$. Since $z = \infty$ lies above a boundary point of $X$, we deduce that $\sigma$ interchanges $\infty$ and $B(\infty)$. Therefore $\sigma(z) = c/z$, where, since $\sigma$ is a symmetry, $\overline{c}/c = 1$, therefore $c$ is real. However, $\sigma B(z) = \sigma(1/z) = z/c$, while $B\sigma(z) = cz$. Therefore $c = \pm 1$ and $\sigma(z) = 1/z$ or $-1/z$. We now show that $\sigma(z) = -1/z$ if the signature of $\Lambda$ is $(1; -; [-]; \{(p)\})$ and $\sigma(z) = 1/z$ if the signature of $\Lambda$ is $(0; +; [-]; \{((-), (p))\})$.

Assume $\sigma(z) = 1/z$ and let $\sigma$ act on the defining equation (27). Using that $\sigma(y) = r(z)y/s(z)$ for some relatively prime polynomials $r(z)$ and $s(z)$, we deduce that $z^4r^2y^2 - s^2(z - 1/\overline{a})(z - \overline{a}) = 0$. As before, we deduce that $r/s = \pm 1/z^2$, so $\sigma(y) = \pm y/z^2$, and that $a = 1/\overline{a}$ or $a = \overline{a}$. Since $\sigma(z) = 1/z$, a point $(\beta, \gamma)$ satisfying (27) will have its first coordinate fixed by $\sigma$ if and only if $|\beta| = 1$. If $a = 1/\overline{a}$, the point $(a, 0)$ is fixed by $\sigma$. Therefore, $a \neq 1/\overline{a}$ and thus $a = \overline{a}$. We know that $\sigma(y) = \pm y/z^2$; we now determine which sign should be chosen. Let $r$ be real and let $P = (e^{ir}, \gamma)$ be a point satisfying (27). We must define $\sigma$ so that $P$ is not a fixed point. However
$P$ is fixed if and only if $\pm y/z^2 - \gamma$ has a zero at $P$. This occurs if and only
if $\pm \gamma/e^{2ir} - \bar{\gamma} = 0$, which implies $\pm \gamma^2/e^{2ir} - |\gamma|^2 = 0$, which implies
that $\pm(e^{ir} - a)(e^{ir} - 1/a)/e^{ir} = |(e^{-ir} - a)(e^{-ir} - 1/a)|$. Simplifying the last
equality yields $\pm(2 \cos(r) - (a^2 + 1)/a) = |2 \cos(r) - (a^2 + 1)/a|$. If $a > 0$, the
left hand side is negative, so the plus sign should be chosen. If $a < 0$, the
left hand side is positive, so the minus sign should be chosen. In this way,
$\sigma$ will have no fixed points on $\tilde{Z}$. In summary, we have that if $\sigma(z) = 1/z$,
then $a$ is real and $\sigma(y) = y/z^2$ if $a > 0$ while $\sigma(y) = -y/z^2$ if $a < 0$.

We now show that if $\sigma(z) = 1/z$, then $\tilde{X}$ has two fixed ovals. From the
definitions $t = z + 1/z$ and $u = y - y/z^2$, we see that $t$ is fixed by $\sigma$ and
$\sigma(u) = \pm u$, where the sign depends on whether $a$ is positive or negative.
Therefore fixed points of $\tilde{X}$ are the points $(r, \gamma)$, satisfying (28) such that
$r$ is real and $\gamma$ is real or pure imaginary, depending on whether $\sigma(u) = u,$
or $-u$ respectively. Considering (28), we see that there will always be two
intervals for $t$ for which $(t - 2)(t + 2)(t - (a^2 + 1)/a)$ is positive and two
intervals for which it is negative. Therefore, independent of the definition
of $\sigma(u)$, there are two ovals of $\tilde{X}$ fixed by $\sigma$.

Now assume $\sigma(z) = -1/z$. We will determine what this implies about
$a$, $\sigma(y)$, and the number of fixed ovals of $\tilde{X}$. As before, applying $\sigma$ to (27)
yields that $\sigma(y) = ry/s$ and $r^2y^2 + s^2(1/z)(-1/z - \bar{a})(-1/z - 1/\bar{a}) = 0$. This
yields $z^4r^2(z - a)(z - 1/a) + s^2(z + \bar{a})(z + 1/\bar{a}) = 0$. Since $a = -1/\bar{a}$ implies
$|a| = 1$, we conclude that $a = -\bar{a}$, so $a = ik$ for some real number $k$ and
$\sigma(y) = \pm iy/z^2$. Since $\sigma(z) = -1/z$, we conclude that no point satisfying (27)
will have the same $z$ coordinate under $\sigma$, therefore $\sigma$ is fixed point free on
$\tilde{Z}$. We now examine its action on $\tilde{X}$. Note that $\sigma(t) = -t$ and $\sigma(u) = \pm iu$.
Let $\eta^2 = i$. Then $\sigma$ fixes $t_1 := it$ and $u_1 := \eta u$ or $\eta^3 u$, depending on whether
$\sigma(u)$ equals $iu$ or $-iu$ respectively. When expressed in terms of $t_1$ and $u_1$, (28) becomes

$$u_1^2 \pm (t_1^2 + 4)(t_1 + (k^2 - 1)/k).$$

Since $t_1^2 + 4 > 0$, this yields only one fixed oval, regardless of the plus or
minus sign. In summary, if $\sigma(z) = -1/z$, then $a = ik$, where $k$ is real,
$\sigma(y) = \pm iy/z^2$, and one oval of $\tilde{X}$ is fixed by $\sigma$.

Recall that the point at infinity of (28) is ramified in $\tilde{W}$ with ramification
index $p$. Therefore the points $Q := (0, 0)$ and the point at infinity $P$ of $\tilde{Z}$
are ramified in $\tilde{W}$. Since the covering of $\tilde{Z}$ by $\tilde{W}$ is cyclic, the field extension
$C(\tilde{W})$ of $C(z, y)$ has a defining equation of the form

$$(29) \quad w^p - h(z, y) = 0,$$

where the multiplicity of each pole and zero of $h(z, y)$ is a multiple of $p$ except
at $P$ and $Q$. At these points, the multiplicity must be relatively prime to
$p$. Note that the function $z$ has the divisor $(z) = 2Q - 2P$ on $\tilde{Z}$. We will
show that, up to isomorphism, powers of $z$ are the only functions we need to consider for $h(z, y)$ in (29).

Assume the divisor for $h$ is $(h) = p(k_1 R_1 + k_2 R_2 + \cdots + k_n R_n) + d_1 P + d_2 Q$, where $d_1$ and $d_2$ are relatively prime to $p$ and each $k_i \in \mathbb{Z}$. Assume that a pole occurs at $P$, and that $d_1 = -(rp + s)$ where $r$ and $s$ are positive. Let $g \in C(\hat{Z})$ be a function which has a simple pole at $P$ and which does not have a zero or pole at $Q$. Then

\[(w/g^p) - h/g^p = 0\]

is also a defining equation for $X$. In addition, the pole divisor at $P$ is $s$, where $s < p$. A similar result holds for $Q$, and zeros at $P$ or $Q$ can also be handled in this manner. In this way, we may assume that $h$ has a pole at $P$ and a zero at $Q$ and the pole and zero each have degree less than $p$. Assume $h$ is not a power of $z$. Since the divisor for $h$ has degree zero we may now assume that the defining equation for $C(\hat{W})$ is (29), where the divisor for $h$ is of the form

\[(h) = p(j_1 R_1 + j_2 R_2 + \cdots + j_m R_m) - p(k_1 S_1 + k_2 S_2 + \cdots + k_n S_n) - dP + dQ,\]

and a zero occurs at each of the $R_i$’s and a pole occurs at each of the $S_i$’s.

Consider the divisor $D := -(j_1 - 1) R_1 - j_2 R_2 - j_3 R_3 \cdots - j_m R_m + k_1 S_1 \cdots + k_n S_n$. This divisor has degree 1. By the Riemann Roch theorem, there is a function $g \in C(\hat{Z})$ such that $(g) + D \geq 0$, in particular, $g$ has a zero of order at least $j_1 - 1$ at $R_1$ and $j_i$ at each $R_i$ for $i > 1$. In addition, it’s pole divisor is contained in the pole divisor of $D$. There are two possibilities. Either the multiplicity of the zeros of $g$ are precisely the same as the zeros in $D$, (and the pole divisor has degree $k_1 + k_2 + \cdots + k_n - 1$), or $g$ also has a zero at a point $T$ and the pole divisor of $g$ agrees precisely with that of $D$. Note that $T$ may be one of the points of (30). In either case, redefining $h = h/g^p$ and $w = w/g$ we have that

\[(h) = pR_1 - pT - dP + dQ.\]

Recall, from (v) of Proposition 3.1, that $B(w) = k(z, y)/w$ for some $k(z, y) \in C(z, y)$. Since $B$ as order 2, we obtain

\[w = B^2(w) = \frac{B(k(z, y))}{k(z, y)} w,\]

so $B$ fixes $k(z, y)$, so $k(z, y) \in C(\hat{X})$. Using the defining equation (29) we obtain that $k^p(z, y)/w^p = B(h)$, thus $k^p(z, y) = hB(h)$. Thus (31) yields that the divisor for $k(z, y)$ is $R_1 + B(R_1) - T - B(T)$. However, since $k(z, y) \in C(\hat{X})$ and $\hat{X}$ is elliptic, we know that $[C(\hat{X}) : C(k)] \geq 2$, thus $[C(\hat{Z}) : C(k)] \geq 4$. On the other hand, since the pole divisor of $k(z, y)$ has degree at most 2, we have $[C(\hat{Z}) : C(k)] \leq 2$ unless $k(z, y)$ is a constant. Thus $k(z, y)$ is a constant, so $B$ switches $R_1$ and $T$. If $R_1 \in \{P, Q\}$, then so is $T$. 

and therefore $h$ is a constant multiple of a power of $z$. We now assume that $R \notin \{P, Q\}$.

Recall that $B$ is a fixed point free automorphism of the elliptic curve $\hat{Z}$. In the group structure of $\hat{Z}$, $B$ corresponds to the map $S \mapsto S + Q$. Therefore, in the group structure $T = R_1 + Q$ and (31) becomes $(h) = pR_1 - p(R_1 + Q) - dP + dQ$. However, a divisor is the divisor of a function in $C(\hat{Z})$ if and only if it has degree zero and its sum is zero in the group structure of $\hat{Z}$. Since $p$ is odd and $Q$ has order two, we obtain that $d$ is odd. In addition, there is a function $h_1$ with divisor $(h_1) = R_1 - T + P - Q = R_1 - (R_1 + Q) + P - Q$. But then $h/h_1^p$ has divisor $(p + d)Q - (p + d)P$. Since $d$ is odd, this yields that $h/h_1^p$ is a constant multiple of a power of the function $z$. We may redefine $h_1$ to absorb this constant, so (29) yields that $(w/h_1)^p - z^j = 0$, for some positive integer $j$. From (ii) of Proposition 3.1, we may assume $j = 1$ and $\tilde{W}$ has defining equations of the form

$$w^p - z = 0, \quad y^2 - z(z-a)(z-1/a),$$

where $a$ is real if $\sigma(z) = 1/z$, and $a = ik$, with $k$ real, if $\sigma(z) = -1/z$.

We now determine the automorphisms of $\tilde{W}$. The map $A(w) = \epsilon w$, $A(z) = z$, $A(y) = y$, where $\epsilon$ is a primitive $p$th root of unity is clearly an automorphism of $\tilde{W}$. From (v) of Proposition 3.1, $\sigma(w) = k_1/w$ and $B(w) = k_2/w$ for some $k_1$ and $k_2$ in $C(\hat{Z})$.

If $\sigma(z) = 1/z$, applying $\sigma$ to (32) yields $k_1^p - 1 = 0$, therefore $k_1$ is a $p$th root of unity. We may redefine $w = \epsilon^j w$, where $j$ is an appropriate integer, to obtain $\sigma(w) = 1/w$. This does not change the defining equations (32) or any results concerning $\hat{Z}$. Applying $B$ to (32) yields that $k_2 = \epsilon^j$ for some integer $j$. Given this, it is trivial to check that $\sigma B(w) = B \sigma(w)$. We now redefine $B$ as $A^j B$ to obtain the simplification $B(w) = 1/w$. This change merely concerns the representation of the dihedral group $\langle A, B \rangle$ and does not change the defining equations or the action of $B$ on $\hat{Z}$.

In a similar manner, if $\sigma(z) = -1/z$, then $\tilde{W}$ has defining equations (32), and possesses the automorphism $B(w) = 1/w$ and the symmetry $\sigma(w) = -1/w$.

In summary, if $\tilde{W}$ is defined as in (32), then $\tilde{W}$ has the automorphisms

$$A(w) = \epsilon w \quad A(z) = z \quad A(y) = y$$

$$B(w) = 1/w \quad B(z) = 1/z \quad B(y) = -y/z^2.$$ 

If $\Lambda$ has signature $(0; +; [-]; \{(-), (p)\})$ then $a$ in (32) is real and $\tilde{W}$ has the symmetry

$$\sigma(w) = 1/w, \quad \sigma(z) = 1/z, \quad \sigma(y) = \pm y/z^2$$ 

where the plus sign is chosen if $a > 0$ and the minus sign is chosen if $a < 0$. If $\Lambda$ has signature $(1; -; [-]; \{(p)\})$, then in (32), $a = ik$, where $k$ is real. In
addition, $\hat{W}$ has the symmetry

\begin{equation}
\sigma(w) = -1/w, \quad \sigma(z) = -1/z, \quad \sigma(y) = \pm iy/z^2.
\end{equation}

In (35), either sign can be chosen in the definition of $\sigma(y)$.

### 3.2.2. The signature $(1; -; [2, 2p]; \{-\})$.

Assume $\Lambda$ has the above signature. Then $X$ has genus 0, and $\sigma$ is fixed point free on $\hat{X}$. Thus the points of $X$ of ramification index 2 and $2p$ each have two points of $\hat{X}$ lying over them. Let us choose coordinates so that $x = 0$ and $x = \infty$ are the points of $\hat{X}$ which have ramification index $2p$ and $x = -1$ is one of the points with ramification index 2. Then $\sigma$ interchanges $x = 0$ and $x = \infty$, so $\sigma(x) = c/x$, where $c$ is a real number. It is easy to see that $c$ must be negative, otherwise $\sigma$ has fixed points on $\hat{X}$. Defining the real number $k$ by $-k^2 = c$, we redefine $x$ as $x = x/k$. With this change of coordinates, $\sigma(x) = -1/x$, $x = 0$ and $x = \infty$ each have ramification index $2p$ in $\hat{W}$, and $x = -1/k$ has ramification index 2. In addition, the other point with ramification index 2 is $x = k$.

From Proposition 3.1, we may assume a defining equation for $\hat{W}$ is of the form

\begin{equation}
w^{2p} - x(x - k)^p(x + 1/k)^p.
\end{equation}

We note that $\hat{W}$ possess the automorphism $A$ which maps $w \mapsto \varepsilon w$ and $x \mapsto x$, where $\varepsilon$ is a primitive $2p$th root of unity. From (v) of Proposition 3.1 we observe that $\sigma(w) = h/w$, for some $h \in C(x)$. Note that $\sigma(x - k) = -1/x - k = (-k/x)(x + 1/k)$ and $\sigma(x + 1/k) = -1/x + 1/k = (1/(kx))(x - k)$. This yields that

\begin{align*}
h^{2p} &= (-1/x)(-1/x - k)^p(-1/x + 1/k)^p x(x - k)^p(x + 1/k)^p \\
&= (x - k)^{2p}(x + 1/k)^{2p}/x^{2p}.
\end{align*}

Therefore, $\sigma(w) = \varepsilon^j (x - k)(x + 1/k)/wx$, for some integer $j$. Since $\sigma$ has order two and fixes $(x - k)(x + 1/k)/x$, we deduce that $\varepsilon = \pm 1$. Regardless of the sign, $\sigma$ will not have fixed points. Therefore $\hat{W}$ has (36) as a defining equation, possesses the automorphism $A$ defined above, and possesses the symmetry

\begin{equation}
\sigma(i) = -i, \quad \sigma(x) = -1/x, \quad \sigma(w) = \pm (x - k)(x + 1/k)/wx,
\end{equation}

where either sign can be chosen in (37).

### 3.2.3. The signature $(0; +; [2, 2p]; \{-\})$.

Let $\Lambda$ have the above signature. In this case $G \cong Z_{2p} = \langle A \mid A^{2p} = 1 \rangle$. Note that $\hat{X}$ has genus 0 and let coordinates be chosen so that the action of $\sigma$ on $\hat{X}$ is complex conjugation. Through a real transformation similar to that
in Section 3.1.1, we may assume that \( C(x) \) is ramified at the points \( x = \pm i \) and \( x = \pm ki \), where \( k \neq \pm 1 \) is a nonzero real number, and the ramification indices are 2, 2, 2p and 2p respectively in \( C(\hat{W}) \). A defining equation for \( \hat{W} \) is

\[
(38) \quad w^{2p} - (x - i)^p(x + i)^p(x - ki)(x + ki)^{2p-1} = 0,
\]

and \( \hat{W} \) possesses the automorphism \( A \) which maps \( w \mapsto \epsilon w \) and \( x \mapsto x \), where \( \epsilon \) is a primitive 2pth root of unity. It is easy to deduce that \( \hat{W} = W/\langle \sigma \rangle \) where

\[
\sigma(x) = x, \quad \sigma(w) = \lambda(x^2 + 1)(x^2 + k^2)/w,
\]

for some 2pth root of unity \( \lambda \). Since \( \sigma \) is a symmetry, \( \sigma^2(w) = w \) implies that \( \lambda = \pm 1 \). We will show that \( \lambda = -1 \), since \( \sigma \) is fixed point free on \( \hat{W} \). Note that if a point \((r, s)\) satisfying (38) is fixed by \( \sigma \), then \( r \) must be real. Therefore, \( s^{2p} = (r^2 + 1)^p(r - ki)(r + ki)^{2p-1} = (r^2 + 1)^p(r + ki)^{2p}(r - ki)/(r + ki) = e^{it}(r^2 + 1)^p(r + ki)^{2p} \), where \( t \) is real and \( e^{it} = (r - ki)/(r + ki) \). Therefore, points \((r, s)\) with \( r \) real, which satisfy (38) have \( s = \epsilon \sqrt{r^2 + 1}e^{it/(2p)}(r + ik) \) for some 2pth root of unity \( \epsilon \). Now assume \( \sigma(w) = \lambda(x^2 + 1)(x^2 + k^2)/w \), where \( \lambda = \pm 1 \). If the point \((r, s) = (r, \epsilon \sqrt{r^2 + 1}e^{it/(2p)}(r + ik))\) is fixed under \( \sigma \), then we have that \( \sigma(w - s) = \sigma(w) - \hat{s} \) must have a zero at \((r, s)\). This yields that

\[
0 = \lambda(r^2 + 1)(r^2 + k^2)/\left(\epsilon \sqrt{r^2 + 1}e^{it/(2p)}(r + ik)\right) - \epsilon \sqrt{r^2 + 1}e^{-it/(2p)}(r - ik).
\]

It is easy to see that \( \lambda = 1 \) implies that such a point is always a fixed point, and \( \lambda = -1 \) implies that such a point is never a fixed point. Therefore \( \sigma(w) = -(x^2 + 1)(x^2 + k^2)/w \).

References


Received January 10, 2000 and revised July 7, 2000. The first author was partially supported by DGICYT PB98-0017. The second author was partially supported by DGICYT through a sabbatical grant for 1997-1998.

Facultad de Ciencias
UNED
28040 Madrid
Spain
E-mail address: eb@mat.uned.es

University of Gdansk
Wita Stwosza 57
Gdansk
Poland
E-mail address: gregrom@ksinet.univ.gda.pl

Purdue University Calumet
Hammond, IN 46323
E-mail address: turbek@nwi.calumet.purdue.edu
ON COMPOSITION OPERATORS WHICH PRESERVE BMO

YASUHIRO GOTOH

Dedicated to Professor Közō Yabuta on his sixtieth birthday

We characterize the Lebesgue measurable maps between Euclidean spaces which preserve BMO.

1. Introduction.

For a subdomain $D$ of $\mathbb{R}^n$, $n \geq 1$, let $\text{BMO}(D)$ be the space of all locally integrable functions $f$ on $D$ satisfying

$$\|f\|_* = \|f\|_{\text{BMO}(D)} = \sup_{Q \subset D} |Q|^{-1} \int_Q |f - f_Q| \, dx < \infty,$$

where $|Q|$ is the $n$-dimensional Lebesgue measure of $Q$, $f_Q = |Q|^{-1} \int_Q f \, dx$, and the supremum is taken over all closed cubes $Q \subset D$ with sides parallel to the coordinate axes.

Let $D$ and $D'$ be subdomains of $\mathbb{R}^m$ and $\mathbb{R}^n$, $m, n \geq 1$, respectively. We say that a map $F : D \to D'$ is measurable if $F^{-1}(E)$ is measurable for each measurable subset $E$ of $D'$. We say that a measurable map $F : D \to D'$ is a BMO map if i) for each null set $E \subset D'$, $F^{-1}(E)$ is also a null set, and furthermore, ii) for each BMO($D'$) function $f$, $C_F(f) = f \circ F$ belongs to BMO($D$). The condition i) guarantees the uniqueness of the function $f \circ F$. From the closed graph theorem each BMO map $F$ induces a bounded operator $C_F$ between BMO spaces.

Various partial results are known for the characterization of BMO maps. It seems, however, that we do not know almost anything yet for non-continuous BMO maps. The main purpose of the present paper is to give a characterization of BMO maps $F : \mathbb{R}^m \to \mathbb{R}^n$, $m, n \geq 1$ (Theorem 3.1).

Our argument depends on the following two celebrated results for BMO; a growth estimation for BMO functions due to John-Nirenberg, and the existence of certain extremal BMO functions due to Uchiyama (Propositions 4.1 and 4.2).

The present paper is organized as follows. First, we give various examples of BMO maps in §2. The main results of the present paper are given in §3. The following §4 is devoted to their proofs. Finally, in §5 we give a remark on BMO maps which are homeomorphisms between intervals.
In the following, a cube implies a closed cube with sides parallel to the coordinate axes, \( tQ \) denotes the cube with the same center as \( Q \) and expanded by a constant factor \( t > 0 \), and we use the letter \( C \) to denote a positive constant which may vary from place to place unless stated otherwise, that is, \( f \leq 2C \) implies \( f \leq C \), on the other hand, \( f \leq 2C_2 \) does not necessarily mean \( f \leq C_2 \). Also we sometimes write “\( F : D \to D' \)” even if \( F(D) \not\subset D' \) under the assumption that both \( A = F(D) \setminus D' \) and \( F^{-1}(A) \) are null sets. For instance, we may write \( F : \mathbb{R} \to (0, \infty), \, F(x) = |x| \), instead of \( F : [0, \infty), \, F(x) = |x| \).

The author would like to thank the referee for his helpful comments and suggestions.

2. Examples.

In the present section we give various examples of BMO maps.

**Example 2.1.** a) Let \( F : D_1 \times D_2 \to D_1, \, D_1 \subset \mathbb{R}^m, \, D_2 \subset \mathbb{R}^n \), be the canonical projection. Then \( F \) is a BMO map satisfying \( \|CF\| = 1 \). In particular, if \( D_2 = \mathbb{R}^n \), then \( \|CF\|_* = \|f\|_* \) holds for each \( f \in \text{BMO}(D_1) \).

b) Let \( F : D \to D' \) be the inclusion map. Then \( F \) is a BMO map satisfying \( \|CF\| = 1 \).

**Example 2.2.** Let \( F : D \to D' \) be a homeomorphism between subdomains of \( \mathbb{R}^n, \, n \geq 2 \). If \( F \) is quasiconformal, then \( F \) is a BMO map satisfying \( \|CF\| \leq C(n, K_F) \), where \( K_F \) is the maximal dilatation of \( F \). Conversely, if \( F \) is a BMO map satisfying i) for each null set \( E, \, F^{-1}(E) \) is also a null set, ii) \( F \) is ACL, iii) \( F \) is differentiable a.e., then \( F \) is a quasiconformal map satisfying \( K_F \leq C(n, \|CF\|) \) (Reimann [13]).

**Example 2.3.** Let \( F \) be a homeomorphism of \( \mathbb{R} \). Then \( F \) is a BMO map if and only if we can take constants \( K, \, \alpha > 0 \) so that

\[
\begin{align*}
\frac{|F^{-1}(E \cap I)|}{|F^{-1}(I)|} & \leq K \left( \frac{|E \cap I|}{|I|} \right)^\alpha
\end{align*}
\]  

(2) holds for each pair of a measurable subset \( E \) of \( \mathbb{R} \) and an interval \( I \) (Jones [11]). Note that (2) holds if and only if \( F^{-1} \) is absolutely continuous and its derivative \( (F^{-1})' \) (or \( -(F^{-1})' \)) is an \( A_\infty \) weight (cf. (4)). In this case \( F^{-1} \) also satisfies the same condition, and so \( F \) induces a bijection of BMO(\( \mathbb{R} \)).

Jones gave no explicit relation between the constants \( K, \, \alpha \) above and \( \|CF\| \). In §3 we show, however, that his argument implicitly gives the following estimations: If (2) holds, then \( \|CF\| \leq CK/\alpha \) for some universal constant \( C > 0 \); conversely, if \( F \) is a BMO map, then we can take constants \( K, \, \alpha \) so that \( K = C_1 \) and \( \alpha = C_2/\|CF\| \), where \( C_k > 0, \, k = 1, 2 \), are universal constants. Hence \( \|CF\| \) and \( \inf(K/\alpha) \) are comparable with universal
constant factors, where the infimum is taken over all pairs of $K$, $\alpha$ satisfying (2) (Theorem 5.3) (cf. Mayer-Zinsmeister [12]).

Fominykh [3] gave a sufficient condition for spherically continuous maps between (finite or infinite) open intervals to be BMO maps, which partially extends Jones’ result.

**Example 2.4.** Let $F : D \to D'$ be a nonconstant holomorphic map between plane domains. Then $F$ is a BMO map if and only if we can take an integer $p > 0$ so that for each disk $B$ satisfying $2B \subset D$, $F$ is $p$-valent on $B$. In particular, a holomorphic map $F : C \to C$ is a BMO map if and only if it is a polynomial (Gotôh [8]).

Thus, whether a given nonconstant holomorphic map $F : D \to D'$ between plane domains is a BMO map or not is independent of the choice of its target $D'$. The following example shows that this does not extend to hold for general maps.

**Example 2.5.** a) Let $D = \{x \in \mathbb{R}^2 \mid 1 < |x| < 2\}$, $I = \{(0, x_2) \in \mathbb{R}^2 \mid -2 < x_2 < -1\}$, and $D_0 = D \setminus I$. Let $F$ satisfy $F(x) = x$ on $D_0$ and $F(x) \in D_0$ on $I$. Then $F : D \to D$ is a BMO map, and $F : D \to D_0$ is not a BMO map, because BMO($D$) $\neq$ BMO($D_0$).

b) Let $F : D \to D'$ be a BMO map. Let $D'_0$ be a subdomain of $D'$ satisfying $F(D) \subset D'_0$. Assume that each BMO($D'_0$) function is the restriction of some BMO($D'$) function. (Such domains $D'_0$ are characterized as relative uniform domains with respect to $D'$ (Gotôh [7]).) For instance, uniform domains $D'_0$ satisfy this condition (Proposition 3.4). In this case $F : D \to D'$ is a BMO map if and only $F : D \to D'_0$ is a BMO map.

**Example 2.6.** Let $D = \mathbb{R}^{n-1} \times (0, \infty)$ be the upper half space. Then for each $f \in \text{BMO}(D)$, its symmetric extension $g$, $g(x, y) = f(x, y)$ on $D$ and $g(x, y) = f(x, -y)$ on $D \setminus D$, is a BMO($\mathbb{R}^n$) function satisfying $\|f\|_{\text{BMO}^n} \leq C\|f\|_{\text{BMO}^D}$, where $C > 0$ is a universal constant, which is called a reflection principle for BMO. In other words, the two-sheeted folding map $F : \mathbb{R}^n \to D$, $F(x, y) = (x, |y|)$, is a BMO map satisfying $\|CF\| \leq C$ (cf. Reimann-Rychener [14]).

**Example 2.7.** Let $D$ be a quasidisk, that is, $D$ is the image of the upper half plane under a quasiconformal map of $\mathbb{R}^2 = \mathbb{R}^2 \cup \{\infty\}$. Let $\tau : \mathbb{R}^2 \to \mathbb{R}^2$ be the quasiconformal reflection with respect to $\partial D$. Then from Examples 2.2 and 2.6 the two-sheeted folding map $F : \mathbb{R}^2 \to D$, $F(x) = x$ on $D$ and $F(x) = \tau(x)$ on $\mathbb{R}^n \setminus D$, is a BMO map.

**Example 2.8.** a) Let $D = \{r < |x| < r'\} \subset \mathbb{R}^n$, $n \geq 2$. Let $a = r'/r$ and set $D_k = \{a^k r < |x| < a^{k+1} r\}$, $k \in \mathbb{Z}$. We define an infinite-sheeted folding map $F : \mathbb{R}^n \to D$ as follows: Set $F(x) = x/a^{2k}$ on $D_{2k}$ and $F(x) = F(\tau_k(x))$ on $D_{2k+1}$, where $\tau_k$ is the reflection with respect to the sphere
\(|x| = a^{2k+1}r\). Then \(F\) is a BMO map and \(\|C_F\| \leq \frac{aC}{n+1}\), where \(C = C(n) > 0\). This is a consequence of the reflection principle, the removability of one point for BMO, and Proposition 5.2 below.

b) We define an infinite-sheeted folding map \(F : \mathbb{R} \to (0, 1)\) as follows: Set \(F(x) = x\) on \([0, 1]\), \(F(x) = 2 - x\) on \([1, 2]\), and \(F(x) = F(x - 2k), 2k \leq x \leq 2k + 2, k \in \mathbb{Z}\). Then \(F\) is a BMO map. On the other hand, \(F \times id_{\mathbb{R}} : \mathbb{R} \times \mathbb{R} \to (0, 1) \times \mathbb{R}\) is not a BMO map: Let \(f(x_1, x_2) = x_2\). Then \(C_{F \times id_{\mathbb{R}}}(f)(x_1, x_2) = x_2\). Thus \(f \in \text{BMO}(\{(0, 1) \times \mathbb{R}\})\) and \(C_{F \times id_{\mathbb{R}}}(f) \notin \text{BMO}(\mathbb{R} \times \mathbb{R})\).

There are essentially non-continuous BMO maps.

**Example 2.9.** a) Let \(\tau\) be a Möbius transformation of \(\mathbb{R}^n = \mathbb{R}^n \cup \{\infty\}\), \(n \geq 2\). Let \(D\) be an arbitrary subdomain of \(\mathbb{R}^n\) and \(D' = \tau(D) \setminus \{\infty\}\). Then \(F = \tau|D : D \to D'\) is a BMO map. This is a consequence of Example 2.2 and the removability of one point for BMO. (cf. Remann-Rychener [14]. Also see Lemma 5.1 below.) For instance, \(x \mapsto x/|x|^2\), which is discontinuous at the origin under the Euclidean topology, induces a bijection between \(\text{BMO}(\{|x| < 1\})\) and \(\text{BMO}(\{|x| > 1\})\).

b) Let \(F : \mathbb{R} \to (0, 1)\) be the infinite-sheeted folding map in Example 2.8 b). Then \(G : \mathbb{R} \to (0, 1), G(x) = F(1/x)\), is a BMO map which is discontinuous at the origin even under the spherical topology.

Moreover, there are BMO maps between plane domains with essential singularities.

**Example 2.10.** Let \(F(z) = \mathcal{P}(1/z)\), where \(\mathcal{P}\) is the Weierstrass \(\mathcal{P}\)-function. Then \(F : \mathbb{C} \to \mathbb{C}\) is a BMO map having the origin as an essential singularity. Another example is given by the Blaschke product \(F : \mathbb{C} \to \mathbb{C}\),

\[
F(z) = \prod_{k=0}^{\infty} \frac{z - 2^{-k}i}{z + 2^{-k}i} \prod_{i=1}^{\infty} \frac{2^ki - z}{2^ki + z}.
\]

(See the next example.) Moreover, for an arbitrary plane domain \(D\) and an arbitrary sequence \(\{z_k\} \subset D, z_k \to \partial D\), there exists a BMO meromorphic map \(F : D \to \mathbb{C}\) having \(\{z_k\}\) as simple poles satisfying \(\|C_F\| \leq C\), where \(C > 0\) is a universal constant (Gotoh [6], [8]).

Contrary to the case of holomorphic maps between plane domains (Example 2.4), it seems difficult to estimate the operator norms for rational maps \(F : \mathbb{C} \to \mathbb{C}\). As to this we only know the following.

**Example 2.11.** Let \(F\) be a finite Blaschke product on the unit disk \(\Delta\). Let \(t_\zeta, \zeta \in \Delta\), denote the Carleson constant associated with the zeros of the Blaschke product \((F - \zeta)/(1 - \zeta F)\). Let \(s_F = \sup_{\zeta \in \Delta} t_\zeta\). Then for the operator norm \(\|C_F\|\) of the map \(F : \mathbb{C} \to \mathbb{C}\), we have \(\|C_F\| \leq C_1(s_F)\) and \(s_F \leq C_2(\|C_F\|)\). In particular, we can show that there exists a sequence of
rational maps $F_k : \mathbb{C} \to \mathbb{C}$, deg $F_k = k$, satisfying $\|C_{F_k}\| \leq C$, where $C > 0$ is a universal constant (Gotoh [6]).

For related topics, see Astala [1], Smith [15], and Mayer-Zinsmeister [12].

3. Main theorem.

We say that a domain $D \subset \mathbb{R}^n$ is admissible if $D$ is an increasing limit of some sequence of cubes. For instance, $\mathbb{R}^n$, half spaces with sides parallel to the coordinate axes, and open cubes are admissible.

**Theorem 3.1** (Main Theorem). For a measurable map $F : D \to D', D \subset \mathbb{R}^m$, $D' \subset \mathbb{R}^n$, we consider the following conditions:

(a) We can take constants $K$, $\alpha > 0$ so that for an arbitrary pair of measurable subsets $E_1, E_2$ of $D'$ we have

$$\sup_{Q \subset D} \min_{k=1, 2} \frac{|F^{-1}(E_k) \cap Q|}{|Q|} \leq K \left( \sup_{Q' \subset D'} \min_{k=1, 2} \frac{|E_k \cap Q'|}{|Q'|} \right)^\alpha,$$

where the suprema are taken over all cubes $Q \subset D$ and $Q' \subset D'$ respectively;

(b) We can take constants $\gamma$, $0 < \gamma < 1/4$, and $\lambda > 0$ so that for an arbitrary pair of measurable subsets $E_1, E_2$ of $D'$ satisfying

$$\sup_{Q' \subset D'} \min_{k=1, 2} \frac{|E_k \cap Q'|}{|Q'|} < \lambda,$$

we have

$$\sup_{Q \subset D} \min_{k=1, 2} \frac{|F^{-1}(E_k) \cap Q|}{|Q|} < \gamma,$$

where the suprema are taken over all cubes $Q \subset D$ and $Q' \subset D'$ respectively;

(c) $F$ is a BMO map.

Then we have (a) $\Rightarrow$ (b) $\Rightarrow$ (c). Moreover, if we can take an admissible domain $D'_0$ satisfying $F(D) \subset D'_0 \subset D'$, then all these conditions are equivalent.

In particular, all the conditions above are equivalent if $D'$ is admissible. The implication (a) $\Rightarrow$ (b) is trivial. We show that (a) implies (c) with $\|C_F\| \leq C(m, n)K/\alpha$ (Lemma 4.8). Furthermore, if we can take an admissible domain $D'_0$ satisfying $F(D) \subset D'_0 \subset D'$, then we show that (c) implies (a) with constants $K = K(m, n)$ and $\alpha = C(m, n)/\|C_F\|$ (Lemma 4.10). Thus we have:

**Corollary 3.2.** If we can take an admissible domain $D'_0$ satisfying $F(D) \subset D'_0 \subset D'$, then the operator norm $\|C_F\|$ and inf$(K/\alpha)$ are comparable with
constant factors depending only on \( m \) and \( n \), where the infimum is taken over all pairs of constants \( K, \alpha \) satisfying the estimation (3).

Let \( F \) be a homeomorphism of \( \mathbb{R} \). Then the condition (2) in Example 2.3 implies (3) with the same constants \( K, \alpha \). And so we may regard the Main Theorem as an extension of Jones’ result.

We say that a weight \( w \) is an \( A_\infty \) weight on \( D \) if we can take a constant \( \alpha, K > 0 \) so that

\[
\frac{\int_{E \cap Q} w \, dx}{\int_Q w \, dx} \leq K \left( \frac{|E \cap Q|}{|Q|} \right)^\alpha
\]

holds for each pair of a measurable set \( E \subset D \) and a cube \( Q \subset D \). A weight \( w \) is an \( A_\infty \) weight if and only if we can take constants \( \varepsilon, \delta, 0 < \varepsilon, \delta < 1 \), so that for each pair of \( E, Q \) satisfying \( |E \cap Q|/|Q| < \delta \), we have

\[
\frac{\int_{E \cap Q} w \, dx}{\int_Q w \, dx} < \varepsilon.
\]

The equivalence of the conditions (a) and (b) of the Main Theorem implies the corresponding result holds for BMO maps.

Recall that for a weight \( w \) on \( D \), \( f = \log w \) belongs to \( \text{BMO}(D) \) if and only if \( w^\gamma \) is an \( A_\infty \) weight on \( D \) for some \( \gamma > 0 \). Hence,

**Corollary 3.3.** We can add the condition

(d) For each \( A_\infty \) weight \( w \) on \( D' \), \( w^\gamma \circ F \) is an \( A_\infty \) weight on \( D \) for some \( \gamma > 0 \).

to the list of the Main Theorem in the sense that (c) \( \Leftrightarrow \) (d) holds.

We say that a domain \( D \subset \mathbb{R}^n \) is uniform if

\[
k_D(x, y) \leq C \log \left( \frac{d(x, \partial D) + d(y, \partial D) + |x - y|}{\min\{d(x, \partial D), d(y, \partial D)\}} \right), \quad x, y \in D,
\]

holds for some \( C > 0 \), where \( k_D \) is the quasihyperbolic metric on \( D \). Uniform domains are invariant under quasiconformal maps on \( \overline{\mathbb{R}}^n \). Half spaces are uniform domains. In the case of a simply connected plane domain, \( D \) is uniform if and only if \( D \) is a quasidisk. The uniformness can be characterized by the BMO extension property.

**Proposition 3.4** (Jones [10]). A domain \( D \subset \mathbb{R}^n \) is uniform if and only if each \( \text{BMO}(D) \) function is the restriction of some \( \text{BMO}(\mathbb{R}^n) \) function.

In this case, for each \( g \in \text{BMO}(D) \) we can take \( f \in \text{BMO}(\mathbb{R}^n) \), \( f|D = g \), so that \( \| f \|_\ast \leq C \| g \|_\ast \), where \( C > 0 \) is a constant depending only on \( n \) and the constant of uniformness.

**Corollary 3.5.** Let \( D \) and \( D' \) be subdomains of \( \mathbb{R}^m \) and \( \mathbb{R}^n \) respectively. Assume that \( D' \) is uniform. Then for a measurable map \( F : D \to D' \), the following conditions are equivalent:

...
(a) We can take constants $K, \alpha > 0$ so that for an arbitrary pair of measurable subsets $E_1, E_2$ of $D'$ we have

$$\sup_{Q \subset D} \min_{k=1,2} \frac{|F^{-1}(E_k) \cap Q|}{|Q|} \leq K \left( \sup_{Q' \subset \mathbb{R}^n} \min_{k=1,2} \frac{|E_k \cap Q'|}{|Q'|} \right)^\alpha,$$

where the suprema are taken over all cubes $Q \subset D$ and $Q' \subset \mathbb{R}^n$ respectively;

(b) We can take constants $\gamma, 0 < \gamma < 1/4$, and $\lambda > 0$ so that for an arbitrary pair of measurable subsets $E_1, E_2$ of $D'$ satisfying

$$\sup_{Q' \subset \mathbb{R}^n} \min_{k=1,2} \frac{|E_k \cap Q'|}{|Q'|} < \lambda,$$

we have

$$\sup_{Q \subset D} \min_{k=1,2} \frac{|F^{-1}(E_k) \cap Q|}{|Q|} < \gamma,$$

where the suprema are taken over all cubes $Q \subset D$ and $Q' \subset \mathbb{R}^n$ respectively;

(c) $F$ is a BMO map;

(d) $G = i \circ F : D \to \mathbb{R}^n$ is a BMO map, where $i : D \to \mathbb{R}^n$ is the inclusion map.

We cannot replace the condition "$Q' \subset \mathbb{R}^n$" in (a) (and in (b)) above with "$Q' \subset D'$" (Example 4.11).

**Remark 3.6.** We may replace the assertion (a) of the Main Theorem with

(a') For each $N \geq 2$ we can take constants $K, \alpha > 0$ so that for arbitrary measurable subsets $E_1, \ldots, E_N$ of $D'$ we have

$$\sup_{Q \subset D} \min_{1 \leq k \leq N} \frac{|F^{-1}(E_k) \cap Q|}{|Q|} \leq K \left( \sup_{Q' \subset D'} \min_{1 \leq k \leq N} \frac{|E_k \cap Q'|}{|Q'|} \right)^\alpha,$$

where the suprema are taken over all cubes $Q \subset D$ and $Q' \subset D'$ respectively,

or

(a'') The assertion (a') holds for some $N \geq 2$.

Similarly, we may replace the assertion (b) of the Main Theorem with

(b') For each $N \geq 2$ we can take constants $\gamma, 0 < \gamma < 1/4$, and $\lambda > 0$ so that for arbitrary measurable subsets $E_1, \ldots, E_N$ of $D'$ satisfying

$$\sup_{Q' \subset D'} \min_{1 \leq k \leq N} \frac{|E_k \cap Q'|}{|Q'|} < \lambda,$$
we have
\[ \sup_{Q \subset D} \min_{1 \leq k \leq N} \frac{|F^{-1}(E_k) \cap Q|}{|Q|} < \gamma, \]
where the suprema are taken over all cubes \( Q \subset D \) and \( Q' \subset D' \) respectively,
or
(b') The assertion (b') holds for some \( N \geq 2 \).

The implications (a') \( \Rightarrow \) (a), (b') \( \Rightarrow \) (b), (a) \( \Rightarrow \) (b), (a') \( \Rightarrow \) (b'), and (a'') \( \Rightarrow \) (b'') are trivial. Furthermore, we obtain (a'') \( \Rightarrow \) (a) and (b'') \( \Rightarrow \) (b) by setting \( E_2 = E_3 = \cdots = E_N \). In the next section we show (b) \( \Rightarrow \) (c), and (c) \( \Rightarrow \) (a) (under the additive assumption). It is easy to check that we can show (c) \( \Rightarrow \) (a') in the same way.

Note that we can also rewrite Corollaries 3.2 and 3.5 similarly.

4. Proofs of the Main Theorem and Corollary 3.5.

The following two results play fundamental roles in the proof of the Main Theorem. The latter one shows that the growth estimation of BMO functions given by the former one is remarkably precise.

**Proposition 4.1** (John-Nirenberg \[9\]). Let \( f \in \text{BMO}(D) \), \( D \subset \mathbb{R}^n \), and \( Q \subset D \) be a cube. Then
\[ |\{ x \in Q : |f(x) - f_Q| \geq t \}| \leq C_1|Q| \exp \left( -C_2 \frac{t}{\|f\|_*} \right), \quad t \geq 0, \]
where \( C_1, C_2 > 0 \) are constants depending only on \( n \).

**Proposition 4.2** (Uchiyama \[17\], cf. Garnett-Jones \[4\]). Let \( D \) be an admissible subdomain of \( \mathbb{R}^n \). Let \( N \geq 2 \), \( t > 1 \), and \( E_1, \ldots, E_N \) be measurable subsets of \( D \) satisfying
\[ \sup_{Q \subset D} \min_{1 \leq k \leq N} \frac{|E_k \cap Q|}{|Q|} \leq 2^{-nt}, \]
where the supremum is taken over all cubes \( Q \subset D \). Then there exist \( \text{BMO}(D) \) functions \( f_1, \ldots, f_N \) satisfying \( \sum_{k=1}^N f_k = 1 \) and
\[ 0 \leq f_k \leq 1, \quad f_k = 0 \text{ on } E_k, \quad \|f_k\|_* \leq C/t, \quad (0 \leq k \leq N), \]
where \( C = C(n, N) > 0 \).

Garnett-Jones showed the assertion when \( D = \mathbb{R}^n \), \( N = 2 \) and \( E_1 \subset Q \), \( E_2 = \mathbb{R}^n \setminus 2Q \). Uchiyama extended their result to the form above.

First, we give a variant of the John-Nirenberg Theorem.
Lemma 4.3. Let \( f \in \text{BMO}(D) \), \( D \subset \mathbb{R}^n \), and \( Q \subset D \) be a cube. Then
\[
\min\{\{|x \in Q| f(x) \geq t\}, \{|x \in Q| f(x) \leq s\}\} \\
\leq C_1|Q|\exp\left(-C_2\frac{t-s}{\|f\|_*}\right), \quad -\infty < s \leq t < \infty,
\]
where \( C_1, C_2 > 0 \) are constants depending only on \( n \).

Proof. We may assume \( f_Q \leq (s + t)/2 \). Then from the John-Nirenberg theorem we have
\[
\left|\{x \in Q| f(x) \geq t\}\right| \leq \left|\left\{x \in Q| |f(x) - f_Q| \geq \frac{t-s}{2}\right\}\right| \\
\leq C|Q|\exp\left(-C_2\frac{t-s}{\|f\|_*}\right).
\]

Conversely,

Lemma 4.4. Let \( f : \mathbb{R} \to [0,1] \) be a nonconstant, non-decreasing function. Assume that there exist constants \( C_1, C_2 > 0 \) such that for each cube \( Q \subset D \) we have
\[
\min\{\{|x \in Q| f(x) \geq t\}, \{|x \in Q| f(x) \leq s\}\} \\
\leq C_1|Q|e^{-C_2(t-s)}, \quad -\infty < s \leq t < \infty.
\]

Then \( f \) is a \( \text{BMO}(D) \) function satisfying \( \|f\|_* \leq 4(C_1 + 1)C_2^{-1}\exp(2C_2) \).

This is a direct consequence of the following.

Lemma 4.5. Let \( \lambda : \mathbb{R} \to [0,1] \) be a nonconstant, non-decreasing function. Assume that there exist constants \( C_1, C_2 > 0 \) such that
\[
\min(\lambda(s), 1 - \lambda(t)) \leq C_1e^{-C_2(t-s)}, \quad -\infty < s \leq t < \infty.
\]

Then we can take \( t_0 \in \mathbb{R} \) so that
\[
\max(\lambda(t_0 - t), 1 - \lambda(t_0 + t)) \leq (C_1 + 1)e^{2C_2}e^{-C_2t}, \quad t \geq 0.
\]

Proof. Since \( \lambda \) is nonconstant, \( \lambda(t) \to 0 \) (\( t \to -\infty \)), and \( \lambda(t) \to 1 \) (\( t \to \infty \)). Let \( s_k = \sup\{t \mid \lambda(t) \leq 1 - \lambda(t + k)\}, k \geq 1 \). Then \( s_k \) is non-increasing, \( s_k + k \) is non-decreasing, and
\[
\lambda(s_k - 1) \leq 1 - \lambda(s_k + k - 1), \quad \lambda(s_k + 1) > 1 - \lambda(s_k + k + 1).
\]

Set \( t_0 = s_1 \). First, assume \( k \leq t < k + 1, k \geq 2 \). Then
\[
1 - \lambda(t_0 + t) \leq 1 - \lambda(s_{k-1} + k) \leq C_1e^{-C_2(k-1)} \leq C_1e^{2C_2}e^{-C_2t},
\]
\[
\lambda(t_0 - t) \leq \lambda(s_{k-1} - 1) \leq C_1e^{-C_2(k-1)} \leq C_1e^{2C_2}e^{-C_2t}.
\]

Next, if \( 0 \leq t < 2 \), then \( \max\{\lambda(t_0 - t), 1 - \lambda(t_0 + t)\} \leq 1 \leq e^{2C_2}e^{-C_2t} \). \( \square \)
Proof of Lemma 4.4. Set \( \lambda(t) = \frac{\{|x \in Q \mid f(x) \leq t\}|}{|Q|} \). Then \( \lambda \) satisfies the assumption of Lemma 4.5 with the same constants \( C_1, C_2 \). Thus we can take \( t_0 \) so that

\[
\max\{\lambda(t_0 - t), 1 - \lambda(t_0 + t)\} \leq (C_1 + 1)e^{2C_2}e^{-C_2t}, \quad t \geq 0.
\]

And so

\[
\mu(t) := \left\{\begin{array}{ll}
|\{x \in Q \mid f(x) - t_0 \geq t\}| \leq 2(C_1 + 1)|Q|e^{2C_2}e^{-C_2t}, & t \geq 0.
\end{array}\right.
\]

Hence

\[
\int_Q |f - f_Q|dx \leq 2\int_Q |f - f_0|dx = 2\int_0^\infty \mu(t)dt \leq 4(C_1 + 1)C_2^{-1}e^{2C_2}|Q|.
\]

\[\square\]

Lemma 4.6. Let \( F : D \to D' \) satisfy the condition (a) of the Main Theorem. Then \( K \geq 1 \) and \( \alpha \leq 1 \).

Proof. We obtain \( K \geq 1 \) by setting \( E_1 = E_2 = D' \).

Next, assume \( \alpha > 1 \). Let \( Q_0 = [p,q] \times P_0 \subset R \times R^{n-1} = R^n \) be a cube in \( D' \). Let \( l = q - p \). Let \( I_1 = [p, p + l/4], I_2 = [q - l/4, q] \). We decompose \( I \) into \( 2^s \) subintervals \( J_k = [p + 2^{-s}(k - 1)l, p + 2^{-s}kl], 1 \leq k \leq 2^s \), where \( s \) is a sufficiently large integer. Let \( E_k = (J_k \times R^{n-1}) \cap D' \). Let \( Q \) be a cube in \( D \). Let \( k_0 \) be the integer \( k \) which maximizes \( |F^{-1}(E_k) \cap Q|, 1 \leq k \leq 2^s \).

Then

\[
\sup_{Q' \subset D'} \min\left\{ \frac{|E_k \cap Q'|}{|Q'|}, \frac{|E_{k_0} \cap Q'|}{|Q'|} \right\} \leq \frac{4}{2^s},
\]

holds for each \( k \in \Sigma_1 \) or for each \( k \in \Sigma_2 \), where \( \Sigma_1 = \{k \mid J_k \subset I_1\} \) and \( \Sigma_2 = \{k \mid J_k \subset I_2\} \). Thus from the assumption we have

\[
\frac{|F^{-1}(E_k) \cap Q|}{|Q|} \leq K \left( \frac{4}{2^s} \right)^\alpha,
\]

for each \( k \in \Sigma_1 \) or for each \( k \in \Sigma_2 \). It follows from \( \sharp \Sigma_1 = \sharp \Sigma_2 = 2^s - 2 \) that

\[
\min_{j=1,2} |F^{-1}((I_j \times R^{n-1}) \cap D') \cap Q| \leq 2^{s-2}K \left( \frac{4}{2^s} \right)^\alpha |Q| \to 0, \quad s \to \infty.
\]

Therefore, the measure \( \mu(S) = |F^{-1}((S \times R^{n-1}) \cap D') \cap Q| \) on \([p,q]\) is absolutely continuous and satisfies \( \mu([a, a + t]) = 0 \) or \( \mu([a + 3t, a + 4t]) = 0 \) for each \( a, t \) with \( p \leq a < a + 4t \leq q \). Thus \( \mu([p,q]) = 0 \), and so \( |F^{-1}(Q_0) \cap Q| = 0 \). Since \( Q \) and \( Q_0 \) are arbitrary, we have \( |F^{-1}(D')| = 0 \), which is a contradiction. \[\square\]

Lemma 4.7. Let \( F : D \to D', D \subset R^n, D' \subset R^n \), be a BMO map. Then \( \|C_F\| \geq 1 \).
Proof. Note that if \( E \) is a measurable subset of \( D \) satisfying \(|E| > 0\), \(|D \setminus E| > 0\), then the characteristic function \( f \) of \( E \) satisfies \( \|f\|_* = \|f\|_{*,D} = 1/2 \).

Let \( F \) be a BMO map. Fix a cube \( Q \subset D \) and set \( \lambda(E) = |F^{-1}(E) \cap Q| \). Then \( \lambda \) is an absolutely continuous finite measure on \( D' \), thus we can take \( t_0 \in \mathbb{R} \) so that \( \lambda(E_0) = \lambda(D' \setminus E_0) = |Q|/2 \), where \( E_0 = ((-\infty, t_0] \times \mathbb{R}^{n-1}) \cap D' \). Let \( f \) be the characteristic function of \( E_0 \). Then \( f \circ F \) is the characteristic function of \( F^{-1}(E_0) \), and so from the first paragraph we have \( \|f\|_* = \|f \circ F\|_* = 1/2 \), which implies the assertion. \( \square \)

The following lemma shows \((a) \Rightarrow (c)\) of the Main Theorem with an estimation of the operator norm.

**Lemma 4.8.** Let \( F : D \to D' \), \( D \subset \mathbb{R}^m \), \( D' \subset \mathbb{R}^n \), be a measurable map. Assume that there exist constants \( K, \alpha > 0 \) such that for each pair of measurable subsets \( E_1, E_2 \) of \( D' \), we have

\[
\sup_{Q \subset D} \min_{k=1,2} \frac{|F^{-1}(E_k) \cap Q|}{|Q|} \leq K \left( \sup_{Q' \subset D'} \min_{k=1,2} \frac{|E_k \cap Q'|}{|Q'|} \right)^\alpha.
\]

Then \( F \) is a BMO map satisfying \( \|C_F\| \leq CK/\alpha \), where \( C = C(m,n) > 0 \).

**Proof.** Assume that \( F \) satisfies the assumption of the lemma. Then the inverse image of a null set is trivially a null set.

Let \( f \in \text{BMO}(D') \), \( E_1 = \{x \in D' \mid f(x) \leq s\} \), \( E_2 = \{x \in D' \mid f(x) \geq t\} \), \(-\infty < s \leq t < \infty\), and \( Q' \subset D' \). From Lemma 4.3 we have

\[
\min\{|E_1 \cap Q'|, |E_2 \cap Q'\| \leq C|Q'| \exp\left(-\frac{C(t-s)}{\|f\|_*}\right).
\]

It follows from the assumption and the fact \( \alpha \leq 1 \) (Lemma 4.6) that for an arbitrary cube \( Q \subset D \) we have

\[
\min\{|F^{-1}(E_1) \cap Q|, |F^{-1}(E_2) \cap Q| \} \leq CK|Q| \exp\left(-\frac{C\alpha(t-s)}{\|f\|_*}\right).
\]

Now, \( F^{-1}(E_1) = \{x \in D \mid f(F(x)) \leq s\} \), \( F^{-1}(E_2) = \{x \in D \mid f(F(x)) \geq t\} \), and so from Lemma 4.4 and the fact \( K \geq 1 \) (Lemma 4.6) we have \( F \circ f \in \text{BMO}(D) \) and

\[
\|f \circ F\|_* \leq \frac{CK}{\alpha} \|f\|_* \exp\left(\frac{C\alpha}{\|f\|_*}\right).
\]

Finally, applying this estimation to \( tf, t > 0 \), and letting \( t \to \infty \), we obtain \( \|f \circ F\|_* \leq \frac{CK}{\alpha} \|f\|_* \). \( \square \)

Next, to show \((b) \Rightarrow (c)\) of the Main Theorem, we need the following.
Proposition 4.9 (Strömberg [16]). Let $f$ be a measurable function on $D \subset \mathbb{R}^n$. Assume that we can take constants $\gamma$, $0 < \gamma < 1/2$, and $\lambda > 0$ so that for each cube $Q \subset D$ we have

$$\inf_{c \in \mathbb{R}} \{x \in Q \mid |f(x) - c| \geq \lambda\} \leq \gamma |Q|.$$ 

Then $f$ is a BMO($D$) function satisfying $\|f\|_* \leq C \lambda$, where $C = C(n, \gamma) > 0$.

Proof of Main Theorem (b) $\Rightarrow$ (c). Assume that $F$ satisfies the condition (b) of the Main Theorem. Then the inverse image of a null set is trivially a null set.

Let $f \in \text{BMO}(D')$. We may assume $\|f\|_* = 1$. Let $-\infty < s < t < \infty$, $E_1 = \{x \in D' \mid f(x) \leq s\}$, $E_2 = \{x \in D' \mid f(x) \geq t\}$, and $g = C_F(f)$. Then from Lemma 4.3 there exists $C_1 > 0$ such that if $t - s \geq C_1$, then

$$\sup_{Q' \subset D'} \min_{k=1,2} \frac{|E_k \cap Q'|}{|Q'|} \leq Ce^{-(t-s)/C_1} < \lambda,$$

and so

$$\sup_{Q \subset D} \min_{k=1,2} \frac{|F^{-1}(E_k) \cap Q|}{|Q|} < \gamma.$$

For $Q \subset D$ we set

$$s_Q = \sup \left\{ s \in \mathbb{R} \mid |\{x \in Q \mid g(x) \leq s\}| \leq |\{x \in Q \mid g(x) \geq s + C_1\}| \right\}.$$

Since $g \neq \pm \infty$ (a.e.), we have $s_Q \neq \pm \infty$. Thus

$$|\{x \in Q \mid g(x) \leq s_Q - 1\}| < \gamma, \quad |\{x \in Q \mid g(x) \geq s_Q + C_1 + 1\}| < \gamma,$$

and so if we set $c_Q = s_Q + C_1/2$ and $\delta = 1 + C_1/2$, then

$$|\{x \in Q \mid |g(x) - c_Q| > \delta\}| \leq 2\gamma \quad (< 1/2).$$

Hence $g \in \text{BMO}(D)$ by Proposition 4.9. \hfill \Box

Finally, we show the remaining implication (c) $\Rightarrow$ (a) of the Main Theorem.

Lemma 4.10. Let $F : D \rightarrow D'$, $D \subset \mathbb{R}^m$, $D' \subset \mathbb{R}^n$, be a BMO map. Assume that there exists an admissible domain $D'_0$ satisfying $F(D) \subset D'_0 \subset D'$. Then there exist constants $K$, $\beta > 0$ depending only on $m$ and $n$ such that for an arbitrary pair of measurable subsets $E_1, E_2$ of $D'$ we have

$$\sup_{Q \subset D} \min_{k=1,2} \frac{|F^{-1}(E_k) \cap Q|}{|Q|} \leq K \left( \sup_{Q' \subset D'} \min_{k=1,2} \frac{|E_k \cap Q'|}{|Q'|} \right)^{\beta/\|C_F\|}.$$
Proof. Let \( G : D \to D'_0 \), \( G(x) = F(x) \). Let \( g \in \text{BMO}(D'_0) \). Since admissible domains are uniform domains with uniformly bounded constants of uniformness, from Proposition 3.4 we can take \( f \in \text{BMO}(D') \) so that \( f|_{D'_0} = g \) and \( \| f \|_* \leq C \| g \|_* \). Thus \( \| g \circ F \|_* = \| f \circ F \|_* \leq C \| f \|_* \). Therefore, \( G \) is a BMO map satisfying \( \| C_G \| \leq C \| C_F \| \).

Let \( E_1, E_2 \) be an arbitrary pair of measurable subsets of \( D' \). Let \( E'_k = E_k \cap D'_0 \). Then

\[
2^{-nt} := \sup \min_{Q' \subset D'_0} \frac{|E'_k \cap Q'|}{|Q'|} \leq \sup \min_{Q' \subset D'} \frac{|E_k \cap Q'|}{|Q'|}.
\]

Because of Lemma 4.7, we may assume \( t > 1 \). From the Uchiyama theorem there exist BMO(\( D'_0 \)) functions \( f_1, f_2 \) satisfying \( f_1 + f_2 = 1 \) and

\[
0 \leq f_k \leq 1, \quad f_k = 0 \text{ on } E'_k, \quad \| f_k \|_* \leq C/t, \quad (k = 1, 2).
\]

Let \( g_k = f_k \circ G \). Then \( g_1 + g_2 = 1 \) and

\[
0 \leq g_k \leq 1, \quad g_k = 0 \text{ on } G^{-1}(E'_k), \quad \| g_k \|_* \leq C \| C_G \| / t, \quad (k = 1, 2).
\]

Let \( Q \) be an arbitrary cube in \( D \). Since \( (g_1)_Q + (g_2)_Q = 1 \), we may assume \( (g_1)_Q \geq 1/2 \). Then from the John-Nirenberg theorem we have

\[
|G^{-1}(E'_1) \cap Q| \leq \{|x \in Q | (g_1)_Q - (g_1)_Q \geq 1/2| \}
\leq C|Q| \exp \left( -\frac{C}{\| g_1 \|_*} \right) \leq C|Q| \exp \left( -\frac{Ct}{\| C_G \|} \right),
\]

and so we obtain

\[
\sup_{Q \subset D} \min_{k=1,2} \frac{|E^{-1}(E'_k) \cap Q|}{|Q|} = \sup_{Q \subset D} \min_{k=1,2} \frac{|G^{-1}(E'_k) \cap Q|}{|Q|} \leq C \left( \sup_{Q' \subset D'_0} \min_{k=1,2} \frac{|E'_k \cap Q'|}{|Q'|} \right)^\frac{\| C_G \|}{C/t} \leq C \left( \sup_{Q' \subset D'} \min_{k=1,2} \frac{|E_k \cap Q'|}{|Q'|} \right)^\frac{\| C_G \|}{\| C_F \|}.
\]

\( \Box \)

**Proof of Corollary 3.5.** (a) \( \Rightarrow \) (b) is trivial. (c) \( \Leftrightarrow \) (d) follows from Proposition 3.4. (d) \( \Rightarrow \) (a) is a consequence of the Main Theorem.

Finally, assume that (b) holds. Let \( E_1 \) and \( E_2 \) be measurable subsets of \( \mathbb{R}^n \) satisfying

\[
\sup_{Q' \subset \mathbb{R}^n} \min_{k=1,2} \frac{|E_k \cap Q'|}{|Q'|} < \lambda.
\]
Then $E'_k = E_k \cap D'$ satisfies the assumption. Thus
\[
\sup_{Q \subset D} \min_{k=1,2} \frac{|G^{-1}(E_k) \cap Q|}{|Q|} = \sup_{Q \subset D} \min_{k=1,2} \frac{|F^{-1}(E'_k) \cap Q|}{|Q|} < \gamma,
\]
and so $G$ satisfies the condition (b) of the Main Theorem. Hence (d) follows.

\[\square\]

**Example 4.11.** Let $D' = \{ x \in \mathbb{R}^2 \mid |x_1| < 4, |x_2| < 5 \} \setminus \{ x \in \mathbb{R}^2 \mid x_1 \geq 0, |x_2| < 1 \}$, where $x = (x_1, x_2)$. Then $D'$ is a uniform domain. Let $a_1 = (2, 3), a_2 = (2, -3), E_1 = \{|x-a_1| \leq 1\}$, and $E_2 = \{|x-a_2| \leq 1\}$. Let $D = \{ x \in \mathbb{R}^2 \mid |x_1| < 1, |x_2| < 1 \}$. Let $F : D \to D'$ be a conformal map. Then $F$ is a BMO map. On the other hand, $\min\{|E_1 \cap Q'|, |E_2 \cap Q'|\} = 0$ holds for each cube $Q' \subset D'$. Thus we cannot replace the condition “$Q' \subset \mathbb{R}^n$” in Corollary 3.5 with $Q' \subset D'$.

5. **Homeomorphisms between intervals.**

The Main Theorem gives a characterization of BMO maps between (finite or infinite) open intervals. Jones result (Example 2.3) implies that in the case of homeomorphisms of $\mathbb{R}$ we can reduce the condition to a much simpler form. The purpose of the present section is to show that his argument really characterizes BMO maps which are homeomorphisms between general open intervals with an explicit estimation on operator norms.

Recall that the space BMO($\mathbb{R}$) is invariant under Möbius transformations of $\overline{\mathbb{R}} = \mathbb{R} \cup \{ \infty \}$ (cf. Riemann-Rychener [14]). More generally:

**Lemma 5.1.** Let $I_1$ and $I_2$ be open intervals on $\mathbb{R}$. Let $\tau$ be a Möbius transformation of $\overline{\mathbb{R}}$ satisfying $\tau(I_1) = I_2$. Then $F = \tau|I_1 : I_1 \to I_2$ is a BMO map satisfying $C^{-1} \leq \|C_F(f)\|_\ast / \|f\|_\ast \leq C$, $f \in \text{BMO}(I_2)$, where $C > 0$ is a universal constant.

**Proof.** Let $f \in \text{BMO}(I_2)$. Let $I$ be an interval satisfying $2I \subset I_1$. Then $\max_I |F'| \leq C \min_I |F'|$, thus
\[
|I|^{-1} \int_I |f \circ F - f_{F(I)}|dx \leq C |F(I)|^{-1} \int_{F(I)} |f - f_{F(I)}|dx \leq C \|f\|_\ast,
\]
and so $\|C_F(f)\| \leq C \|f\|_\ast$ from the proposition below. \[\square\]

**Proposition 5.2** (cf. Riemann-Rychener [14]). Let $f \in L^1_{\text{loc}}(D), D \subset \mathbb{R}^n$, and $t \geq 1$. Assume that
\[
\sup_Q |Q|^{-1} \int_Q |f - f_Q|dx \leq \lambda
\]
holds for each $Q$ satisfying $tQ \subset D$. Then $f \in \text{BMO}(D)$ and $\|f\|_\ast \leq Ct\lambda$, where $C = C(n) > 0$. 
By virtue of the proposition above, repeating the argument of Jones, we can easily extend his result as follows.

**Theorem 5.3** (cf. Jones [11]). Let $F : I_1 \rightarrow I_2$ be a homeomorphism between open intervals.

(a) If $I_1 = \mathbb{R}$ and $I_2 \neq \mathbb{R}$, then $F$ is not a BMO map.

(b) If $I_1 \neq \mathbb{R}$ or $I_2 = \mathbb{R}$, then the following conditions are equivalent:

(i) $F$ is a BMO map;

(ii) There exist constant $K_0$, $\alpha_0 > 0$ such that for each measurable subset $E$ of $I_2$ and each subinterval $I$ of $I_2$ satisfying $2F^{-1}(I) \subset I_1$, we have

$$\frac{|F^{-1}(E \cap I)|}{|F^{-1}(I)|} \leq K_0 \left( \frac{|E \cap I|}{|I|} \right)^{\alpha_0}. \tag{7}$$

Moreover, if (7) holds, then $\|C_F\| \leq CK_0/\alpha_0$ for some universal constant $C > 0$, and conversely, if $F$ is a BMO map, then we can take constants $K_0$, $\alpha_0$ so that $K_0 = C_1$ and $\alpha_0 = C_2/\|C_F\|$, where $C_k > 0$, $k = 1, 2$, are universal constants.

In particular, $\|C_F\|$ and inf$(K_0/\alpha_0)$ are comparable with universal constant factors, where the infimum is taken over all pairs of $K_0$, $\alpha_0$ satisfying (7).

**Lemma 5.4.** Let $I_0$ be an interval. Let $J_1$ and $J_2$ be mutually disjoint subintervals of $I_0$. Then

$$\sup_{I \subset I_0} \min_{k=1,2} \frac{|J_k \cap I|}{|I|} = \frac{s}{d(J_1, J_2) + 2s},$$

where $s = \min_{k=1,2} |J_k|$.

**Proof of Theorem 5.3.** First, assume $I_1 = \mathbb{R}$ and $I_2 \neq \mathbb{R}$. From the Möbius invariance of BMO we may assume $I_2 = (0, \infty)$ and $F$ is sense preserving. Let $f(x) = \log x$. Then $f \in \text{BMO}(I_2)$. On the other hand, since $g = f \circ F$ is an increasing function satisfying $\lim_{x \to -\infty} g(x) = \infty$, $\lim_{x \to -\infty} g(x) = -\infty$, if we set $J_k = [-k, k]$ and $E_k = [-k, -k/2] \cup [k/2, k]$, then

$$|J_k|^{-1} \int_{J_k} |g - g_{J_k}| dx \geq (2k)^{-1} \int_{E_k} \geq 4^{-1} (g(k/2) - g(-k/2)) \to \infty,$$

as $k \to \infty$. Thus $F$ is not a BMO map.

Next, assume that $F$ satisfies the condition (ii) of (b). Then $K_0 \geq 1$ and $\alpha_0 \leq 1$. Let $I$ be a subinterval of $I_2$ satisfying $2F^{-1}(I) \subset I_1$. Let $f \in \text{BMO}(I_2)$, $I' = F^{-1}(I)$, $g = f \circ F$, and $E_t = \{x \in I \mid |f(x) - f_I| \geq t\}$, $t \geq 0$. Then from the John-Nirenberg theorem we have $|E_t| \leq C |I| \exp(-Ct/\|f\|_s)$,
thus
\[
\mu(t) := |\{x \in I' \mid |g(x) - f_t| \geq t\}| = |F^{-1}(E_t)| \\
\leq CK_0|I'| \exp \left( -\frac{C\alpha_0 t}{\|f\|_*} \right)
\]
and so
\[
|I'|^{-1} \int_{I'} |g - f_t| dx = \int_0^\infty \mu(t) dt \leq \frac{CK_0}{\alpha_0} \|f\|_*.
\]
Hence, from Proposition 5.2 \(\|g\|_* \leq CK_0\alpha_0^{-1}\|f\|_*\).

Finally, assume that \(F\) satisfies the condition (i) of (b). We may assume that \(F\) is sense preserving. Let \(I_k = (p_k, q_k), k = 1, 2\). Let \(E\) be a measurable subset of \(I_2\). Let \(I = [a, b]\) be an interval satisfying \(2F^{-1}(I) \subset I_1\). Let \(l = |I|\) and \(l' = |F^{-1}(I)|\). Let \(E_1 = E \cap [a, a + l/2]\) and \(E_2 = [b, q_2]\). Then \(s := \min_{k=1,2} |F^{-1}(E_k)| \geq |F^{-1}(E_1)|/2\), thus from Lemma 5.4
\[
\sup_{I' \subset I_2} \min_{k=1,2} \frac{|E_k \cap I'|}{|I'|} \leq \frac{\min_{k=1,2} |E_k|}{l/2 + 2\min_{k=1,2} |E_k|} \leq \frac{2|E_1|}{l} \leq \frac{2|E \cap I|}{l},
\]
\[
\sup_{I' \subset I_1} \min_{k=1,2} \frac{|F^{-1}(E_k) \cap I'|}{|I'|} \geq \frac{s}{(l' - |F^{-1}(E_1)|)} + 2s \geq |F^{-1}(E_1)|/2l'.
\]
And so from the Main Theorem we have
\[
\frac{|F^{-1}(E_1)|}{l'} \leq C \left( \frac{|E \cap I|}{l} \right)^{C/\|C_F\|}.
\]
Since the same estimation holds for \(E'_1 = E \cap [a + l/2, b]\), we obtain (7). \(\square\)

Contrary to the case of homeomorphisms of \(\mathbb{R}\), no homeomorphism \(F : \mathbb{R} \to (0, \infty)\) is a BMO map even if \(F^{-1}\) is a BMO map. Moreover, there exists a homeomorphism \(F : (0, \infty) \to (0, \infty)\) such that \(F\) is a BMO map and \(F^{-1}\) is not a BMO map. One such example is given by \(F(x) = \log(1 + x)\) (cf. Corollary 5.9). This example also shows that we cannot drop the condition \(2F^{-1}(I) \subset I_1\) in the statement above: Assume that the estimation (7) holds for \(I = [\log 2, \log(a + 1)]\) and \(E = [\log(a/2 + 1), \log(a + 1)], a > 1\). Then
\[
\frac{a/2}{a - 1} \leq K_0 \left( \frac{\log((a + 1)/(a/2 + 1))}{\log((a + 1)/2)} \right)^{\alpha_0}, \quad a > 1,
\]
which is a contradiction. Another such example is given by \(F : (0, \infty) \to (0, \infty), F(x) = 1/x\).

Applying (7) to subintervals \(I = [a, b]\) and \(E = [x, y]\) of \(I_2, a \leq x \leq y \leq b\), we obtain:

**Corollary 5.5.** Let \(F : I_1 \to I_2\) be a homeomorphism between open intervals. Assume that \(F\) is a BMO map. Then \(F^{-1}\) is locally a Hölder continuous function of order \(C/\|C_F\|\), where \(C > 0\) is a universal constant.
The homeomorphism $F(x) = |x|^p \text{sgn} x$, $p \geq 1$, of $\mathbb{R}$ shows that the estimation above is best possible. (See Example 5.8 below.)

Recall that for a homeomorphisms $G : J_1 \to J_2$ between (finite) closed intervals, $G$ is absolutely continuous and $G'$ is an $A_\infty$ weight on $J_1$ if and only if $G^{-1}$ is absolutely continuous and $(G^{-1})'$ is an $A_\infty$ weight on $J_2$ (cf. Coifman-Fefferman [2]). Thus, under the assumption of the corollary above, $F'$ (or $-F'$) satisfies the $A_\infty$ condition uniformly on each $I$ satisfying $2I \subset I_1$. In particular, $F' \in L^p_{\text{loc}}(I_1)$ holds for some $p > 1$, and so $F$ is also locally Hölder continuous. We do not know, however, whether the similar estimation holds or not for the order of $F'$. Moreover, from Proposition 5.2 we have:

**Corollary 5.6.** Let $F : I_1 \to I_2$ be a homeomorphism between open intervals. Assume that $F$ is a BMO map. Then we have $\| \log |F'| \|_* \leq C$, where $C = C(\| CF \|) > 0$.

It is easy to see that the corresponding result does not hold for $\log |(F^{-1})'|$.

In the rest of the present section, we give a remark on the global behavior of BMO maps. For an open interval $I = (a, b)$ ($\neq \mathbb{R}$), the hyperbolic metric $d_{sh}$ is defined by

$$d_{sh}(x) = \frac{(b-a)dx}{(b-x)(x-a)}.$$ 

The hyperbolic metric is invariant under Möbius transformations of $\mathbb{R}$ and comparable with the quasihyperbolic metric $dx/d(x, \partial I)$. Let $d_h(x, y)$ denote the hyperbolic distance between $x$ and $y$.

**Lemma 5.7.** Let $F : I_1 \to I_2$ ($I_1, I_2 \neq \mathbb{R}$) be a homeomorphism between open intervals. Assume that $F$ is a BMO map. Then

$$d_h(F(x), F(y)) \leq C \| CF \|(d_h(x, y) + 1), \quad x, y \in I_1,$$

where $C > 0$ is a universal constant.

**Proof.** Since both the quasihyperbolic metric and BMO are invariant under Möbius transformations, we may assume that $I_1 = I_2 = (0, \infty)$, $F$ is sense preserving, $x = 1 = F(1)$, and $y = a > 1$. Applying the Main Theorem with $E_1 = (0, 1]$ and $E_2 = [F(a), \infty)$, and utilizing Lemma 5.4, we get

$$\frac{1}{a+1} \leq C \left( \frac{1}{F(a)+1} \right)^{C/\| CF \|},$$

hence $F(a) \leq Ca^{C/\| CF \|}$, which implies the assertion. □

Each hyperbolically Lipschitz continuous function is a BMO function and $F$ preserves the space of all hyperbolically Lipschitz continuous functions if and only if $F$ is hyperbolically Lipschitz continuous. The lemma above shows that the similar result holds for BMO. Note that quasiconformal maps
satisfy the corresponding estimation with respect to the quasihyperbolic metric. The following example shows that the estimation (8) is best possible.

**Example 5.8.** Let $F : (0, \infty) \to (0, \infty)$, $F(x) = x^p$, $p \geq 1$. $F$ is the hyperbolic dilation centered at the point 1: $d_h(F(x), F(y)) = pd_h(x, y)$, $x, y \in (0, \infty)$. A simple calculation shows $F$ satisfies (7) with $K_0 = 1$ and $\alpha_0 = 1/p$. On the other hand, if we set $f(x) = \log x$, then $C_F(f) = pf$. Hence, $\|C_F\|$ and $p$ are comparable with universal constant factors. Note that as to the antisymmetric extension $F_1 : \mathbb{R} \to \mathbb{R}$, $F_1(x) = |x|^p \text{sgn} x$, of $F$, $\|C_{F_1}\|$ and $p$ are comparable with universal constant factors similarly.

**Corollary 5.9.** Let $F : I_1 \to I_2$, $I_1, I_2 \neq \mathbb{R}$, be a homeomorphism between open intervals. Assume that $F$ is a BMO map. Then $C_F$ is a bijection between $\text{BMO}(I_2)$ and $\text{BMO}(I_1)$ if and only if we can take a constant $C > 0$ so that for each interval $I \subset I_1$ satisfying $d(I, \partial I_1) = |I|$, we have $d(F(I), \partial I_2) \leq C|F(I_2)|$.

Note that we can take such a constant $C > 0$ if and only if

$$C^{-1} \leq \frac{d_h(F(x), F(y)) + 1}{d_h(x, y) + 1} \leq C, \quad x, y \in I_1,$$

holds for some $C \geq 1$.

**Proof.** Assume that we can take such a constant $C > 0$. Let $f \in \text{BMO}(I_1)$. Then for each $I \subset I_1$ satisfying $d(I, \partial I_1) = |I|$, $\|f \circ F^{-1}\|_{\text{BMO}(I)} \leq C$ holds, thus from Proposition 5.2 we have $f \circ F^{-1} \in \text{BMO}(I_2)$, and so $F^{-1}$ is a BMO map. The converse assertion easily follows from Lemma 5.7. \hfill $\Box$

**References**


COMPOSITION OPERATORS WHICH PRESERVE BMO


Received December 21, 1999 and revised June 3, 2000.

DEPARTMENT OF MATHEMATICS
NATIONAL DEFENSE ACADEMY
HASHIRIMIZU 1-10-25
YOKOSUKA 239
JAPAN
E-mail address: gotoh@cc.nda.ac.jp
ANALYTIC CONTINUATION OF CONVEX BODIES AND FUNK’S CHARACTERIZATION OF THE SPHERE

ERIC L. GRINBERG AND ERIC TODD QUINTO

A celebrated theorem of P. Funk, 1916, states that an origin-centered star body in $\mathbb{R}^3$ is determined by the areas of its central hyperplane cross-sections. In particular, if all these concurrent sections have the same area then the body must be a ball (its boundary is a sphere). It is natural to try to strengthen the theorem by using a smaller class of planes. It is evident that a lower-dimensional class of hyperplanes, e.g., planes passing through an axis, does not suffice, but a proper open subset of planes appears plausible. The class of planes at a small angle relative to an axis has been considered in the literature. We show that this class does not characterize the body. We then show that if a body is known to osculate a ball centered at the origin to infinite order along one hyperplane through the axis, then the proper open class of planes above does determine whether the body is a ball. We generalize our theorem to arbitrary origin centered star bodies and to any open connected collection of planes that fills out $\mathbb{R}^n$. We have counterexamples to the theorem for every finite order of osculation. We have similar theorems for the cosine transform and projection areas.

0. Introduction.

The goal of this paper is to use Radon transforms to answer specific questions about star-shaped and convex sets. Funk showed that a radially symmetric convex body in $\mathbb{R}^3$ is determined by areas of its intersections with all planes through the origin. A natural question is whether there are subclasses of planes for which the areas of intersection are sufficient to determine the body. One might think that planes that are within an angle of $\Theta \in (0, \pi/2)$ of an axis might determine the object, and this claim is in the literature [T]. This is reasonable since the set of all planes near an axis is a full-dimensional set, and every point in the object meets an infinite number of such planes.

In fact, as we show in Proposition 2.2, this does not determine if an object is a ball (i.e., its boundary is a sphere). In fact, Proposition 2.2 shows that even if we assume the object is a ball to any finite order along the hyperplane, it does not have to be a ball everywhere. As a positive result, we prove that
if an origin-centered star-shaped body is a ball to infinite order along one hyperplane meeting the axis and if its areas are constant on all hyperplanes sufficiently near the axis, then the body is a ball (Theorem 1.2). In fact, our theorem holds for arbitrary origin-centered star bodies that agree to infinite order along a hyperplane through the axis (Theorem 1.3) and for any open connected set of hyperplanes through the origin that fills out $\mathbb{R}^n$. We prove similar uniqueness theorems for the cosine transform and projection areas (Theorems 3.1–3.3).

The authors thank Richard Gardner for useful comments and helpful references, in particular to [SW]. This work has also benefited from discussions with Mark Agranovsky, Fulton Gonzalez, Joram Lindenstrauss, and Larry Zalcman. Proposition 2.3 came out of a suggestion of Michael Gage and Frank Morgan, and Maria Moszyńska provided helpful editorial comments. A pleasant conversation and insightful questions by Wolfgang Weil led to a simplification of the proof of Theorem 1.1 and to an improved Proposition 2.1. The authors thank Boris Rubin for specific information about [Mi]. The second author thanks the Humboldt Stiftung for its support and Prof. A.K. Louis and the Fachbereich Mathematik der Universität Saarlandes for their hospitality.

1. Radon transforms and cross-sections of star bodies.

A star body is a body in $\mathbb{R}^n$ that is star-shaped about the origin; that is, the origin is connected to every point in the body by a segment (which could be just a point) contained in the body. A body $K$ is origin centered if $K = -K$. The sets we consider will be assumed to be compact. We will say two sets are equal when they agree up to a set of measure zero. There is a natural one-to-one correspondence between a star body and its bounding surface, which we will use when convenient.

There is an identification between star shaped bodies in $\mathbb{R}^n$ and certain functions on the sphere $S^{n-1}$. If $K$ is a star body in $\mathbb{R}^n$ then its radial function $\rho_K$ is defined by

$$\rho_K(\omega) = \max\{t \in [0, \infty) \mid t\omega \in K\}$$

where $\omega \in S^{n-1}$. The number $\rho_K(\omega)$ is just the (maximum) distance from the origin to the boundary of $K$ in the direction $\omega$.

For $\omega \in S^{n-1}$ (or for $\omega \in \mathbb{RP}^{n-1}$) let $P(\omega)$ be the hyperplane through the origin perpendicular to $\omega$ and let $\omega^\perp = P(\omega) \cap S^{n-1}$ be the great sphere in $S^{n-1}$ perpendicular to $\omega$.

The Funk-Radon transform for functions $f \in C(S^{n-1})$ is defined by

$$Rf(\omega) = \int_{\eta \in \omega^\perp} f(\eta) \, d\eta.$$
Thus $R$ operates by integration on great $(n-2)$ spheres in $S^{n-1}$ with respect to the normalized rotation invariant measure. Call a function $f(\omega)$ even if $f(-\omega) = f(\omega)$. By inspection, $Rf$ is always an even function. Minkowski [Mi] proved injectivity of this transform for functions with integrable second derivative. Funk [F] proved an inversion formula. Both results imply the following uniqueness theorem.

**Funk-Minkowski Theorem ([F, Mi]).** $R$ is injective on even functions, $L^2_{\text{even}}(S^{n-1})$.

It is obvious that $Rf$ is zero if $f$ is an odd function. This theorem shows that the null space of $R$ is the set of odd functions. Since $Rf$ is always even, we will sometimes view $R$ as an injective transform from $L^2(RP^{n-1})$ to $L^2(RP^{n-1})$.

The link between cross-sections of star bodies, $K$, and the transform $R$ is given by the polar coordinate area formula

\[(1.3) \quad \text{Area}(K \cap P(\omega)) = \frac{1}{n} \int_{\eta \in \omega^\perp} (\rho_K(\eta))^{n-1} d\eta = R\left((1/n) \rho_K^{n-1}\right)(\omega).\]

This, in turn, follows from integration in polar coordinates. For background information about star and convex bodies see [G3]. In particular, since star bodies are determined by their radial functions and the Funk-Radon transform is injective on even functions we can make the following assertion.

*Origin-centered star bodies in $\mathbb{R}^n$ are determined by the areas of their central cross-sections.*

Funk’s original proof of his theorem uses spherical harmonic expansions and is valid on $S^2$. After many years it remains the standard proof though others have emerged. See, e.g., [BEGM]. For even functions, Helgason [He1] proved this theorem and inversion formulas for odd dimensional spheres and projective spaces, and Semyanistyi [Se] proved injectivity for spheres of all dimensions using Fourier techniques. Grinberg [Gr] proved injectivity and the inversion formula for spheres of all dimensions (and projective spaces) using spherical harmonics and group invariance.

We need two definitions about asymptotic behavior of functions and sets near a given hypersurface.

**Definition 1.1.** Let $f$ be a function on a Riemannian manifold and let $S$ be a closed hypersurface. The function $f$ is zero to order $k \in \mathbb{N}$ on $S$ if and only if $f$ is continuous near $S$ and $f(x) = O((\text{dist}(x,S))^k)$ where $\text{dist}(x,S)$ is the minimum geodesic distance from $x$ to $S$. The function $f$ is zero to infinite order on $S$ if and only if $f$ is zero to order $k$ for all $k \in \mathbb{N}$.

**Definition 1.2.** *Order of osculation among star bodies.* Let $K_1$ and $K_2$ be star bodies in $\mathbb{R}^n$, and let $H_0$ be a hyperplane through the origin. Assume $K_1 \cap H_0 = K_2 \cap H_0$. Let $k \in \mathbb{N} \cup \{\infty\}$. Then, $K_1$ and $K_2$ osculate to order...
Along $H_0$ if and only if the difference of radial functions, $\rho_{K_1} - \rho_{K_2}$ is zero to order $k$ on $S^{n-1} \cap H_0$ according to Definition 1.1.

If the bodies $K_1$ and $K_2$ have smooth boundaries, then the condition in Definition 1.2 is equivalent to the boundaries of $K_1$ and $K_1$ being tangent to order $k$ along $K_1 \cap H_0$.

Our first theorem, a support theorem for the Funk-Radon transform, is the key to the subsequent theorems.

**Theorem 1.1.** Let $A$ be an open connected subset of $S^{n-1}$. Assume $f$ is an even function in $C_c(S^{n-1})$ with $Rf(\omega) = 0$ for all $\omega \in A$ and assume, for some $\omega_0 \in A$, $f$ is zero to infinite order on $P(\omega_0)$. Then, $f$ is zero on $\cup_{\omega \in A} \omega^\perp$.

This theorem can be proven using microlocal techniques [GS] for any generalized Funk-Radon transform (in which the measure $d\eta$ (1.2) is replaced by any nowhere zero real-analytic weight). The transform (1.2) is shown to satisfy a microlocal condition, the Bolker Condition, in [Q1]. Theorem 1.1 is a special case of Theorem 2.2 in [Q3], which is true for all Radon transforms satisfying the Bolker Condition.

However, the proof we give is elementary and it involves the standard projection from a half sphere in $S^{n-1}$ to $\mathbb{R}^{n-1}$. This projection has been used to prove support theorems for the hyperplane and the great-sphere Radon transforms in [Q2] and [Bo], and it was noted by Gelfand in the sixties.

**Proof of Theorem 1.1.** We rotate $S^{n-1} \subset \mathbb{R}^n$ so that $\omega_0^\perp$ is on the $(x_1, \ldots, x_{n-1})$ hyperplane. Then, we consider the projection from the origin through each point in the open upper half sphere $S^+$ to the $(n-1)$-plane $x_n = 1$. We identify this plane with $\mathbb{R}^{n-1}$. Under this projection each point $\omega \in \text{int} S^+$ is mapped to the point $x = \frac{1}{\omega_n} \omega$ which is identified with its first $n-1$ coordinates, $x' = (x_1, \ldots, x_{n-1})$. The great sphere perpendicular to $\omega$ is mapped to the $n-2$ dimensional hyperplane on $\mathbb{R}^{n-1}$ normal to $\omega' = (\omega_1, \ldots, \omega_{n-1})$ and of distance $\omega_n/\sqrt{1 - \omega_n^2}$ from the origin (in the opposite direction to $\omega'$).

Let $R$ be the classical hyperplane transform on $\mathbb{R}^{n-1}$. If $f$ is a continuous even function on $S^{n-1}$ and $f$ is its projected function on $\mathbb{R}^{n-1}$, and $\omega \in S^{n-1}$, then

\begin{align*}
Rf(\omega) &= \sqrt{1 + p^2} R(\tilde{f}(x)) (1 + |x|^2)^{-(n-1)/2} (\tau, p) \\
\text{where} \quad \tilde{f}(x) &= f \left( \frac{(x, 1)}{\| (x, 1) \|} \right), \quad \tau = \frac{\omega'}{|\omega'|}, \quad p = \frac{-\omega_n}{\sqrt{1 - \omega_n^2}}.
\end{align*}

The $p$ and $x$ factors in (1.4a) come from the change of coordinates.
Note that if \( f \) is zero to infinite order on \( \omega_0^\perp \), then \( \tilde{f}(x)(1 + |x|^2)^{-(n-1)/2} \) is rapidly decreasing at infinity. Let \( \mathcal{A} \) be the set of \( n-2 \) dimensional hyperplanes in \( \mathbb{R}^{n-1} \) that correspond to the set of great spheres parameterized by \( A \). As \( Rf = 0 \) on \( A \), \( R(\tilde{f}(1 + |x|^2)^{-(n-1)/2}) \) is zero for all hyperplanes in \( \mathcal{A} \). As \( A \) is open and \( \omega_0 \in A \), a whole neighborhood of \( \omega_0 \) is in \( \mathcal{A} \). Since the great sphere \( \omega_0^\perp \) corresponds to the hyperplane at infinity in \( \mathbb{R}^{n-1} \), \( \mathcal{A} \) contains all hyperplanes outside of some ball, \( B \), in \( \mathbb{R}^{n-1} \). Now, we can use Helgason’s support theorem for the hyperplane transform [He2, Theorem 2.6], which is true for rapidly decreasing functions that are continuous near infinity, to conclude \( \tilde{f} = 0 \) outside \( B \). Finally, we can use the main theorem of [BQ] (or a generalization of [He2, Lemma 2.11]) to conclude \( \tilde{f} \) is zero on \( \cup \mathcal{A} \). Therefore, \( f \) is zero on \( \cup (\omega \in \mathcal{A}) \omega^\perp \). □

Theorem 1.1 has the following consequence for star bodies in \( \mathbb{R}^n \).

**Theorem 1.2.** Let \( K \) be an origin centered star body in \( \mathbb{R}^n \) and let \( A \) be an open connected set of hyperplanes through the origin so that \( \mathbb{R}^n = \cup A \). If for some \( H_0 \in A \), \( K \) osculates a ball centered at the origin to infinite order along \( H_0 \) and \( K \) has constant cross-sectional areas when sliced by planes in \( A \), then \( K \) is that ball.

Note that the collection of planes in \( \mathbb{R}^3 \) within an angle of \( \Theta \in (0, \pi/2) \) of an axis satisfies the hypotheses of Theorem 1.2, so this theorem shows that any set that osculates a ball centered at the origin to infinite order on one 2-plane through the axis and with constant cross-sectional areas on these planes is a ball. Thus, under the additional osculation assumption, the theorem discussed in the introduction [T] holds.

**Proof of Theorem 1.2.** Let \( A \) be the connected set of directions in \( RP^{n-1} \) normal to hyperplanes in \( \mathcal{A} \), \( \mathcal{A} = \{ P(\omega) \mid \omega \in A \} \). Let \( K \) be an origin centered star body. Assume \( K \) osculates a ball centered at the origin to infinite order along a plane \( H_0 \), then \( \rho_K^{n-1} \) is constant to infinite order on \( H_0 \cap S^{n-1} \). If the cross-sectional areas of \( K \) are constant on all planes in \( \mathcal{A} \), including \( H_0 \), then the Funk transform of \( \rho_K^{n-1} \) agrees with that of a ball on \( A \). By Theorem 1.1, we can conclude the radial function of \( K \) is constant on all unit vectors on planes in \( \mathcal{A} \). By the assumption that \( \mathbb{R}^n = \cup A \), the radial function of \( K \) must be constant on \( S^{n-1} \) and so \( K \) is a ball. □

This proof is valid for any pair of origin-centered star bodies that osculate each other to infinite order along a hyperplane, so we obtain the following theorem.

**Theorem 1.3.** Let \( K_1 \) and \( K_2 \) be origin-centered star bodies in \( \mathbb{R}^n \) and let \( A \) be an open connected set of hyperplanes through the origin so that \( \mathbb{R}^n = \cup A \). If for some \( H_0 \in A \), \( K_1 \) and \( K_2 \) osculate each other to infinite order along \( H_0 \), then \( K_1 \) and \( K_2 \) are congruent.
order along $H_0$, and their cross-sectional areas agree when sliced by planes in $\mathcal{A}$, then $K_1 = K_2$.

Simple counterexamples (related to the non-uniqueness of the interior Radon transform on lines) show that the connectedness assumptions in these theorems are necessary.

2. Counterexamples.

In this section we explore the hypotheses of the characterization Theorems 1.2 and 1.3, and show through counterexamples that they are needed. First we ask if the ‘starter’ hypothesis (the condition that the sets agree to infinite order along a hyperplane) can be eliminated entirely. This can be answered using mid-decade developments involving intersection bodies. If $K$ is a star body in $\mathbb{R}^n$ then the Funk-Radon transform of its radial function defines another body $IK$, the intersection body of $K$:

$$\rho_{IK}(\omega) = R(\rho^n_K)(\omega),$$

up to normalizations. H. Busemann proved that if $K$ is a centered convex set, then $IK$ is convex [G3]. Given another star body $L$ it is natural to ask if $L$ is expressible as $IK$ for a star body $K \subset \mathbb{R}^n$. In dimension three the answer is affirmative for origin centered bodies [G2]. In higher dimensions the answer is negative [Ko1, G1, G2, Zh, GZ]. These results played a critical role in the solution of the celebrated Busemann-Petty problem. This problem involves the comparison of volumes of two convex bodies by means of comparison of areas of lower-dimensional cross-sections.

In the present context the surjectivity of the intersection body transform in dimension three shows that the ‘starter’ hypothesis above cannot be entirely eliminated. For the values of central cross-sectional areas of the star body $K$ give the values of the radial function of $IK$. If one could remove the starter hypothesis in Theorem 1.2, then radial functions of star bodies would enjoy a continuation property. Since, in dimension three, every centered convex body is an intersection body, no such continuation can exist. In particular, it is not possible to determine a star body from its central cross-sections by planes with relative angle less than $\Theta$ with respect to an axis since this would imply that an intersection body is determined by its intersection with a central slab. Though intersection bodies form a proper subclass in higher dimensions it is nonetheless possible to show that no suitable continuation property can exist there either.

We now turn to the more delicate question involving the severity of the starter assumption. We will construct a body with constant cross-sections for planes at a small angle relative to an axis which osculates the unit ball to finite order along a plane containing the axis.
Proposition 2.1. Let $k \in \mathbb{N}$ and let $\Theta \in (0, \pi/2)$. There is a non-zero function $f \in C^\infty_{\text{even}}(S^{n-1})$ which vanishes to order $k$ on the great sphere $e_n^\perp$ (which contains the $x_1$-axis) and whose integrals vanish on all great spheres within angle $\Theta$ of the $x_1$-axis.

Proof. We use the relation, (1.4), between the great sphere transform and the classical Radon transform on hyperplanes in $\mathbb{R}^{n-1}$, $\mathbb{R}$, and then a range theorem of Solmon for functions that satisfy a finite number of moment conditions.

Let $H(\tau, p)$ be the hyperplane in $\mathbb{R}^{n-1}$ perpendicular to $\tau \in S^{n-2}$ and $p \geq 0$ units from the origin. Then, under the correspondence (1.4),

$$H(\tau, p) \text{ corresponds to the great sphere } \omega^\perp \subset S^{n-1}$$

where $\omega = \left( \frac{1}{\sqrt{1+p^2}} \tau, \frac{-p}{\sqrt{1+p^2}} \right) \in S^{n-1}$.

So, the set of great spheres within $\Theta$ radians of the $x_1$-axis correspond to $\omega \in S^{n-1}$ with $|\omega_1| < \cos \Theta$. Under (1.4) and (2.1), this set of great spheres corresponds to hyperplanes for $(\tau, p)$ in a subset of the set $\{(\tau, p) \mid |\tau_1| < a \text{ or } |p| > b\}$ for some $a > 0$, $b > 0$.

We now construct a function $g$ satisfying:

(2.2a) $g \in C^\infty(S^{n-2} \times \mathbb{R})$

(2.2b) $g(\tau, p) = g(-\tau, p) = g(-\tau, -p)$

(2.2c) $g$ is supported in $C = \{(\tau, p) \mid |\tau_1| > a \text{ and } 0 < |p| < b\}$

(2.2d) $\int_{p=\pm\infty} g(\tau, p)p^m dp = 0 \ \forall \tau \in S^{n-2}, m = 0, 1, \ldots, k$.

Let $g_1(\tau)$ be a smooth even function that is supported in $\{\tau \in S^{n-2} \mid |\tau_1| > a\}$. It is easy to construct a function $g_2(p) \in C^\infty((0, b))$ that satisfies $\int_{p=\pm\infty} g_2(p)p^m dp = 0$, $\forall m = 0, \ldots, k$. For example, one can just take an appropriate linear combination of $k+2$ translates of a small bump function supported in $(0, b/(k+4))$. Now, extend $g_2$ to $\mathbb{R}$ to be even. Let $g(\tau, p) = g_1(\tau)g_2(p)$ and then $g$ satisfies (2.2a)–(2.2d).

Now, we use Theorem 7.7, p. 376 of [So] to conclude that $g = \mathcal{R} h$ for some $h \in C^\infty(\mathbb{R}^{n-1})$. Furthermore, because $g$ satisfies the first $k$ moment conditions, (2.2d),

(2.3) $h = \mathcal{O}(|x|^{-n-k})$ at infinity.

Next, we let $f : S^{n-1} \rightarrow \mathbb{R}$ be defined by

(2.4) $f(\omega) = h \left( \frac{\omega'}{\omega_n} \right) |\omega_n|^{-(n-1)}$
Then $f$ is zero to order $k$ on $e_{n}^{\perp}$ by the relation of $f$ to $h$, (2.4), and the growth condition on $h$, (2.3) (and since $|x| = 1/|\omega_n|$). Also, $f$ is continuous on $S^{n-1}$ and $f$ is smooth away from $e_{n}^{\perp}$.

Finally, we look back at (1.4) to see that

$$Rf(\omega) = (1 - \omega_n^2)^{-1/2}g(\omega'/|\omega'|, -\omega_n/\sqrt{1 - \omega_n^2}).$$

Because of (1.4) and the definition of $g$, $Rf(\omega) = 0$ for $|\omega_1| < \cos \Theta$. By (2.2b)–(2.2c) $g(\tau, p)$ is zero near $p = 0$ ($\omega_n = 0$) and $g$ is zero near $p = \infty$ ($\omega_n = 1$) and $g$ is even. Therefore, $Rf$ is smooth and even on $S^{n-1}$. Since $R$ is bijective on $C_{even}^{\alpha}(S^{n-1})$ (and injective on $C(S^{n-1})$), $f$ is smooth. This finishes the construction. □

Let $B_n$ be the unit ball in $\mathbb{R}^n$. Our next proposition is an application Proposition 2.1 to convex sets.

**Proposition 2.2.** Let $k \in \mathbb{N}$ and let $\Theta \in (0, \pi/2)$. Let $H_0$ be the hyperplane in $\mathbb{R}^n \times_n = 0$. There is a strictly convex origin-centered set, $K$, with smooth boundary that is not a ball but that osculates the unit ball, $B_n$, to order $k$ along $H_0$ and which has cross-sectional areas of $\pi$ on all hyperplanes through the origin and within $\Theta$ radians of the $x_1$-axis.

**Proof.** Let $f$ be an even non-zero function satisfying the conclusions of Proposition 2.1. Let $\epsilon_0 > 0$ be so small that $1 + \epsilon_0 f > 0$. By the choice of $\epsilon_0$, for any $\epsilon \in (0, \epsilon_0)$, $\rho_\epsilon(\omega) = \sqrt{1 + \epsilon f(\omega)}$ is the radial function of an origin-centered set, $K_\epsilon$ with smooth boundary. Since the unit sphere has strictly positive curvature, for sufficiently small $\epsilon \in (0, \epsilon_0)$, $K_\epsilon$ will be strictly convex; see the end of the proof of Theorem 3.1 of [G1] and also [O]. □

These counterexamples give the following strong non-uniqueness results.

**Proposition 2.3.** Let $A$ be any set in $S^{n-1}$ that omits an even open subset of $S^{n-1}$. Then, $f \in C_{even}^{\infty}(S^{n-1})$ is not determined by data $Rf(\omega)$ for $\omega \in A$. Furthermore, origin centered convex bodies are not determined by cross-sectional areas on subspaces perpendicular to vectors in $A$.

**Proof.** Any set $A$ that omits an even open subset of $S^{n-1}$ can be rotated so that the omitted subset contains the $x_1$-axis. Proposition 2.1 provides the non-uniqueness result for the Funk-Radon transform. Proposition 2.2 provides the non-uniqueness result for cross-sectional areas. □

3. Projections of convex bodies and the cosine transform.

Now, we explain how the projection areas of a convex body give the Cosine transform of the surface area measure of the body. We review the relation between the Cosine and Funk-Radon transform and note that the two are related by an invertible real analytic elliptic differential operator and hence
have the same microlocal support properties. We can then state analogs of
the cross-section theorems above for projection areas ("shadows").

We would like to formulate analogs of the results of Section 1 for pro-
jections or shadows instead of cross-sections. If $K$ is a star body in $\mathbb{R}^n$ let
$K|_{P(\omega)}$ denote its orthogonal projection into the hyperplane, $P(\omega)$, through
the origin which has normal vector $\omega$. The quantity $\text{Area}(K|_{P(\omega)})$ is called
the *brightness* of $K$ in the direction $\omega$. Elementary examples show that we
can only hope to recover the convex hull of $K$ from its brightness function
and, in keeping with tradition, we restrict the discussion to convex bod-
ies $K$. There is a partial duality relating cross-sections of star bodies
and projections of convex bodies.

With each smooth convex body $K$ we can associate a measure on $S^{n-1}$.
To define it, we first define the *Gauss map* to be the map $g_K: \text{bd} K \to S^{n-1}$
taking $x \in \text{bd} K$ to the outer unit normal to $\text{bd} K$ at $x$. If $\text{bd} K$ is $C^2$ and
strictly convex, then $g_K$ is continuous and bijective [G3, p. 24-5]. If $\text{bd} K$ is
$C^2$ and strictly positively curved everywhere then $g_K$ is a $C^1$ function with
$C^1$ inverse (the inverse is $C^1$ because strict positive curvature implies the
derivative of $g_K$ has only positive eigenvalues). If $\text{bd} K$ is $C^\infty$ and strictly
positively curved everywhere then $g_K$ is a $C^\infty$ function with $C^\infty$ inverse.

This allows us to define a measure, $S_K$, on $S^{n-1}$, the *infinitesimal surface
area measure* of $K$. Let $E \subset S^{n-1}$ be measurable and let $\lambda_{n-1}$ be the surface
area measure on $\text{bd} K$. Then the measure is defined by

\begin{equation}
S_K(E) = \lambda_{n-1}(g_K^{-1}(E)).
\end{equation}

In integrals, we will denote this measure by $dS_K$.

Note that the infinitesimal surface area measure of the unit ball is the
standard measure on the sphere because the Gauss map is the identity. If
$\text{bd} K$ is strictly positively curved and $C^2$, then calculation (3.7) below shows
that $dS_K$ is absolutely continuous with respect to Lebesgue measure on the
sphere. However, the infinitesimal surface measure can be defined even if
the Gauss map is not well-defined (e.g., if $K$ is a polytope). See [G3] for
details.

This measure is important because the projection area

\begin{equation}
\text{Area}(K|_{P(\omega)}) = \int_{S^{n-1}} |\omega \cdot \eta| dS_K(\eta)
\end{equation}

where $\omega \cdot \eta$ is the Euclidean inner product of these vectors.

If $K$ is origin-centered then $K$ is determined by its infinitesimal surface
area measure. This follows from the Alexandrov projection and uniqueness
theorems [G3, §3.3]. If $K$ is smoothly bounded and positively curved
throughout, then $dS_K$ is absolutely continuous with respect to Lebesgue
measure on $S^{n-1}$ and its Radon-Nikodym derivative is the $(n-1)^{st}$ elemen-
tary symmetric function of the principal radii of curvature of $K$ [S2, §2.5,
4.2. The projection areas of the convex body $K$ define a centered body $\Pi(K)$ called the projection body of $K$, by duality with intersection bodies above. We refer to [G3] for general background on these matters.

Equation (3.1) suggests an integral operator called the Cosine transform. Its definition is:

$$T : C(S^{n-1}) \rightarrow C_{\text{even}}(S^{n-1})$$

$$(3.3) \quad Tf(\omega) = \int_{S^{n-1}} |\omega \cdot \eta| f(\eta) \, d\eta,$$

where $d\eta$ is the rotation invariant normalized measure on the sphere. The cosine transform is a special Blaschke-Levy Representation, and many authors, including Alexandrov, Gardner, Goodey, Groemer, Koldobsky, Rubin, Schneider, and Weil have proven important properties including inversion formulas [Ko2] and inversion formulas using wavelets [Ru1] and relations to other transforms [Ru2]. See [G3, Ko2] for more information.

There is an identity linking the Funk-Radon and Cosine transforms [GW]:

$$\Delta + 2(n - 1) Tf(\omega) = 2Rf(\omega), \quad \forall \omega \in S^{n-1}. \tag{3.4}$$

Here $\Delta$ denotes the spherical Laplacian. Thus $R$ and $T$ are related by an invertible real-analytic-elliptic operator and so they have the same microlocal transformation properties. This means, in particular, that $Rf$ is smooth or real-analytic whenever $Tf$ is. More precisely, $Tf$ and $Rf$ have the same wavefront set in both the smooth [Hö] and real-analytic [Tr] categories. The relation (3.4) gives us analogous theorems to Theorem 1.2 and 1.3 for the Cosine Transform. For $\omega \in S^{n-1}$, recall that $P(\omega)$ is the hyperplane through the origin in $\mathbb{R}^n$ perpendicular to $\omega$ and $\omega^\perp = P(\omega) \cap S^{n-1}$.

**Theorem 3.1.** Let $K_1$ and $K_2$ be origin centered convex bodies in $\mathbb{R}^n$ with $C^2$ boundaries that are strictly positively curved. Let $A$ be an open connected set in $S^{n-1}$ such that $\mathbb{R}^n = \bigcup_{\omega \in A} P(\omega)$. Assume that for some $\eta_0 \in A$, the infinitesimal surface area measures of $K_1$ and $K_2$ agree to infinite order\(^1\) along the equator $\eta_0^\perp$, and assume the projection areas (3.2) of $K_1$ and $K_2$ agree for $\omega \in A$. Then $K_1 = K_2$.

The assumptions about $K_j$ in the theorem ensure that the Radon-Nikodym derivatives of the infinitesimal surface area measures with respect to Lebesgue measure exist.

**Proof.** For $j = 1, 2$, because $K_j$ is $C^2$ and strictly positively curved, its infinitesimal surface measure is absolutely continuous with respect to Lebesgue measure on the sphere (see (3.7)). Let $f_j$ be its Radon-Nikodym derivative with respect to Lebesgue measure; then $f_j$ is a continuous function [G3]. The identities (3.2) and (3.4) can be used to show that the Funk-Radon

\(^1\)That is, the Radon-Nikodym derivatives of $dS_{K_1}$ and $dS_{K_2}$ with respect to Lebesgue measure on the sphere are functions that agree to infinite order according to Definition 1.1.
transforms of the functions $f_1$ and $f_2$ are identical on $A$. Since $f_1$ and $f_2$
agree to infinite order on the great sphere $\eta_0^\perp$ and $\eta_0 \in A$, Theorem 1.1 can
be used to finish the proof. □

Our next theorem follows from Theorem 3.1, and it has a more geometric
starter condition.

**Theorem 3.2.** Let $K$ be an origin centered convex body in $\mathbb{R}^n$ with a $C^2$
boundary that is strictly positively curved. Let $A$ be an open connected set
in $S^{n-1}$ such that $\mathbb{R}^n = \bigcup_{(\omega \in A)} P(\omega)$. Assume that for some $\eta_0 \in A$, $K$
oscillates a disk, $D$, centered at the origin to infinite order along $P(\eta_0)$ and
assume the projection areas (3.2) of $K$ agree on $A$ with those of $D$. Then
$K = D$.

Schneider [S1] has proven important related theorems. Let $K_1$ be a cen-
tered, convex polytope. Assume that $K_2$ is a centered convex body and the
projection areas of $K_1$ and $K_2$ agree for all $\eta$ in an open set $A$ of arbitrarily
small measure (and containing the normal vectors to each face of $K_1$). He
proves that $K_1$ and $K_2$ are the same. His starter assumption is that $K_1$
is a polytope, but he does not assume $\text{bd} K_1$ and $\text{bd} K_2$ agree anywhere, and
his set $A$ is different from ours.

Schneider and Weil prove a related theorem in [SW] for odd dimensions.
Their starter assumption is that each body has at least one vertex and both
sets share a supporting plane, $\mathcal{P}$, that touches a vertex of each body. They
assume the projection areas of $K_1$ and $K_2$ are the same for normal vectors
in an equatorial belt perpendicular to the normal vectors to $\mathcal{P}$. Then, they
conclude $K_1 = K_2$. They also show that either if $n$ is even or if only one of
the sets has a vertex supporting plane, then the theorem is false. This very
intriguing theorem has a different flavor from ours since convex sets with
vertices do not have smooth boundaries, so our theorem does not apply.
Also, their set $A$ is different from ours since it does not include the vector
for which starter information is given (the normal to $\mathcal{P}$).

**Proof.** We can assume the disk $D$ is the unit disk in $\mathbb{R}^n$. We show that the
infinitesimal surface area measure of $K$ agrees on $\eta_0^\perp$ to infinite order with
that of $D$. Then, we use Theorem 3.1 to conclude that $K = D$.

To compare $dS_K$ to $d\eta = dS_D$, we use (3.2) to write the surface area on
$\text{bd} K$, $\lambda_{n-1}$, as a measure on $S^{n-1}$ using the radial map $\rho_K : S^{n-1} \to \mathbb{R}$. Recall that the Gauss map $g_K : \text{bd} K \to S^{n-1}$ takes points on $\text{bd} K$ to their
unit outer normals. Let $h : S^{n-1} \to \text{bd} K$ be defined by

$$h(\omega) = \rho_K(\omega) \omega.$$  

(3.5)

A straightforward exercise shows for $F \subset \text{bd} K$ that

$$\lambda_{n-1}(F) = \int_{h^{-1}(F)} \frac{\rho_K^{n-1}(\omega)}{|\omega \cdot g_K(h(\omega))|} d\omega.$$  

(3.6)
Another exercise using a change of variables (where $J$ is the Jacobian determinant) shows that if $E \subset S^{n-1}$ measurable, then

$$S_K(E) = \int_{\omega \in (g_K \circ h)^{-1}(E)} \frac{\rho_K^{n-1}(\omega)}{|\omega \cdot g_K(h(\omega))|} \, d\omega$$

(3.7)

$$= \int_{\eta \in E} \frac{\rho_K^{n-1}(\omega)}{|\omega \cdot g_K(h(\omega))|} \frac{1}{|J(g_K \circ h)(\omega)|} \, |\omega = (g_K \circ h)^{-1}(\eta)| \, d\eta.$$ 

It should be pointed out that $J(g_K \circ h)$ is non-zero since $K$ is strictly convex and $C^2$. Since $K$ osculates the unit disk along $P(\eta_0)$ to infinite order, this implies $\rho_K = 1$ to infinite order on the great sphere $\eta_0^\perp$ and that $g_K \circ h$ is the identity map to infinite order on $\eta_0^\perp$. This implies that the expression in brackets in (3.7) is equal to one to infinite order on $\eta_0^\perp$. But, this expression is just the Radon-Nikodym derivative of $S_K$. Therefore, the hypotheses of Theorem 3.1 hold and we can use that to prove our theorem. □

A similar result to Theorem 3.2 can be stated for arbitrary strictly convex sets $K_1$ and $K_2$ that agree to infinite order on a set that becomes a great sphere under the Gauss map.

**Theorem 3.3.** Let $K_j$ be an origin centered convex body in $\mathbb{R}^n$ with a $C^\infty$ boundary that is strictly positively curved for $j = 1, 2$. Let $A$ be an open connected set in $S^{n-1}$ so that $\mathbb{R}^n = \bigcup_{\omega \in A} P(\omega)$. Let $\eta_0 \in A$, and assume $K_1$ osculates $K_2$ to infinite order along the set $H = g_{K_1}^{-1}(\eta_0^\perp)$. Assume the projections (3.2) of $K_1$ and $K_2$ agree on $A$. Then $K_1 = K_2$.

Since $K_j$ is strictly convex with $C^\infty$ boundary, the set $H = g_{K_1}^{-1}(\eta_0^\perp)$ is smooth. The osculation condition in Theorem 3.3 is defined in terms of the radial functions of $K_1$ and $K_2$ in a similar way as in Definitions 1.1 and 1.2.

The proof is similar to the proof of Theorem 3.2. One observes that, if $K_1$ osculates $K_2$ to infinite order along $H$, then the radial maps $\rho_{K_j}$ agree to infinite order on $h_j^{-1}(H \cap \text{bd } K_j) = H \cap S^{n-1}$ (where $h_j : S^{n-1} \to \text{bd } K_j$ is the map defined by (3.5)) and the Gauss maps $g_{K_j}$ agree to infinite order on $H \cap \text{bd } K_j$. Then, one can use (3.7) to observe that the infinitesimal surface area measures agree to infinite order on $\eta_0^\perp$. Finally, one uses Theorem 3.1.

**References**


Received February 16, 1999. The first author was partially supported by NSF grant 9971828, and the second author was partially supported by NSF grants 9622947 and 9877155.
AN ENDPOINT ESTIMATE FOR SOME MAXIMAL OPERATORS ASSOCIATED TO SUBMANIFOLDS OF LOW CODIMENSION

Ya Ryong Heo

We show that the maximal operator

$$Mf(x) = \sup_{j \in \mathbb{Z}} \left| \int_{\mathbb{R}^d} f(x - 2^j y) \, d\mu(y) \right|$$

maps $H^1$ into $L^{1,\infty}$ under certain assumptions on the decay of $\hat{\mu}$ and the geometry of $\text{supp}(\mu)$.

1. Introduction and statement of results.

In this paper we consider the lacunary maximal operator $M$ defined by

(1) $$Mf(x) = \sup_{j \in \mathbb{Z}} \left| \int_{\mathbb{R}^d} f(x - 2^j y) \, d\mu(y) \right|.$$ 

Here $d \geq 1$ is an integer. When $\mu$ is a finite positive Borel measure on $\mathbb{R}^d$, it is proved in [DR] that if the Fourier transform of $\mu$ satisfies

$$|\hat{\mu}(\xi)| \leq c (1 + |\xi|)^{-\alpha}$$

for some $\alpha > 0$, then (1) is bounded on $L^p(\mathbb{R}^d)$ for $1 < p \leq \infty$. Also when $\alpha = \frac{d}{2}$, it is proved in [O] that (1) maps $H^1(\mathbb{R}^d)$ into $L^{1,\infty}(\mathbb{R}^d)$. Here $H^1$ denotes the usual real-variable Hardy space. On the other hand, Theorem 4 in [C2] states that if $\mu$ is the Lebesgue measure $\sigma_{d-1}$ on the unit sphere $\sum_{d-1}$ in $\mathbb{R}^d$ then (1) maps $H^1(\mathbb{R}^d)$ into $L^{1,\infty}(\mathbb{R}^d)$. The purpose of this paper is to prove a result which includes the results in [O] and Theorem 4 in [C2] as special cases and which also applies to maximal operators associated to some submanifolds of codimension greater than 1. The method of proof is an adaptation of the argument in [O], which is based on the basic approach in [C2].

For each bounded subset $A$ of $\mathbb{R}^d$ and $0 < \epsilon < 1$, define $N(A, \epsilon)$ as the smallest number of $\epsilon$-balls needed to cover $A$, i.e.,

$$N(A, \epsilon) = \min \left\{ m : A \subset \bigcup_{i=1}^m B(x_i, \epsilon) \text{ for some } x_i \in \mathbb{R}^d \right\}.$$ 

Now we state our main result.
Theorem 1. Suppose $\mu$ is a finite positive Borel measure on $\mathbb{R}^d$ with compact support such that for $0 < \epsilon < 1$

$$N(\text{supp}(\mu), \epsilon) \leq c \epsilon^{-n}, \quad |\hat{\mu}(\xi)| \leq c (1 + |\xi|)^{-\frac{n}{2}}$$

then (1) maps $H^1(\mathbb{R}^d)$ into $L^{1,\infty}(\mathbb{R}^d)$ when $0 < n \leq d$.

In particular if $n = d$, then we obtain the result of [O]. Moreover we have the following.

Corollary 2. Suppose $M \subset \mathbb{R}^d$ is a $C^1$ submanifold of dimension $n$ equipped with a finite positive Borel measure $\mu$ which has compact support. If the Fourier transform of $\mu$ satisfies the decay estimate

$$|\hat{\mu}(\xi)| \leq c (1 + |\xi|)^{-\frac{n}{2}}$$

then (1) maps $H^1(\mathbb{R}^d)$ into $L^{1,\infty}(\mathbb{R}^d)$ when $0 < n \leq d$.

Proof. Let $A$ be a bounded subset of $\mathbb{R}^n$ and $f : A \mapsto \mathbb{R}^d$ be a Lipschitz map. Then it is easy to show that

$$N(f(A), \epsilon) \leq c N(A, \epsilon) \leq c \epsilon^{-n}. \quad (3)$$

If $M$ is a $C^1$ submanifold of $\mathbb{R}^d$, then we can view $M$ locally as the graph of a vector-valued $C^1$ function defined on its tangent plane. Hence by (3) and compactness of $\text{supp}(\mu)$, we have $N(\text{supp}(\mu), \epsilon) \leq c \epsilon^{-n}$. By applying Theorem 1, we obtain the conclusion. \qed

In particular if $M$ is $\sum_{d-1}$ and $\mu$ is $\sigma_{d-1}$, then we obtain Theorem 4 in [C2]. Also, as was treated in [CDMM] and [CM], if $M$ is a smooth compact convex hypersurface of finite type in $\mathbb{R}^{1+n}$, with Gaussian curvature $\kappa$ and surface measure $\mu$, then the Fourier transform $\hat{\kappa^{1/2}} \mu(\xi)$ decays as $|\xi|^{-\frac{n}{2}}$ as $|\xi|$ goes to infinity. Hence Corollary 2 holds for $\kappa^{1/2} \mu$ when $n \geq 1$.

Our proof follows the methods of [C2] and [O]. What is different from [O] is the use of the geometry of $\text{supp}(\mu)$. We use the geometry of $\text{supp}(\mu)$ in proving Lemma 5. The use of geometry of $\text{supp}(\mu)$ allows us to put a weaker decay condition on $\hat{\mu}$. Littman [L] showed that, if $M \subset \mathbb{R}^{1+n}$ is a smooth submanifold of dimension $n$ and has at least $l$ nonzero principal curvatures everywhere on $\text{supp}(\mu)$, where $\mu$ is smooth and compactly supported, then

$$|\hat{\mu}(\xi)| \leq c (1 + |\xi|)^{-\frac{l}{2}}.$$ 

Hence when $l = n \geq 1$, Corollary 2 can be applied.

As was indicated in [C3], the proof of Littman’s theorem goes unchanged to establish the following. Suppose that $M \subset \mathbb{R}^d$ is a smooth manifold of dimension $n$, and $\mu$ is a smooth compactly supported measure on $M$. For fixed $b \in M$, we can view $M$ locally as a graph of a vector-valued function
ψ(x) defined on its tangent plane. Let \( N_b(M) \) be a collection of a unit vector normal to \( M \) at \( b \) then for each \( v \in N_b(M) \) the function \( \langle ψ(x), v \rangle \) has a critical point at \( x = b \). Suppose that for all \( b \in M \) in some neighborhood of \( \text{supp}(μ) \) and for all \( v \in N_b(M) \) we have

\[
\det D^2 \langle ψ(x), v \rangle |_{x=b} \neq 0.
\]

(4)

Then

\[
|\hat{μ}(ξ)| \leq c (1 + |ξ|)^{-\frac{n}{2}}.
\]

(5)

Hence Corollary 2 can be applied in this case also. The condition (4) is controlled by the second-order terms in the Taylor expansion of \( ψ \) at \( b \). We give some examples which satisfy (5).

Example 3.

(3.1) For \( n = 2m \) and \( d = n + 2 \), let \( x, y \in \mathbb{R}^m \) and \( M \) be the manifold described by \( (x, y; |x|^2 - |y|^2, x \cdot y) \), then a smooth measure \( μ \) supported in a sufficiently small neighborhood of the origin satisfies (5) when \( m \geq 1 \). So Corollary 2 holds for this \( μ \) when \( m \geq 1 \).

(3.2) For \( n = 4m \) and \( d = n + 2 \), let \( x, y, z, u \in \mathbb{R}^m \) and \( M \) be the manifold described by \( (x, y, z, u; x \cdot z + y \cdot u, x \cdot u - y \cdot z) \), then a smooth measure \( μ \) supported in a sufficiently small neighborhood of the origin satisfies (5) when \( m \geq 1 \). So Corollary 2 holds for this \( μ \) when \( m \geq 1 \).

(3.3) For \( n = 4m \) and \( d = n + 3 \), let \( x, y, z, u \in \mathbb{R}^m \) and \( M \) be the manifold described by \( (x, y, z, u; |x|^2 - |y|^2 - |z|^2 + |u|^2, x \cdot y - z \cdot u, x \cdot z + y \cdot u) \), then a smooth measure \( μ \) supported in a sufficiently small neighborhood of the origin satisfies (5) when \( m \geq 1 \). So Corollary 2 holds for this \( μ \) when \( m \geq 1 \).

2. Preliminaries.

Notation. If \( Q \) is a dyadic cube in \( \mathbb{R}^d \) with side-length \( 2^j \), we write \( σ(Q) = j \). For \( σ \in \mathbb{Z} \), \( \mathcal{R}_σ \) denotes the collection of dyadic cubes \( Q \in \mathbb{R}^d \) with \( σ(Q) = σ \). And for \( Q \in \mathcal{R}_σ \), \( Q^σ \) denotes \( Q + [-2^σ, 2^σ]^d \). \( |\cdot| \) denotes the Lebesgue measure.

The following Lemma is taken from [O] (see Lemma 1).

Lemma 4. Suppose \( \alpha > 0 \) is given, and given any finite collection of dyadic cubes \( \{Q\}_{Q \in \mathcal{C}} \) in \( \mathbb{R}^d \), and corresponding collection of positive numbers \( \{λ_Q\}_{Q \in \mathcal{C}} \) there exists a finite collection of pairwise disjoint dyadic cubes \( \{S\}_{S \in \mathcal{S}} \) such that each \( Q \in \mathcal{C} \) is contained for some \( S \in \mathcal{S} \) and

\[
\sum_{Q \subset S} λ_Q \leq 3^d α |S|
\]

(4.1)

\[
\sum_{S \in \mathcal{S}} |S| \leq \frac{1}{α} \sum λ_Q
\]

(4.2)

\[
\left\| \sum_{Q: \text{ not contained in any } S} λ_Q |Q|^{-1} \chi_Q \right\|_{L^∞} \leq α.
\]

(4.3)
Lemma 5 (cf. [C2, Lemma 5.1]). Suppose given the following: 0 < n ≤ d, a Borel measure μ defined on a compact subset of $\mathbb{R}^d$ with $N(\operatorname{supp}(\mu), \epsilon) \leq c \epsilon^{-n}$ for $0 < \epsilon < 1$, some $\alpha > 0$, a finite collection $S$ of pairwise disjoint dyadic cubes $S \subset \mathbb{R}^d$, a finite collection $\mathcal{C}$ of dyadic cubes $Q \subset \mathbb{R}^d$ such that each $Q \in \mathcal{C}$ is contained in some $S = S(Q) \in S$ and for each $Q \in \mathcal{C}$ a positive number $\lambda_Q$ is assigned. Then there exist a function $K : \mathcal{C} \rightarrow \mathbb{Z}$ and a measurable set $E$ such that

(5.1) $|E| \leq c \left( \frac{1}{n} \sum \lambda_Q + \sum |S| \right)$

(5.2) $\{ Q + 2^j \operatorname{supp}(\mu) \} \subset E$ if $j < K(Q)$ and $Q \in \mathcal{C}$

(5.3) $\sigma(S(Q)) < K(Q)$ (Q \in \mathcal{C})$

(5.4) For each $\tau, \sigma \in \mathbb{Z}$ with $\sigma \leq \tau$, and any $q \in \mathcal{R}_\sigma$

$$\sum_{Q \subset q, K(Q) \leq \tau} \lambda_Q \leq 2^n \alpha 2^{(d-n)\sigma+n\tau}.$$ 

Proof. The proof is a stopping-time argument controlled by two parameters $\tau$ and $\sigma$ as in the proof of Lemma 5.1 in [C2]. Let $m = \min \{ \sigma(Q) : Q \in \mathcal{C} \}$. Select an integer $\tau_0$ such that

$$\tau_0 > \max \{ \sigma(Q) : Q \in \mathcal{C} \}, \quad \sum_{Q \in \mathcal{C}} \lambda_Q < \alpha 2^{(d-n)m+n\tau_0}.$$ 

For each fixed $\tau \in \mathbb{Z}$ with $\tau \leq \tau_0$, we define a sequence of functions $\Lambda_{\tau, \sigma} : \mathcal{R}_\sigma \mapsto \mathbb{R}$ by a descending induction on $\sigma \in \mathbb{Z}$ with $\sigma \leq \tau$. And proceed with the same construction by a descending induction on $\tau$. At each step, we divide $\mathcal{C}$ into disjoint subcollections $\mathcal{C}_1$ and $\mathcal{C}_2$ which will increase as we proceed. Let $\mathcal{C}_1, \mathcal{C}_2 \subset \mathcal{C}$ and $\tau \in \mathbb{Z}$ be fixed for the moment, and we define [Inner Loop] as

[Inner Loop] Define $\Lambda_{\tau, \sigma} : \mathcal{R}_\sigma \mapsto \mathbb{R}$ with $\sigma \leq \tau$. For each $q \in \mathcal{R}_\sigma$ define

$$\Lambda_{\tau, \sigma}(q) = \sum_{Q \subset q, Q \notin \mathcal{C}_1 \cup \mathcal{C}_2} \lambda_Q.$$ 

First, begin with $\sigma = \tau$. If $\Lambda_{\tau, \sigma}(q) > \alpha 2^{(d-n)\sigma+n\tau}$ then we say that “$q$ is selected at step $(\tau, \sigma)$” and put into $\mathcal{C}_1$ every $Q$ such that $Q \subset q$ and for such a $Q$ define $K(Q) = 1 + \tau$. Next replace $\sigma$ by $\sigma - 1$ and repeat the process. Repeat until $\sigma < m$. Actually this part of process terminates once $\sigma$ is smaller than $m$. Finally, put into $\mathcal{C}_2$ every $Q \in \mathcal{C} \setminus \mathcal{C}_1$ such that $\sigma(Q) \geq \tau$ and for such a $Q$ define $K(Q) = 1 + \sigma(S(Q))$. Actually every $Q \in \mathcal{C} \setminus \mathcal{C}_1 \cup \mathcal{C}_2$ satisfies $\sigma(Q) \leq \tau - 1$.

Perform [Inner Loop] with $\mathcal{C}_1 = \mathcal{C}_2 = \emptyset$ and $\tau = \tau_0$. Next replace $\tau$ by $\tau - 1$ and repeat [Inner Loop]. Repeat until $\tau = m - 1$. After this process, we obtain $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$, and clearly all selected $q$ are disjoint, and $K$
is well-defined. Note that there is the usual stopping-time condition

\begin{equation}
\Lambda_{\tau,\sigma}(q) \leq 2^n \alpha 2^{(d-n)\sigma+n\tau}
\end{equation}

which holds for all \( q \in \mathbb{R}_\sigma \) when \( \sigma \leq \tau \leq \tau_0 \). This is because, if \( \tau = \tau_0 \) then the condition is clear from the initial condition on \( \tau_0 \). And when \( \sigma \leq \tau < \tau_0 \), suppose this fails. Then \( \Lambda_{\tau+1,\sigma}(q) \geq \Lambda_{\tau,\sigma}(q) > \alpha 2^{(d-n)\sigma+n(\tau+1)} \). This means \( q \) is selected at step \( (\tau+1,\sigma) \), hence \( \Lambda_{\tau,\sigma}(q) = 0 \) and we have contradiction.

Next we show (5.4), which says that for each \( q \in \mathbb{R}_\sigma \) with \( \sigma \leq \tau \leq \tau_0 \)

\begin{equation}
\sum_{Q \subset q} \lambda_Q \leq 2^n \alpha 2^{(d-n)\sigma+n\tau}.
\end{equation}

When \( \tau \geq \tau_0 \), then the condition is clear from the initial condition of \( \tau_0 \). When \( \tau \leq \tau_0 \), then we note the fact that for each \( q \in \mathbb{R}_\sigma \) with \( \sigma \leq \tau \leq \tau_0 \)

\begin{equation}
\Lambda_{\tau,\sigma}(q) = \sum_{Q \subset q, K(Q) \leq \tau} \lambda_Q \geq \sum_{Q \subset q, K(Q) \leq \tau} \lambda_Q.
\end{equation}

Combining (6) and (7), we have (5.4) when \( \sigma \leq \tau \leq \tau_0 \). (7) will follow from the definition

\begin{equation}
\Lambda_{\tau,\sigma}(q) = \sum_{Q \subset q} \lambda_Q
\end{equation}

and the fact that if \( Q \in C_1 \cup C_2 \) at the beginning of step \( (\tau,\sigma) \) then \( K(Q) > \tau \). This is because, if \( Q \in C_1 \) then \( K(Q) \geq 1 + \tau > \tau \), and if \( Q \in C_2 \) then \( K(Q) = 1 + \sigma(S(Q)) \geq 1 + (1 + \tau) > \tau \). Hence we have (5.4).

Next, we construct an exceptional set \( E \). If \( q \) is selected at step \( (\tau,\sigma) \), then we define \( \tau(q) = \tau \) and

\[
T(q) = \bigcup_{j \leq \tau(q)+1} \{ q + 2^j \text{supp}(\mu) \}
\]

\[
E = E_1 \cup E_2, \quad E_1 = \bigcup_{S \in S} S^*, \quad E_2 = \bigcup_{q \text{selected}} T(q).
\]

Thus we have

\[
|E_1| \leq c \sum |S|
\]

and

\[
T(q) = \bigcup_{j \leq \tau(q)+1} \{ q + 2^j \text{supp}(\mu) \}
\]

\[
= \bigcup_{j \leq \sigma(q)} \{ q + 2^j \text{supp}(\mu) \} \bigcup_{\sigma(q) < j \leq \tau(q)+1} \{ q + 2^j \text{supp}(\mu) \}.
\]

Because \( \text{supp}(\mu) \) is compact, if we regard \( q^* \) as a proper expansion of \( q \) then \( \bigcup_{j < \sigma(q)} \{ q + 2^j \text{supp}(\mu) \} \subset q^* \). And for \( j > \sigma(q) \), if \( x_0 \) is the center
of \( q \), then by using translation invariance and dilation property of Lebesque measure, we have

\[
\left| \{ q + 2^j \text{supp}(\mu) \} \right| \leq \left| \left\{ B(x_0, 2^\sigma(q)) + 2^j \text{supp}(\mu) \right\} \right|
= \left| \left\{ B(0, 2^\sigma(q)) + 2^j \text{supp}(\mu) \right\} \right|
= 2^d \left[ B(0, 2^\sigma(q) - j) + \text{supp}(\mu) \right]
\leq c 2^d 2^{d(\sigma(q) - j)} N(\text{supp}(\mu), 2^\sigma(q) - j)
\leq c 2^{(d-n)\sigma(q) + nj}.
\]

Hence

\[
|T(q)| \leq c \left( |q| + \sum_{\sigma(q) < j \leq \tau(q) + 1} 2^{(d-n)\sigma(q) + nj} \right) \leq c 2^{(d-n)\sigma(q) + n\tau(q)}
\]

and we have

\[
|E_2| \leq \sum_{q \in \text{selected}} |T(q)|
\leq c \sum_{q \in \text{selected}} 2^{(d-n)\sigma(q) + n\tau(q)}
\leq \frac{c}{\alpha} \sum_{q \in \text{selected}} \Lambda_{\tau,\sigma}(q)
\leq \frac{c}{\alpha} \sum \lambda_Q.
\]

So we obtain (5.1). For (5.2), observe that if \( Q \in C_1 \) then \( Q \) belongs to some selected \( q \), hence

\[
\bigcup_{j < K(Q)} \{ Q + 2^j \text{supp}(\mu) \} \subset \bigcup_{j \leq K(Q) = \tau(q) + 1} \{ q + 2^j \text{supp}(\mu) \} = T(q) \subset E_2
\]

and if \( Q \in C_2 \) then \( Q \) belongs to some \( S = S(Q) \in S \), hence

\[
\bigcup_{j < K(Q) = 1 + \sigma(S(Q))} \{ Q + 2^j \text{supp}(\mu) \} \subset S^* \subset E_1
\]

if we regard \( S^* \) as a proper expansion of \( S \). For (5.3), we replace \( K \) by \( K' \) and define

\[
K(Q) = \max \left\{ K'(Q), 1 + \sigma(S(Q)) \right\}.
\]

Then (5.1) and (5.3) are satisfied. We must check (5.2) and (5.4). For (5.2), if \( K(Q) = K'(Q) \) then there is no problem. If \( K(Q) = 1 + \sigma(S(Q)) > j \)
then the argument is the same as above. For (5.4)

$$\sum_{Q \subseteq q; K(Q) \leq \tau} \lambda_Q \leq \sum_{Q \subseteq q; K'(Q) \leq \tau} \lambda_Q$$

and Lemma 5 follows.

□

3. Proof of Theorem 1.

Let \( f \in H^1(\mathbb{R}^d) \) have the form of a finite sum

$$f = \sum \lambda_Q a_Q$$

where \( \lambda_Q > 0 \) and \( a_Q \), supported in \( Q \), satifies

$$\|a_Q\|_{L^\infty} \leq \frac{1}{|Q|}, \quad \int a_Q = 0.$$

As was pointed out in [C2], a device of Garnett and Jones involving auxiliary dyadic grids allows us to assume that each \( Q \) is dyadic. For \( \alpha > 0 \), it is enough to show

$$|\{x : Mf(x) > 2\alpha\}| \leq \frac{c}{\alpha} \sum \lambda_Q. \quad (8)$$

Let \( S \) be as in Lemma 4 and define

$$b = \sum_{S \in \mathcal{S}} \sum_{Q \subseteq S} \lambda_Q a_Q, \quad g = f - b.$$

Then \( \|g\|_{L^\infty} \leq \alpha \) from (4.3) and so \( |Mg| \leq \alpha \) (by assuming \( \mu \) has mass 1). Thus (8) will follow from

$$|\{x : Mb(x) > \alpha\}| \leq \frac{c}{\alpha} \sum \lambda_Q. \quad (9)$$

Let \( \mathcal{S} \) be as above and \( \mathcal{C} \) be the collection of \( Q \)'s appearing in the definition of \( b \). With \( K \) and \( E \) as in Lemma 5, it is enough to prove

$$\|Mb\|_{L^2(\mathbb{R}^d \setminus E)}^2 \leq c\alpha \sum \lambda_Q. \quad (9)$$

Let \( \mu_j \) be the dilate of \( \mu \) defined by

$$\langle \phi, \mu_j \rangle = \int_{\mathbb{R}^d} \phi(2^j x) \, d\mu(x)$$

then

$$Mb(x) = \sup_{j \in \mathbb{Z}} |b \ast \mu_j(x)|.$$
If \( Q \in C \), then by (5.3) \( a_Q * \mu_j \) is supported in \( E \) unless \( j \geq K(Q) \). Thus for \( x \notin E \), we have

\[
|M b(x)|^2 \leq \sum_j |b * \mu_j(x)|^2
\]

\[
= \sum_j \left| \left( \sum_{K(Q) \leq j} \lambda_Q a_Q \right) * \mu_j(x) \right|^2
\]

\[
= \sum_j \sum_{j-s}^\infty \left( \sum_{K(Q)=j-s} \lambda_Q a_Q \right) * \mu_j(x) \right|^2.
\]

So for \( x \notin E \), by Minkowski’s inequality

\[
|M b(x)| \leq \sum_{s=0}^\infty \left[ \sum_j \left| \left( \sum_{K(Q)=j-s} \lambda_Q a_Q \right) * \mu_j(x) \right| \right]^{2\frac{1}{2}}.
\]

Now (9) will follow from

\[
\left\| \left[ \sum_j \left( \sum_{K(Q)=j-s} \lambda_Q a_Q \right) * \mu_j \right]^{2\frac{1}{2}} \right\|_{L^2} \leq c(s + 3) \alpha 2^{-\epsilon s} \sum \lambda_Q
\]

where \( \epsilon = \min(1, n) \). And so from

\[
(10) \left\| \left( \sum_{K(Q)=j-s} \lambda_Q a_Q \right) * \mu_j \right\|_{L^2}^2 \leq c(s + 3) 2^{-\epsilon s} \sum_{K(Q)=j-s} \lambda_Q.
\]

By scaling we may take \( j = 0 \). And (10) will follow from

\[
(11) \left\| \left( \sum_{K(Q)=-s} \lambda_Q a_Q \right) * \mu \right\|_{L^2}^2 \leq c(s + 3) 2^{-\epsilon s} \sum_{K(Q)=-s} \lambda_Q.
\]

Next as in Lemma 3 in [O], for each positive integer \( N \), we define a sequence of functions \( h_N \) and \( L_N \). First we define \( h_N \) by

\[
\hat{h}_N(\xi) = \frac{\chi_{\{\xi\leq N(\xi)\}}}{(1 + |\xi|)^n}.
\]

Choose a radial function \( \rho \in C_c^\infty(\mathbb{R}^d) \) such that

\[
\int \rho = 1, \quad \text{supp}(\rho) \subset [-1, 1]^d, \quad \hat{\rho} \geq 0.
\]
Now let $L_N = \rho h_N$ and
\[ \hat{L}(\xi) = \lim_{N \to \infty} \hat{L}_N(\xi) = \int \frac{\hat{\rho}(y)dy}{(1 + |\xi - y|)^n}. \]

**Lemma 6.** We have the following:

(6.1) $\text{supp}(L_N) \subset [-1, 1]^d$
(6.2) $\hat{L}_N(\xi) \geq \frac{c}{(1 + |\xi|)^n}$ if $|\xi| \leq N - 1$
(6.3) For each $\beta$, we have
\[ \left| \partial_\xi^\beta \hat{L}(\xi) \right| \leq \frac{A_\beta}{(1 + |\xi|)^{n + |eta|}}. \]

**Proof.** It is easy to check (6.1), (6.2). For (6.3), first we assume $d \geq 2$, then we have
\[ \left| \partial_\xi^\beta \hat{L}(\xi) \right| = \left| \int \frac{\hat{\rho}(y)\partial_\xi^\beta (1 + |\xi - y|)^n}{(1 + |\xi - y|)^n} dy \right| \leq c \int \frac{\hat{\rho}(y)dy}{(1 + |\xi - y|)^{n + |eta|}} \leq \frac{c}{(1 + |\xi|)^{n + |eta|}}. \]

When $d = 1$, we use
\[ \hat{L}(\xi) = \int_\xi^\infty \frac{\hat{\rho}(y)dy}{(1 + y - \xi)^n} + \int_{-\infty}^\xi \frac{\hat{\rho}(y)dy}{(1 + \xi - y)^n}, \]
and do similarly as before. \qed

Next, let $\phi_N$ be the inverse Fourier transform of $(\hat{L}_N)^{\frac{1}{2}}$, then $L_N = \phi_N * \hat{\phi}_N$. And we have
\[ |\hat{\phi}_N(\xi)|^2 \geq \frac{c}{(1 + |\xi|)^n} \quad \text{when} \quad |\xi| \leq N - 1. \]

Therefore, returning to (11) we have
\[
\left\| \left( \sum_{K(Q) = -s} \lambda Q a_Q \right) * \mu \right\|^2_{L^2} = c \int \left\| \left( \sum_{K(Q) = -s} \lambda Q \hat{a}_Q \right)(\xi) \right\|^2 |\hat{\mu}(\xi)|^2 d\xi
\leq c \int \left\| \left( \sum_{K(Q) = -s} \lambda Q \hat{a}_Q \right)(\xi) \right\|^2 \liminf_{N \to \infty} |\hat{\phi}_N(\xi)|^2 d\xi
\leq c \liminf_{N \to \infty} \int \left\| \left( \sum_{K(Q) = -s} \lambda Q \hat{a}_Q \right)(\xi) \right\|^2 |\hat{\phi}_N(\xi)|^2 d\xi
\leq c \liminf_{N \to \infty} \left\| \left( \sum_{K(Q) = -s} \lambda Q a_Q \right) * \phi_N \right\|^2_{L^2}.\]
So (11) will follow from

\[
\liminf_{N \to \infty} \left\| \left( \sum_{K(Q) = -s} \lambda_Q a_Q \right) * \phi_N \right\|^2_{L^2} \\
\leq c \alpha (s + 3) 2^{-\epsilon s} \sum_{K(Q) = -s} \lambda_Q.
\]

Because \( \text{supp}(L_N) \subset [-1, 1]^d \), and for each \( Q, Q' \in \mathcal{C} \) such that \( K(Q) = K(Q') = -s \), we have \( \sigma(Q), \sigma(Q') \leq K(Q) = K(Q') = -s \), hence \( |\langle a_{Q'} * L_N, a_Q \rangle| = 0 \) when \( \text{dist}(Q, Q') > 4 \). So we have

\[
\liminf_{N \to \infty} \left\| \left( \sum_{K(Q) = -s} \lambda_Q a_Q \right) * \phi_N \right\|^2_{L^2} \\
\leq 2 \liminf_{N \to \infty} \sum_{Q, Q', \sigma(Q') \geq \sigma(Q) \text{ dist}(Q, Q') \leq 4} \lambda_Q \lambda_Q' \left| \langle a_{Q'} * L_N, a_Q \rangle \right|
\]

\[
\leq 2 \liminf_{N \to \infty} \sum_{Q, Q', \sigma(Q') \geq \sigma(Q) \text{ dist}(Q, Q') \leq 4} \lambda_Q \lambda_Q' \left| \langle \hat{a}_{Q'} \hat{L}, \hat{a}_Q \rangle \right|
\]

\[
\leq 2 \sum_{Q'} \sum_{Q \subset Q'' \text{ dist}(Q, Q') \leq 4} \lambda_Q \lambda_Q' \left| \langle \hat{a}_{Q'} \hat{L}, \hat{a}_Q \rangle \right|
\]

\[
+ 2 \sum_{Q'} \sum_{Q \cap Q'' \neq \emptyset \text{ dist}(Q, Q') \leq 4} \lambda_Q \lambda_Q' \left| \langle \hat{a}_{Q'} \hat{L}, \hat{a}_Q \rangle \right|
\]

\[
= I + II.
\]

**Lemma 7.** We have the following:

\[
(7.1) \left| \langle \hat{a}_{Q'} \hat{L}, \hat{a}_Q \rangle \right| \leq c 2^{-(d-n)\sigma(Q')}
\]

\[
(7.2) \left| \langle \hat{a}_{Q'} \hat{L}, \hat{a}_Q \rangle \right| \leq c \frac{2^{\sigma(Q)}}{(\text{dist}(Q, Q'))^{d-n+1}} \text{ when } Q \cap Q'' = \emptyset.
\]

**Proof.** For (7.1), we consider as two cases; \( d = n \) and \( d > n \). When \( d = n \), we use the easy estimates,

\[
|\hat{a}_Q(\xi)| \leq c \min(1, |\xi| 2^{\sigma(Q)}), \quad ||\hat{a}_Q||^2_{L^2} \leq c 2^{-d\sigma(Q)}.
\]
Hence we have
\[
\left| \langle \mathring{a}_Q \mathring{L}, \mathring{a}_Q \rangle \right| \leq \left\| \mathring{a}_Q \right\|_{L^\infty} \int \frac{\left| \mathring{a}_{Q'}(\xi) \right|}{(1 + |\xi|)^d} d\xi \\
\leq c \left( \int_{|\xi| < 2^{-\sigma(Q')}} \frac{|\xi|^{2\sigma(Q')}}{(1 + |\xi|)^d} d\xi + \left\| \mathring{a}_{Q'} \right\|_{L^2} \left[ \int_{|\xi| \geq 2^{-\sigma(Q')}} (1 + |\xi|)^{-2d} d\xi \right]^{1/2} \right) \\
\leq c.
\]

When \( d > n \), choose \( \eta \in C_c(\mathbb{R}^d) \) such that \( \eta(\xi) = 1 \) for \( |\xi| \leq 1 \), and \( \eta(\xi) = 0 \) for \( |\xi| \geq 2 \). Define another function \( \delta \) by \( \delta(\xi) = \eta(\xi) - \eta(2\xi) \). Then we have
\[
1 = \eta(\xi) + \sum_{j=1}^{\infty} \delta(2^{-j}\xi), \quad \text{for all } \xi,
\]
and
\[
\mathring{L}(\xi) = \eta(\xi)\mathring{L}(\xi) + \sum_{j=1}^{\infty} \mathring{L}(\xi)\delta(2^{-j}\xi) = m_0(\xi) + \sum_{j=1}^{\infty} m_j(\xi).
\]

We set
\[
K_j(x) = \int e^{2\pi i x \cdot \xi} m_j(\xi) d\xi.
\]

Observe that
\[
\left| (-2\pi i x)^\gamma \partial_x^\beta K_j(x) \right| = \left| \int \partial_\xi^\gamma \left[ (2\pi i \xi)^\beta m_j(\xi) \right] e^{2\pi i x \cdot \xi} d\xi \right|.
\]

By (6.3) and support condition of the integrand, we can show
\[
\left| x^\gamma \partial_x^\beta K_j(x) \right| \leq A_{\gamma,\beta} 2^{j(d-n+|\beta|-|\gamma|)}.
\]

Hence, for each positive integer \( M \), we have
\[
\left| \partial_x^\beta K_j(x) \right| \leq A_{M,\beta} |x|^{-M} 2^{j(d-n+|\beta|-M)},
\]
and so
\[
\sum_{j=0}^{\infty} \left| \partial_x^\beta K_j(x) \right| = \sum_{2^j \leq |x|^{-1}} + \sum_{2^j > |x|^{-1}}.
\]

First with \( M = 0 \), we have
\[
\sum_{2^j \leq |x|^{-1}} \left| \partial_x^\beta K_j(x) \right| \leq A_{\beta} \sum_{2^j \leq |x|^{-1}} 2^{j(d-n+|\beta|)} \leq A'_{\beta} |x|^{-d+n-|\beta|}.
\]
Second with \( M > d - n + |\beta| \), we have

\[
\sum_{2^j > |x|^{-1}} \left| \partial_x^j K_j(x) \right| \leq A_\beta \sum_{2^j > |x|^{-1}} |x|^{-M \cdot 2^j (d - n + |\beta| - M)} \\
\leq A'_\beta |x|^{-(d + n - |\beta|)}.
\]

Hence we have

\[
\sum_{j=0}^{\infty} \left| \partial_x^j K_j(x) \right| \leq A'_\beta |x|^{-(d + n - |\beta|)}.
\]

(14)

Returning to (7.1), by Lebesgue Dominated Convergence Theorem, we have

\[
\left| \langle \hat{a}_Q \hat{L}, \hat{a}_{Q'} \rangle \right| = \left| \sum_{j=0}^{\infty} \langle \hat{a}_{Q'} m_j, \hat{a}_Q \rangle \right| \\
= \left| \sum_{j=0}^{\infty} \langle a_{Q'} K_j, a_Q \rangle \right| \\
\leq \left| \langle |a_{Q'}| \sum_{j=0}^{\infty} |K_j|, |a_Q| \rangle \right| \\
\leq \|a_{Q'}\|_{L^1} \sup_{x \in Q} |a_{Q'}| \sum_{j=0}^{\infty} |K_j(x)| \\
\leq c \|a_{Q'}\|_{L^\infty} \sup_{x \in Q} \int_{Q'} \sum_{j=0}^{\infty} |K_j(x-y)| dy,
\]

and by (14), we have

\[
\sup_{x \in Q} \int_{Q'} \sum_{j=0}^{\infty} |K_j(x-y)| dy \leq c \sup_{x \in Q} \int_{Q'} |x-y|^{-d+n} dy \leq c 2^{n \sigma(Q')}
\]

when \( d > n \). Hence when \( d > n \), we have

\[
\left| \langle \hat{a}_{Q'} \hat{L}, \hat{a}_Q \rangle \right| \leq c 2^{-(d-n)\sigma(Q')},
\]
AN ENDPOINT ESTIMATE FOR SOME MAXIMAL OPERATORS 335

and obtain (7.1). For (7.2), let \( \bar{x} \) be the center of \( Q \), then

\[
\left| \left\langle \hat{a}_{Q'} L, \hat{a}_Q \right\rangle \right| = \left| \sum_{j=0}^{\infty} \left\langle a_{Q'} * K_j, a_Q \right\rangle \right|
\]

\[
= \left| \sum_{j=0}^{\infty} \iint a_{Q'}(y) (K_j(x - y) - K_j(\bar{x} - y)) a_Q(x) \, dxdy \right|
\]

\[
\leq \iint |a_{Q'}(y)| |a_Q(x)| \sum_{j=0}^{\infty} \left| (K_j(x - y) - K_j(\bar{x} - y)) \right| \, dxdy
\]

\[
\leq \iint |a_{Q'}(y)| |a_Q(x)| \sum_{j=0}^{\infty} |x - \bar{x}| \| \nabla K_j(\bar{x} - y) \| \, dxdy,
\]

where \( \bar{x}_j \) lies in the line connecting \( \bar{x} \) and \( x \). By (13), for each positive integer \( M \), we have

\[
\left| \nabla K_j(\bar{x}_j - y) \right| \leq A_M |\bar{x}_j - y|^{-M} 2^j (d - n + 1 - M)
\]

\[
\leq A'_M \text{dist}(Q, Q')^{-M} 2^j (d - n + 1 - M),
\]

when \( Q \cap Q'^* = \emptyset \). Hence, by the same method as in (14), we have

\[
\sum_{j=0}^{\infty} \left| \nabla K_j(\bar{x}_j - y) \right| \leq c \text{ (dist}(Q, Q'))^{-d + n - 1} \text{ when } Q \cap Q'^* = \emptyset.
\]

And so we have

\[
\left| \left\langle \hat{a}_{Q'} L, \hat{a}_Q \right\rangle \right| \leq c \frac{2^{\sigma(Q)}}{\text{ (dist}(Q, Q'))^{d - n + 1}} \iint |a_{Q'}(y)| |a_Q(x)| \, dxdy
\]

\[
\leq c \frac{2^{\sigma(Q)}}{\text{ (dist}(Q, Q'))^{d - n + 1}}
\]

when \( Q \cap Q'^* = \emptyset \). \( \square \)

\[ \bullet \] Estimation of part I:
By (5.4) we have \( \sum_{Q \subset Q^*} \lambda_Q \leq c \alpha 2^{(d-n)\sigma(Q') - ns} \) and use (7.1). So we have

\[
I \leq c \sum_{Q'} \sum_{Q \subset Q^*} \lambda_Q \lambda_{Q'} 2^{-(d-n)\sigma(Q')}
\]

\[
\leq c \left( \sum_{Q'} \lambda_{Q'} 2^{-(d-n)\sigma(Q')} \right) \left( \alpha 2^{(d-n)\sigma(Q') - ns} \right)
\]

\[
\leq c 2^{-ns} \alpha \sum_{K(Q)=-s} \lambda_Q.
\]

\[\bullet\] Estimation of part II:

If \( Q \cap Q^* = \emptyset \), then by (7.2) and \( \sigma(Q) \leq \sigma(Q') \), we have

\[
II \leq c \sum_{Q'} \sum_{Q \supset Q^* = \emptyset} \lambda_Q \lambda_{Q'} \frac{2^{\sigma(Q')}}{\text{dist}(Q, Q')(d-n)+1}
\]

\[
\leq c \left( \sum_{Q'} 2^{\sigma(Q')} \lambda_{Q'} \right) \left( \sum_{Q \supset Q^* = \emptyset} \frac{\lambda_Q}{\text{dist}(Q, Q')(d-n)+1} \right)
\]

\[
\leq c \left( \sum_{Q'} 2^{\sigma(Q')} \lambda_{Q'} \right) \left( \sum_{Q, \text{ dist}(Q, Q') \sim 2^{m+\sigma(Q')}} + \sum_{Q, \text{ dist}(Q, Q') \sim 2^{m+\sigma(Q')}} \right)
\]

\[
\leq c \left( \sum_{Q'} 2^{\sigma(Q')} \lambda_{Q'} \right) (II_1 + II_2).
\]

For each positive integer \( m \), consider the contribution of all \( \lambda_Q \) over all \( Q \) disjoint from \( Q^* \) with \( \sigma(Q) \leq \sigma(Q') \). So we have \( \text{dist}(Q, Q') \sim 2^{m+\sigma(Q')} \). All such \( Q \) are contained in the union of a fixed number of elements of \( \mathbb{R}_{m+\sigma(Q')} \). Hence when \( m + \sigma(Q') \leq -s + 2 \), we can use (5.4) to obtain

\[
II_1 = \sum_{Q, \text{ dist}(Q, Q') \sim 2^{m+\sigma(Q')}} \frac{\lambda_Q}{\text{dist}(Q, Q')(d-n)+1}
\]

\[
\leq c \sum_{m \geq 0} \alpha 2^{-(d-n+1)(m+\sigma(Q'))} 2^{(d-n)(m+\sigma(Q'))-ns}
\]

\[
\leq c \alpha 2^{-\sigma(Q')} 2^{-ns}.
\]
Next, consider all $Q$ with $\text{dist}(Q, Q') \sim 2^{m+\sigma(Q')}$ and $m + \sigma(Q') \geq -s + 3$. Recall that each $Q \in C$ is contained in $S(Q)$ for some $S(Q) \in \mathcal{S}$. Since $K(Q) = -s$ and $K(Q') > \sigma(S(Q))$, we obtain $\text{dist}(S(Q), Q') \geq 2^{-s}$. Also, by (4.1), we have $\sum_{Q \subseteq S} \lambda_Q \leq c\alpha|S|$ for every $S \in \mathcal{S}$, hence we obtain

$$III_2 = \sum_{Q : \text{dist}(Q, Q') \sim 2^{m+\sigma(Q')}} \frac{\lambda_Q}{\text{dist}(Q, Q')^{(d-n)+1}}$$

\[
\leq c \sum_{Q} \frac{\lambda_Q}{\text{dist}(S(Q), Q')^{(d-n)+1}}
\leq c\alpha \sum_{Q} \frac{|S|}{\text{dist}(S, Q')^{(d-n)+1}}
\leq c\alpha \int_{2^{-s} \leq |y| \leq 1} |y|^{-(d-n)+1} dy
\leq c\alpha (s2^{(1-n)s} + 1).
\]

Finally, since $\sigma(Q') < K(Q') = -s$, we obtain

$$II \leq c \sum_{Q'} 2^{\sigma(Q')} \left( \alpha 2^{-\sigma(Q')} 2^{-ns} + \alpha (s2^{(1-n)s} + 1) \right) \lambda_{Q'}$$

\[
\leq c(s + 3)\alpha 2^{-\epsilon s} \sum_{K(Q) = -s} \lambda_Q
\]

where $\epsilon = \min(n, 1)$. This completes the proof of (12) and Theorem 1.

**Acknowledgements.** The results of this paper were obtained during my Ph.D. studies at Pohang University of Science and Technology and will be a part of my thesis. I would like to express deep gratitude to my advisor Jong-Guk Bak for guidance and support.

**References**


Received January 10, 2000.

DEPARTMENT OF MATHEMATICS
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY
POHANG 790-784
KOREA
E-mail address: yaryong@hanmail.net
ON COMPARING THE COHOMOLOGY OF GENERAL LINEAR AND SYMMETRIC GROUPS

ALEXANDER S. KLESHCHEV AND DANIEL K. NAKANO

In this paper the authors explore relationships between the cohomology of the general linear group and the symmetric group. Stability results are given which show that the cohomology of these groups agree in a certain range of degrees.

1. Introduction.

1.1. Let \( k \) be an algebraically closed field, \( \text{GL}_n(k) \) be the general linear group over \( k \), and \( \Sigma_d \) be the symmetric group on \( d \) letters. For \( k = \mathbb{C} \), Frobenius and Schur discovered that the commuting actions of \( \text{GL}_n(k) \) and \( \Sigma_d \) on \( V^\otimes d \) can be used to relate the character theory of these two groups. For modular representations of these groups the relationship between their representation theories is not as direct. However, James \([12]\) showed that the decomposition matrix of the symmetric group is a submatrix of the decomposition matrix of the general linear group, and Erdmann \([9]\) used tilting modules to prove that one can also recover the decomposition numbers of the general linear groups from those of the symmetric groups.

Let \( M(n,d) \) be the category of polynomial representations of \( \text{GL}_n(k) \) of a fixed degree \( d \leq n \). The Schur functor \( \mathcal{F} \) is a certain covariant exact functor from \( M(n,d) \) to modules for the group algebra \( k\Sigma_d \) \([10]\). Some of the aforementioned results can be proved by using \( \mathcal{F} \). So it is natural to hope that the Schur functor can also be used to compare the cohomology theories of \( \text{GL}_n(k) \) and \( \Sigma_d \). In this paper we address the following question. Let \( M, N \in M(n,d) \). When is it true that

\[
\text{Ext}^i_{\text{GL}_n(k)}(M, N) \cong \text{Ext}^i_{k\Sigma_d}(\mathcal{F}(M), \mathcal{F}(N)) \ ?
\]

Here \( \text{Ext}^i_{\text{GL}_n(k)} \) is taken in the category \( M(n,d) \). By \([5, 2.1f]\), this is equivalent to considering extensions in the category of all rational \( \text{GL}_n(k) \)-modules. The case where \( i = 1 \) and \( M, N \) are irreducible representations is of particular importance.

The precise relationship between the two cohomology groups is given by the spectral sequence of \([7]\). It starts with extensions for \( \text{GL}_n(k) \) and converges to extensions for \( k\Sigma_d \). However, to use the spectral sequence one
needs to compute the higher right derived functors
\[ R^\bullet G(\cdot) \cong \text{Ext}^\bullet_{k\Sigma_d}(V^\otimes d, \cdot), \]
where \( G \) is a right adjoint functor to \( F \) (see Section 2.2). These derived functors are used extensively throughout the paper.

1.2. To describe the results of the paper we need some notation. Let us denote the simple (polynomial) \( \text{GL}_n(k) \) module with highest weight \( \lambda \) by \( L(\lambda) \). We also write \( \Delta(\lambda) \) and \( \nabla(\lambda) \) for the standard and costandard \( \text{GL}_n(k) \)-modules with highest weight \( \lambda \), respectively. It is well-known that \( F(L(\lambda)) \) is nonzero if and only if \( \lambda \) is \( p \)-restricted, and
\[ \{ F(L(\lambda)) \mid L(\lambda) \in M(n,d), \text{ and } \lambda \text{ is } p\text{-restricted} \} \]
is a complete set of the simple \( k\Sigma_d \)-modules up to isomorphism. If \( \lambda \) is \( p \)-restricted, we denote \( F(L(\lambda)) \) by \( D\lambda \).

A simple \( k\Sigma_d \)-module is called completely splittable (CS for short) if its restriction to any Young subgroup is semisimple. These modules were introduced in [15], see also [20, 17, 19]. It is possible to say explicitly which of the \( D\lambda \) are CS, see [15] and Section 4.2.

Now we describe the contents of the paper in greater detail. In Section 2, the basic facts about the Schur functor are reviewed. We also define the spectral sequence constructed in [7] and state the elementary properties of the higher right derived functors \( R^\bullet G(\cdot) \). In the following section the image of \( G \) on twisted Specht and twisted Young modules is computed (for \( p > 3 \)). The main result of the section (Theorem 3.2) can be interpreted as the isomorphism (1.1.1) for important classes of modules. Section 4 deals with CS modules. Our result in Theorem 4.4(a) says that
\[ G(D\lambda) = L(\lambda) \]
if \( D\lambda \) is CS and nontrivial (i.e., \( D\lambda \not\cong k \)). We note that the image of \( G \) on the simple \( k\Sigma_d \)-modules is generally very complicated. According to [7, 3.3], \( G(D\lambda) \) can be described as the largest submodule of the injective hull of \( L(\lambda) \) in the category \( M(n,d) \) whose only \( p \)-restricted composition factor is \( L(\lambda) \). Informally, \( G(D\lambda) \) is \( L(\lambda) \) with as many non-\( p \)-restricted modules on top as possible. In the light of this description the property (1.2.1) is quite remarkable. Using this result we show in Theorem 4.4(b) that if \( \lambda \) and \( \mu \) are \( p \)-restricted, \( D\lambda \) is CS, and \( D\lambda \not\cong k \) then
\[ \text{Ext}^1_{\text{GL}_n(k)}(L(\mu), L(\lambda)) \cong \text{Ext}^1_{k\Sigma_d}(D\mu, D\lambda). \]
This again provides us with many cases where (1.1.1) is indeed an isomorphism.

Section 5 contains vanishing results about the cohomology of symmetric groups with coefficients in CS modules. This information is later used to prove the vanishing of some \( R^\bullet G(D\lambda) \). The results of Section 5 allow us to
generalize the isomorphism \((1.2.2)\) to higher Ext-groups in a suitable range of degrees. This result is a portion of two general theorems on stability of extensions given in Sections 6.1 and 6.3, which give further examples where \((1.1.1)\) holds. For example, Corollary 6.1a claims that

$$H^i(\Sigma_d, F(N) \otimes \text{sgn}) \cong \text{Ext}^i_{\text{GL}_d(k)}(\delta, N) \quad \text{for } 0 \leq i \leq p - 2.$$  

Here, sgn is the one-dimensional sign representation for \(\Sigma_d\), \(\delta\) is the one-dimensional determinant representation for \(\text{GL}_d(k)\), and \(N \in M(n, d)\). Note that Cline, Parshall and Scott \([4]\), (12.4) have a similar result in non-describing characteristic relating the cohomology of the finite general linear group with cohomology for the \(q\)-Schur algebra. Their approach relies on using the Deodhar complex and is quite different from ours.

One can improve the stability results on the symmetric group cohomology with coefficients in a simple module if one considers the following setting. Let \(V\) be the natural \(n\)-dimensional \(\text{GL}_n(k)\)-module and \(S^d(V)\) be the \(d\)-th symmetric power of \(V\). Corollary 6.3(b) shows that for \(n \geq d\) and \(0 \leq i \leq 2(p - 2) + 1\),

$$H^i(\Sigma_d, D_\mu) \cong \text{Ext}^i_{\text{GL}_n(k)}(S^d(V), L(\mu)).$$

To interpret this statement in the context of \((1.1.1)\) one needs to note that \(F(S^d(V)) \cong k\).

Theorem 6.1 can also be used to prove a conjecture on the cohomology of dual Specht modules made in \([2]\). In fact, we get an even stronger result (Corollary 6.2) that for any Specht module \(S^\lambda\) over \(k\Sigma_d\) and \(1 \leq i \leq p - 2\),

\((1.2.3)\)

$$H^i(\Sigma_d, (S^\lambda)^*) = 0.$$  

This result can be interpreted as an isomorphism in \((1.1.1)\) with \(M = \wedge^d(V)\) and \(N\) being any costandard module. The vanishing result \((1.2.3)\) along with the work in Section 3 is used in Section 6.4 to prove stability results for extensions of Specht modules. This is followed by similar results in Section 6.5 on extensions of Young modules. Finally, in Section 6.6 we obtain some results on cohomology of the alternating groups.

This work was initiated while the second author was visiting the University of Oregon during the spring of 1998. The second author would like to acknowledge the Department of Mathematics at the University of Oregon for their hospitality and support. This work was completed during the conference on Algebraic Representation Theory held at Aarhus University in August 1998. Both authors thank Henning H. Andersen and Jens C. Jantzen for organizing a productive meeting.

2. Comparing \(\text{GL}_n(k)\) and \(k\Sigma_d\).

2.1. The Schur functor. The basic references for this section are \([10]\), \([11]\). Let \(k\) be an algebraically closed field of characteristic \(p > 0\) and let
The Schur functor $F$ is defined by $\hom A M (\otimes d)$ where $V$ is the natural representation of $G$. The category $M(n,d)$ of polynomial $G$-modules of a fixed degree $d \geq 0$ is equivalent to the category of modules for $S(n,d)$ and we do not distinguish between the two categories from now on. We denote by $\mod (k \Sigma_d)$ (resp. $\mod (k \Sigma_d)$) the category of all (resp. all finite dimensional) $k \Sigma_d$-modules.

Throughout the paper we assume that $n \geq d$. Let $e = \zeta_{(1,2,\ldots,d)(1,2,\ldots,d)}$ be the idempotent in $S(n,d)$ described in [10, (6,1)]. Then $e S(n,d) e \cong k \Sigma_d$. The Schur functor $F$ is the covariant exact functor from $M(n,d)$ to $\mod (k \Sigma_d)$ defined on objects by $F(M) = e M$.

The simple $S(n,d)$-modules are in bijective correspondence with set of partitions of $d$. We will denote this set by $\Lambda = \Lambda^+(n,d)$ and the corresponding simple $S(n,d)$-module by $L(\lambda)$ for $\lambda \in \Lambda$. Note that one can also identify $\Lambda$ as the set of dominant polynomial weights of $G$ of degree $d$. Moreover, if $\lambda \in \Lambda$, let $P(\lambda)$ be the projective cover of $L(\lambda)$ and $T(\lambda)$ be the corresponding tilting module. There exists a duality on $M(n,d)$ fixing simple modules called the transpose dual. This duality will be denoted by $\tau$. The duality $\tau$ and the usual duality $\ast$ in $\mod (k \Sigma_d)$ are compatible in the sense that $e M^{\tau} \cong (e M)^\ast$ for any finite dimensional $M \in M(n,d)$.

A partition $(\lambda_1, \lambda_2, \ldots)$ is called $p$-restricted if $\lambda_i - \lambda_{i+1} \leq p - 1$ for all $i$. As mentioned in Section 1.2, we label the simple $k \Sigma_d$-modules by the $p$-restricted partitions $\lambda \in \Lambda$. The set of the $p$-restricted partitions of $d$ will be denoted by $\Lambda_{\text{reg}}$. A partition $\lambda$ is called $p$-regular if its transpose $\lambda'$ is $p$-restricted. We denote the set of all $p$-regular partitions of $d$ by $\Lambda_{\text{reg}}$. In [11], the simple $k \Sigma_d$-modules are labelled by the $p$-regular partitions and denoted by $D^\lambda$. We will use both parametrizations so note a result from [10, §6]:

\begin{equation}
D^\lambda \cong D_{\lambda'} \otimes \sgn \quad \text{for any } \lambda \in \Lambda_{\text{reg}}.
\end{equation}

The Specht, Young, and permutation modules over $k \Sigma_d$ corresponding to a partition $\lambda \in \Lambda$ are denoted by $S^\lambda$, $Y^\lambda$, and $M^\lambda$, respectively. In particular $M^\lambda$ is the module induced from the trivial module over the Young subgroup $\Sigma_\lambda$. One has following correspondences between $S(n,d)$-modules and $k \Sigma_d$-modules under $F$ (see [10, §6] and [6, 3.5, 3.6]):

$$F(\nabla(\lambda)) = S^\lambda, \ F(\Delta(\lambda)) = (S^\lambda)^\ast, \ F(\varepsilon(\lambda)) = Y^\lambda, \ F(\psi(\lambda)) = Y^{\lambda'} \otimes \sgn.$$
Furthermore, \( G \) is a left inverse to \( F \). Since \( Ae \cong V^\otimes d \), and \( G \) takes injective \( eAe \)-modules to injective \( A \)-modules, one can construct a spectral sequence [7, 2.2]:

**Theorem (A).** Let \( M \in M(n,d) \), \( N \in \text{Mod}(k\Sigma_d) \) with \( n \geq d \). There exists a first-quadrant Grothendieck spectral sequence, with \( E_2 \)-page given by

\[
E^{i,j}_2 = \text{Ext}^i_{S(n,d)}(M, \text{Ext}^j_{k\Sigma_d}(V^\otimes d, N)) \Rightarrow \text{Ext}^{i+j}_{k\Sigma_d}(eM, N).
\]

For \( M \in M(n,d) \) and \( S \) a simple module in \( M(n,d) \), let \([M : S]\) be the multiplicity of \( S \) as a composition factor of \( M \). The following results from [7] provide information on composition factors of the higher right derived functors of \( G \).

**Theorem (B).** Let \( N \in \text{mod}(k\Sigma_d) \), and \( \mu \in \Lambda \). Then:

(i) \([R^jG(N) : L(\mu)] = \dim_k \text{Ext}^j_{k\Sigma_d}(Y^\mu, N)\) for \( j \geq 0 \).

(ii) In particular, \( e(R^jG(N)) = 0 \) for \( j > 0 \).

2.3. A standard spectral sequence argument yields the following useful result.

**Proposition.** Let \( n \geq d \), \( M \in M(n,d) \), and \( N \in \text{Mod}(k\Sigma_d) \). Suppose that \( R^jG(N) = 0 \) for \( 1 \leq j \leq t \). Then

\[
\text{Ext}^i_{S(n,d)}(M, G(N)) \cong \text{Ext}^i_{k\Sigma_d}(eM, N)
\]

for \( 0 \leq i \leq t + 1 \).

**Proof.** Consider the spectral sequence (2.2.1). By assumption,

\[
\text{Ext}^j_{k\Sigma_d}(V^\otimes d, N) = R^jG(N) = 0 \quad \text{for } 1 \leq j \leq t.
\]

Therefore, \( E^{i,j}_2 = 0 \) if \( j > 0 \) and \( 1 \leq i + j \leq t \). The spectral sequence has differentials \( d_r \) with bidegree \((r, 1-r)\). Therefore,

\[
E^{i,0}_2 = \text{Ext}^i_{S(n,d)}(M, G(N)) \cong \text{Ext}^i_{k\Sigma_d}(eM, N)
\]

for \( 0 \leq i \leq t + 1 \). \( \Box \)

3. Twisted Specht and Young modules.

In this section we prove that the image under the functor \( G \) of a twisted Young (resp. Specht) module is a tilting (resp. Weyl) module as long as \( p > 3 \). These results for twisted Young modules can be viewed as dual to the results proved in [3, 5.2.4]. The result in [3] holds for \( p > 2 \) and can be used to prove its dual version (see [8, 6.2]). But we employ a different approach than the one given in [3].
3.1. We deal with Young modules first.

**Lemma.** If $p > 3$ and $\lambda \in \Lambda$ then $G(Y^{\lambda'} \otimes \text{sgn})$ has a good filtration.

**Proof.** Let $M = \Delta(\mu)$ and $N = Y^{\lambda'} \otimes \text{sgn}$. From the five term exact sequence for the spectral sequence (2.2.1), we have

$$0 \to \text{Ext}^1_{S(n,d)}(\Delta(\mu), G(Y^{\prime} \otimes \text{sgn})) \rightarrow \text{Ext}^1_{k\Sigma}(k, G(Y^{\prime} \otimes \text{sgn})).$$

However,

$$\text{Ext}^1_{k\Sigma}(k, G(Y^{\prime} \otimes \text{sgn}) = 0$$

As $Y^{\lambda'}$ is a summand of the permutation module $M^{\lambda'}$, the Frobenius reciprocity implies that $\text{Ext}^1_{k\Sigma}(k, G(Y^{\prime} \otimes \text{sgn}))$ is a summand of $\text{Ext}^1_{k\Sigma}(k, (S^{\mu'})^* \otimes S^{\lambda'})$. Moreover, the restriction $(S^{\mu'})^* |_{\Sigma^{\lambda'}}$ has a Specht filtration by [13], so the last Ext-group is zero in view of [2, 3.4] or [8, 4.2]. Hence, $\text{Ext}^1_{S(n,d)}(\Delta(\mu), G(Y^{\prime} \otimes \text{sgn})) = 0$ for all $\mu \in \Lambda$, thus $G(Y^{\prime} \otimes \text{sgn})$ has a good filtration.

**Theorem.** If $G(Y^{\lambda'} \otimes \text{sgn})$ has a good filtration then $G(Y^{\lambda'} \otimes \text{sgn}) = T(\lambda)$. In particular for $p > 3$ we have $G(Y^{\lambda'} \otimes \text{sgn}) = T(\lambda)$ for all $\lambda \in \Lambda$.

**Proof.** As $d \leq n$, every standard module has a $p$-restricted socle [12]. Hence $T(\lambda)$ also has a $p$-restricted socle. So by [7, 3.3] there exists an exact sequence of the form

$$0 \rightarrow T(\lambda) \rightarrow G(Y^{\prime} \otimes \text{sgn}) \rightarrow X \rightarrow 0$$

where all composition factors of $X$ are not $p$-restricted. Since $T(\lambda)$ and $G(Y^{\prime} \otimes \text{sgn})$ have good filtrations, it follows that $X$ must have a good filtration. Therefore, $X = 0$ because each $\Delta(\mu)$ has a $p$-restricted head by [12] again.

3.2. We now prove that the image under $G$ of a twisted Specht module is a Weyl module.

**Theorem.** Let $p > 3$. Then $\Delta(\lambda) = G(S^{\lambda'} \otimes \text{sgn})$ for all $\lambda \in \Lambda$.

**Proof.** By [12], $\Delta(\lambda)$ has a $p$-restricted socle so in view of [7, 3.3] there exists a short exact sequence of the form

$$0 \rightarrow \Delta(\lambda) \rightarrow G(S^{\lambda'} \otimes \text{sgn}) \rightarrow X \rightarrow 0.$$

Therefore, it will suffice to prove that $\dim_k \Delta(\lambda) = \dim_k G(S^{\lambda'} \otimes \text{sgn})$ or equivalently that $\Delta(\lambda)$ and $G(S^{\lambda'} \otimes \text{sgn})$ have the same composition factors with multiplicities. According to Theorems 2.2B(i), 3.1 and [6, (3.8)], we
have
\[
[\mathcal{G}(S^\lambda \otimes \text{sgn}) : L(\mu)] = \dim_k \text{Hom}_{\Sigma_d}(Y^\mu, S^\lambda \otimes \text{sgn})
\]
\[
= \dim_k \text{Hom}_{\Sigma_d}(\langle S^\lambda \rangle^*, Y^\mu \otimes \text{sgn})
\]
\[
= \dim_k \text{Hom}_{S(n,d)}(\Delta(\lambda'), \mathcal{G}(Y^\mu \otimes \text{sgn}))
\]
\[
= \dim_k \text{Hom}_{S(n,d)}(\Delta(\lambda'), T(\mu'))
\]
\[
= [T(\mu') : \nabla(\lambda')]
\]
\[
= [\Delta(\lambda) : L(\mu)].
\]

3.3. For \( p > 2 \), \( \text{Ext}^1_{\Sigma_d}(Y^\lambda, Y^\mu) = 0 \) for \( \lambda, \mu \in \Lambda \) by [3, 4.6.1] or [8, 6.3]. The following theorem deals with extensions of Specht modules and extensions between Specht and Young modules. We note that part (a) was first observed in [3, 3.8.1].

**Theorem.** Let \( p > 3, \mu \in \Lambda_{\text{res}}, \) and \( \lambda \in \Lambda. \)

(a) \( \text{Ext}^1_{\Sigma_d}(S^\mu, Y^\lambda) = 0. \)

(b) \( \text{Ext}^1_{\Sigma_d}(S^\mu, S^\lambda) \cong \text{Ext}^1_{S(n,d)}(\Delta(\mu'), \Delta(\lambda')). \)

(c) If \( \mu \) does not strictly dominate \( \lambda \) then \( \text{Ext}^1_{\Sigma_d}(S^\mu, S^\lambda) = 0. \)

(d) In particular, \( \text{Ext}^1_{\Sigma_d}(S^\mu, S^\mu) = 0. \)

**Proof.** According to [7, 2.4A(ii)], one can let \( M = \Delta(\mu') \) to get

\[
\text{Ext}^1_{\Sigma_d}(S^\mu \otimes \text{sgn}, N) \cong \text{Ext}^1_{\Sigma_d}(\langle S^\mu \rangle^*, N) \cong \text{Ext}^1_{\Sigma_d}(\Delta(\mu'), \mathcal{G}(N)).
\]

For part (a) set \( N = Y^\lambda \otimes \text{sgn} \) and use Theorem 3.1. Part (b) follows by setting \( N = S^\lambda \otimes \text{sgn} \) and using Theorem 3.2. Note that if \( \mu \) does not strictly dominate \( \lambda \) then \( \lambda' \) does not strictly dominate \( \mu' \), and \( \text{Ext}^1_{S(n,d)}(\Delta(\mu'), \Delta(\lambda')) = 0 \) by [14, II 2.14 (3)] proving part (c), of which (d) is a special case. \( \square \)

4. Completely splittable modules.

4.1. Definition, [15].

A simple \( k\Sigma_d \)-module \( D \) is called **completely splittable** if the restriction \( D|_{\Sigma_\mu} \) is completely reducible for any Young subgroup \( \Sigma_\mu \) of \( \Sigma_d. \)

4.2. We now recall two basic results on CS modules proved in [15]. For a partition \( \lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s > 0) \) define

\[
h(\lambda) := s \quad \text{and} \quad \chi(\lambda) := \lambda_1 - \lambda_s + s.
\]

The invariant \( h(\lambda) \) is often referred to as the **height** of the partition. The first theorem provides necessary and sufficient conditions for a simple module to be CS.
Theorem (A). [15] Let \( \lambda \in \Lambda_{\text{reg}} \). Then \( D^\lambda \) is completely splittable if and only if \( \chi(\lambda) \leq p \).

For a removable node \( A \) of a partition \( \lambda \vdash n \) we denote by \( \lambda_A \) the partition of \( n-1 \) whose Young diagram is obtained from that of \( \lambda \) by removing \( A \). The second result describes how CS representations decompose upon restriction to \( \Sigma_{d-1} \).

Theorem (B). [15] Let \( \lambda \) be a \( p \)-regular partition of \( d \) with \( \chi(\lambda) \leq p \). Then \( D^\lambda|_{\Sigma_{d-1}} \cong \bigoplus D^{\lambda_A} \) where the sum is over all removable nodes \( A \) of \( \lambda \) such that \( \chi(\lambda_A) \leq p \).

4.3. The following is an easy consequence of the branching rule given in Theorem 4.2B.

Proposition. Let \( m > d \), \( \mu \) be any \( p \)-regular partition of \( d \), and \( \lambda \) be a \( p \)-regular partition of \( m \) satisfying \( h(\lambda) \leq s \) and \( \lambda_1 - \lambda_s + s \leq p \). Then \( D^\mu \) appears as a composition factor of \( D^\lambda|_{\Sigma_d} \) only if \( h(\mu) \leq s \) and \( \mu_1 - \mu_s + s \leq p \).

Proof. If \( h(\lambda) = s \) then \( \chi(\lambda) = \lambda_1 - \lambda_s + s \leq p \), and by Theorem 4.2A, \( D^\lambda \) is CS. If \( h(\lambda) < s \) then \( \lambda_s = 0 \), and \( \chi(\lambda) < \lambda_1 - \lambda_s + s \leq p \). So by Theorem 4.2A again, \( D^\lambda \) is CS in this case, too. Now apply Theorem 4.2B.

4.4. A \( k\Sigma_d \)-module is called nontrivial if it is not isomorphic to the trivial module \( k \). Our next result shows that nontrivial CS modules enjoy the remarkable property that they always give simple \( S(n, d) \)-modules upon application of the functor \( G \).

Theorem. Let \( D_\lambda \) be a nontrivial completely splittable \( k\Sigma_d \)-module and \( D_\mu \) be an arbitrary simple \( k\Sigma_d \)-module. Then

(a) \( G(D_\lambda) = L(\lambda) \).

(b) If \( n \geq d \) then \( \text{Ext}^1_{k\Sigma_d}(D_\mu, D_\lambda) \cong \text{Ext}^1_{S(n, d)}(L(\mu), L(\lambda)) \).

Proof. (a) According to Theorem 2.2B(i), we have \( [G(D_\lambda) : L(\nu)] = \dim_k \text{Hom}_{k\Sigma_d}(Y^\nu, D_\lambda) \). If \( \nu \) is \( p \)-restricted then \( Y^\nu \) is a projective \( k\Sigma_d \)-module with simple head \( D_\nu \). So the only \( p \)-restricted composition factor of \( G(D_\lambda) \) is \( L(\lambda) \). Now assume that \( \nu \) is not \( p \)-restricted. Pick \( \gamma \in \Lambda_{\text{reg}} \) so that \( D_\lambda = D^\gamma \). Then \( Y^\nu \) is a direct summand of \( M^\nu \) and

\[ \text{Hom}_{k\Sigma_d}(Y^\nu, D^\gamma) \subseteq \text{Hom}_{k\Sigma_d}(M^\nu, D^\gamma) \cong \text{Hom}_{k\Sigma_d}(k, D^\gamma|_{k\Sigma_\nu}). \]

We claim that the last space is zero. Indeed, \( \Sigma_\nu \cong \Sigma_{\nu_1} \times \Sigma_{\nu_2} \times \ldots \), with \( \nu_1 \geq p \) because \( \nu \) is not \( p \)-restricted. As \( D^\gamma \) is nontrivial, it follows that \( s := h(\gamma) \geq 2 \). Now by Proposition 4.3, \( D^\gamma|_{k\Sigma_\nu_1} \) does not contain a trivial component. This completes the proof of part (a).

(b) By [7, 4.2(i)], we have \( \text{Ext}^1_{k\Sigma_d}(D_\mu, D_\lambda) \cong \text{Ext}^1_{S(n, d)}(L(\mu), G(D_\lambda)) \). But by part (a), \( G(D_\lambda) = L(\lambda) \), which completes the proof. \( \square \)
Martin and the first author conjectured that for any simple module \( D \), \( \text{Ext}^1_{k \Sigma_d}(D, D) = 0 \) as long as \( p > 2 \). The next result shows this holds for CS representations. It is a special case of [16, 2.10 and Remark (iv) on page 2], which show that the conjecture is true for \( D = D^\lambda \) provided \( h(\lambda) \leq p - 1 \). However, we give a proof as our methods here are very different.

**Corollary.** Let \( p \geq 3 \) and \( D \) be any CS module over \( k \Sigma_d \). Then \( \text{Ext}^1_{k \Sigma_d}(D, D) = 0 \).

**Proof.** If \( D_\lambda \not\cong k \) then \( \text{Ext}^1_{k \Sigma_d}(D_\lambda, D_\lambda) = \text{Ext}^1_{S(n,d)}(L(\lambda), L(\lambda)) = 0 \) by Theorem 4.4b and [14, II.2.14]. On the other hand, if \( D_\lambda \cong k \) then \( \text{Ext}^1_{k \Sigma_d}(k, k) \cong \text{H}^1(\Sigma_d, k) = 0 \) as \( p > 2 \). \( \square \)

## 5. Vanishing results.

Theorem 4.4 motivates us to study images of CS modules under the higher derived functors of the functor \( \mathcal{G} \). A vanishing result depending on the height will be proved in Section 5.4. But first we need three lemmas on symmetric group cohomology.

### 5.1. First we calculate cohomology groups of \( \Sigma_p \) with coefficients in simple modules.

**Lemma.** The cohomology of \( \Sigma_p \) with coefficients in simple modules is given as follows.

\[
\begin{align*}
(i) & \quad \text{H}^j(\Sigma_p, D^\lambda) = 0 \quad \text{unless } \lambda \text{ is of the form } (p - i, 1^i) \text{ for } 0 \leq i \leq p - 2. \\
(ii) & \quad \text{For } 1 \leq i \leq p - 2,
\text{H}^j(\Sigma_p, D^{(p-i,1^i)}) \cong \begin{cases} 
  k & \text{if } j \text{ is of the form } 2m(p-1) + i \text{ or } \\
  2m(p-1) + (2p - i - 3) & \text{for some integer } m \geq 0; \\
  0 & \text{otherwise.}
\end{cases}
\text{(iii) For } i = 0,
\text{H}^j(\Sigma_p, k) \cong \begin{cases} 
  k & \text{if } j \text{ is of the form } 2m(p-1) \text{ or } 2m(p-1) - 1 \\
  0 & \text{for some integer } m \geq 0; \\
\end{cases}
\end{align*}
\]

**Proof.** Let \( P_i \) be the projective cover of \( D^{(p-i,1^i)}, 0 \leq i \leq p - 2 \). The module \( P_i \) admits a Specht filtration with factors \( S^{(p-i,1^i)}, S^{(p-i-1,1^{i+1})} \) (starting from the top). Since \( P_i \) is self-dual it also has a filtration with factors \( (S^{(p-i-1,1^{i+1})})^*, (S^{(p-i,1^i)})^* \). Therefore, a minimal projective resolution can be constructed as follows:

\[
\begin{align*}
0 \leftarrow k \leftarrow P_0 \leftarrow P_1 \leftarrow \cdots \leftarrow P_{p-2} \leftarrow P_{p-2} \leftarrow P_{p-3} \leftarrow \cdots \leftarrow P_0 & \leftarrow \cdots.
\end{align*}
\]
Since the resolution is minimal and $D^{(p-1)}_{j}$ is a simple module, we have $H^{j}(\Sigma_{p},D^{(p-1)}) \cong \text{Hom}_{k\Sigma_{p}}(Q_{j},D^{(p-1)})$, where $Q_{j}$ is the $j$th term of the resolution. The lemma follows. \hfill \square

5.2. We now prove a vanishing result on cohomology of $\Sigma_{p^{m}}$ with coefficients in CS modules, using Lemma 5.1 as an induction base. Note that all simple modules for $k\Sigma_{p}$ are CS.

**Lemma.** Let $m \geq 1$ and $D^\lambda$ be a CS module of $\Sigma_{p^{m}}$.

(a) If $D^\lambda \not\cong k$, $h(\lambda) \leq s$ and $\lambda_{1} - \lambda_{s} + s \leq p$, then $H^{j}(\Sigma_{p^{m}},D^\lambda) = 0$ for $0 \leq j \leq s - 2$.

(b) If $D^\lambda \cong k$ then $H^{j}(\Sigma_{p^{m}},D^\lambda) = 0$ for $1 \leq j \leq 2(p - 2)$.

**Proof.** We apply induction on $m$. For $m = 1$ the result follows from Lemma 5.1. Let $m > 1$. The Sylow $p$-subgroup $\text{Syl}_{p}(\Sigma_{p^{m}})$ embeds into the wreath product $A := \Sigma_{p^{m-1}} \wr \Sigma_{p} < \Sigma_{p^{m}}$. Hence $H^{a}(\Sigma_{p^{m}},D^\lambda)$ embeds into $H^{a}(A,D^\lambda)$. Note that $A \cong B \rtimes \Sigma_{p}$ where $B$ is the Young subgroup $\Sigma_{(p^{m-1},p^{m-1},\ldots,p^{m-1})} \cong \Sigma_{p^{m-1}} \times \cdots \times \Sigma_{p^{m-1}} < \Sigma_{p^{m}}$.

Now we apply the Lyndon-Hochschild-Serre spectral sequence

$$E_{2}^{ij} \cong H^{i}(\Sigma_{p},H^{j}(B,D^\lambda)) \Rightarrow H^{i+j}(A,D^\lambda).$$

By the Künneth formula, we get

$$H^{j}(B,D^\lambda) \cong \bigoplus_{(j_{1},\ldots,j_{p}) \in \mathbb{Z}_{+}^{p}} H^{j_{1}}(\Sigma_{p^{m-1}},D_{1}) \otimes \cdots \otimes H^{j_{p}}(\Sigma_{p^{m-1}},D_{p})$$

where the first sum is over all composition factors $D_{1} \otimes \cdots \otimes D_{p}$ of the restriction $D^\lambda|_{B}$. Note that we have used the fact that the restriction is completely reducible. If $D^\lambda \cong k$ and $1 \leq j \leq 2(p - 2)$ then by inductive hypothesis the last expression is zero. So assume that $D^\lambda$ is as given in part (a) and $0 \leq j \leq s - 2$. By Proposition 4.3, none of the modules $D_{q}$ in a composition factor $D_{1} \otimes \cdots \otimes D_{p}$ of $D^\lambda|_{B}$ is trivial. On the other hand, $j_{1} + \cdots + j_{p} = j \leq s - 2 \leq p - 3$ implies that at least one of the indices $j_{q}$ in the sum above must be 0. So for this $q$ we have $H^{0}(\Sigma_{p^{m-1}},D_{q}) = 0$. This shows that $H^{j}(B,D^\lambda) = 0$ under our assumptions.

Finally, we prove part (a). As $H^{j}(B,D^\lambda) = 0$ for $0 \leq j \leq s - 2$, it follows that $E_{2}^{ij} = 0$ for $0 \leq i + j \leq s - 2$ in the spectral sequence (5.2.1). The differentials in this spectral sequence have bidegree $(r,1-r)$, so $H^{a}(A,D^\lambda) = 0$ for $0 \leq n \leq s - 2$. Part (b) can be proved by using a similar line of reasoning and the fact that in this case $H^{j}(B,D^\lambda) = 0$ for $1 \leq j \leq 2(p - 2)$ and $E_{2}^{i,0} = H^{j}(\Sigma_{p},H^{0}(G,k)) = H^{j}(\Sigma_{p},k) = 0$ for $1 \leq i \leq 2(p - 2)$. \hfill \square
5.3.  Now we get a general vanishing result on cohomology of CS modules.

**Lemma.** Let \( D^\lambda \) be a CS module of \( \Sigma_d \).

(a) If \( D^\lambda \not\cong k \), \( h(\lambda) \leq s \) and \( \lambda_1 - \lambda_s + s \leq p \) then \( H^j(\Sigma_d, D^\lambda) = 0 \) for \( 0 \leq j \leq s - 2 \).

(b) If \( D^\lambda \cong k \) then \( H^j(\Sigma_d, k) = 0 \) for \( 1 \leq j \leq 2(p - 2) \).

**Proof.** If \( d = p^1 + p^2 + \ldots + p^k + r \) for \( i_1 \geq i_2 \geq \cdots \geq i_q > 0 \) and \( 0 \leq r < p \), then there is a Sylow \( p \)-subgroup of \( \Sigma_d \) contained in the Young subgroup \( G := \Sigma_{(p^1, \ldots, p^q)} \). This implies that \( H^j(\Sigma_d, D^\lambda) \) embeds into \( H^j(G, D^\lambda) \). By the K"unneth formula, we get

\[
H^j(G, D^\lambda) \cong \bigoplus_{(j_1, \ldots, j_q) \in (\mathbb{Z}^+)^q} H^{j_1}(\Sigma_{p^1}, D^{\mu_1}) \otimes \cdots \otimes H^{j_q}(\Sigma_{p^q}, D^{\mu_q})
\]

where the first sum is over all composition factors \( D^{\mu_1} \otimes \cdots \otimes D^{\mu_q} \) of the restriction \( D^\lambda|_G \) (we have used the fact that the restriction is completely reducible). Furthermore, by Proposition 4.3, each \( \mu(l) \) satisfies \( h(\mu(l)) \leq s \) and \( \mu(l)_1 - \mu(l)_s + s \leq p \). So, by Lemma 5.2, \( H^j(G, D^\lambda) = 0 \) for \( 0 \leq j \leq s - 2 \) if \( D^\lambda \not\cong k \), and \( H^j(G, k) = 0 \) for \( 1 \leq j \leq 2(p - 2) \). \( \square \)

5.4.  Finally, we obtain an information on vanishing of the higher derived functors of \( G \).

**Theorem.** Let \( D^\lambda \) be a CS module of \( \Sigma_d \).

(a) If \( D^\lambda \not\cong k \), \( h(\lambda) \leq s \) and \( \lambda_1 - \lambda_s + s \leq p \) then \( R^jG(D^\lambda) = 0 \) for \( 1 \leq j \leq s - 2 \).

(b) If \( D^\lambda \cong k \) then \( R^jG(k) = 0 \) for \( 1 \leq j \leq 2(p - 2) \).

**Proof.** (a) Using the definition of \( G \) and the decomposition of \( V^{\otimes d} \) as a \( k\Sigma_d \)-module, we have

\[
R^jG(D^\lambda) \cong \bigoplus_{\mu \leq d} \text{Ext}^j_{k\Sigma_d}(V^{\otimes n}, D^{\lambda}) \cong \bigoplus_{\mu \leq d} \text{Ext}^j_{k\Sigma_d}(M^{\mu}, D^{\lambda})
\]

where the sums are over all compositions \( \mu \) of \( d \). For \( 1 \leq j \leq s - 2 \) and a composition \( \mu = (\mu_1, \ldots, \mu_a) \) of \( d \) we prove that \( \text{Ext}^j_{k\Sigma_d}(k, D^{\lambda}) \). Let \( D^{\lambda(1)} \otimes \cdots \otimes D^{\lambda(a)} \) be a composition factor of \( D^{\lambda}|_{\Sigma_\mu} \). According to Proposition 4.3, every partition \( \lambda(i) \) satisfies \( h(\lambda(i)) \leq s \) and \( \lambda(i)_1 - \lambda(i)_s + s \leq p \). Now \( \text{Ext}^j_{k\Sigma_\mu}(k, D^{\lambda(1)} \otimes \cdots \otimes D^{\lambda(a)}) = 0 \) by Lemma 5.3a and the K"unneth formula. Hence, \( \text{Ext}^j_{k\Sigma_\mu}(k, D^{\lambda}|_{\Sigma_\mu}) = 0 \).

(b) As in (a), \( R^jG(k) \cong \bigoplus \text{Ext}^j_{k\Sigma_\mu}(k, k) \) where the sum is over all compositions \( \mu \) of \( d \). But, \( \text{Ext}^j_{k\Sigma_\mu}(k, k) = 0 \) for \( 1 \leq j \leq 2(p - 2) \) by the K"unneth formula and Lemma 5.3b. \( \square \)
Remark. The sign representation is CS and nontrivial if \( p > 2 \). It is equal to some \( D^p \) with \( h(\varepsilon) = p - 1 \) if \( d \geq p - 1 \) ([11, 24.5(iii)]). As \( k\Sigma_d \) is semisimple for \( d < p \), the theorem above implies

\[(5.4.1) \quad R^jG(\text{sgn}) = 0 \quad \text{for } 1 \leq j \leq p - 3.\]

6. Applications.

6.1. Stability of Extensions I. The algebra \( S(n, d) \) is quasi hereditary and thus has finite global dimension. On the other hand, \( k\Sigma_d \) has infinite global dimension. So there is no hope for the extension groups of these algebras to agree in all degrees. Nevertheless, often one can show that the extension groups for \( S(n, d) \) and \( k\Sigma_d \) coincide for a certain range of degrees.

**Theorem.** Let \( D_\lambda = D^\nu \) be a nontrivial CS module with \( h(\nu) \geq 3 \) and let \( M \in M(n,d) \).

(a) For \( 0 \leq i \leq h(\nu) - 1 \) we have \( \text{Ext}^i_{S(n,d)}(M, L(\lambda)) \cong \text{Ext}^i_{k\Sigma_d}(eM, D_\lambda) \).

(b) In particular for any \( \mu \in \Lambda \) and \( 0 \leq i \leq h(\nu) - 1 \) we have:

(i) \( \text{Ext}^i_{S(n,d)}(L(\mu), L(\lambda)) \cong \begin{cases} \text{Ext}^i_{k\Sigma_d}(D_\mu, D_\lambda) & \text{if } \mu \in \Lambda_{\text{res}}; \\ 0 & \text{otherwise.} \end{cases} \)

(ii) \( \text{Ext}^i_{S(n,d)}(\nabla(\mu), L(\lambda)) \cong \text{Ext}^i_{k\Sigma_d}(S^\mu, D_\lambda) \).

(iii) \( \text{Ext}^i_{S(n,d)}(\Delta(\mu), L(\lambda)) \cong \text{Ext}^i_{k\Sigma_d}((S^\mu)^*, D_\lambda) \).

(iv) \( \text{Ext}^i_{S(n,d)}(T(\mu), L(\lambda)) \cong \text{Ext}^i_{k\Sigma_d}(Y^{\mu'}, D^{\lambda'}) \).

(v) \( \text{Ext}^i_{k\Sigma_d}(Y^\mu, D_\lambda) \cong \begin{cases} k & \text{if } \lambda = \mu \text{ and } i = 0; \\ 0 & \text{otherwise.} \end{cases} \)

**Proof.** (a) By Theorem 4.4a, \( G(D_\lambda) = L(\lambda) \), and by Theorem 5.4, \( R^jG(D_\lambda) = 0 \) for \( 1 \leq j \leq h(\nu) - 2 \). The result now follows from Proposition 2.3.

(b) follows from (a). Indeed, for part (ii) (resp. (iii)) set \( M = \nabla(\mu) \) (resp. \( M = \Delta(\mu) \)) and using the results stated in Section 2.1. For (i) set \( M = L(\mu) \) and use the fact that \( eL(\lambda) = 0 \) for \( \lambda \notin \Lambda_{\text{res}} \). To prove part (iv), set \( M = T(\mu) \) and use (2.1.1). For part (v), set \( M = P(\mu) \) where \( P(\mu) \) is the projective cover of \( L(\mu) \). Observe that \( \text{Ext}^i_{S(n,d)}(P(\mu), L(\lambda)) \) is zero if \( i > 0 \) or \( \lambda \neq \mu \) and isomorphic to \( k \) if \( i = 0 \) and \( \lambda = \mu \).

Remark. (a) Note that even in the case \( i = 0 \) the theorem above is saying something new. Part (v) implies that a nontrivial CS module \( D_\lambda \) appears in the head of a Young module \( Y^\mu \) if \textit{and only if} \( \lambda = \mu \). If \( \mu \) is not \( p \)-restricted, the head of \( Y^\mu \) need not be simple.

(b) Parts (iv) and (v) of the theorem can be used to prove some vanishing results for \( \text{Ext}^i_{S(n,d)}(T(\mu), L(\lambda)) \). We leave the formulation of the corresponding results to the reader.
Corollary. Let $M \in M(d, d)$ and $\delta = L(1^d)$ be the determinant representation for $S(d, d)$.

(a) For $0 \leq i \leq 2$, $H^i(\Sigma_d, eM \otimes \text{sgn}) \cong \text{Ext}^i_{S(d, d)}(\delta, M)$.
(b) In particular for any $\mu \in \Lambda$ and $0 \leq i \leq 2$ we have:

(i) $H^i(\Sigma_d, D^\mu) \cong \text{Ext}^i_{S(d, d)}(\delta, L(\mu))$ for $\mu \in \Lambda_{\text{res}}$.

(ii) $H^i(\Sigma_d, S^\mu) \cong \text{Ext}^i_{S(d, d)}(\delta, V(\mu))$.

(iii) $H^i(\Sigma_d, (S^\mu)^*) \cong \begin{cases} k & \text{if } \mu = (1^d) \text{ and } i = 0; \\ 0 & \text{otherwise.} \end{cases}$

(iv) $H^i(\Sigma_d, Y^\mu \otimes \text{sgn}) \cong \begin{cases} k & \text{if } \mu = (1^d) \text{ and } i = 0; \\ 0 & \text{otherwise.} \end{cases}$

Proof. (a) Let $D_\lambda = \text{sgn}$. Then by Theorem 6.1a and (5.4.1) we have for $0 \leq i \leq 2$,

$H^i(\Sigma_d, eM \otimes \text{sgn}) \cong \text{Ext}^i_{k\Sigma_d}((eM)^*, \text{sgn}) \cong \text{Ext}^i_{k\Sigma_d}(eM^\tau, \text{sgn}) \cong \text{Ext}^i_{S(n,d)}(M^\tau, \delta) \cong \text{Ext}^i_{S(n,d)}(\delta, M)$.

(b) By (2.1.1), $D^\mu = D_\mu \otimes \text{sgn}$, and $(S^\mu)^* = S^\mu \otimes \text{sgn}$ so parts (i)-(iii) follow. Part (iv) follows from Theorem 6.1b(v). \qed

6.2. BKM conjecture. We now note that a conjecture made in [2, Conj. 6.2] (stated in the corollary below) is a rather special case of Corollary 6.1b(iii).

Corollary. For a fixed $i > 0$ there exists a constant $C$ depending only on $i$ such that for $p > C$ we have $H^i(\Sigma_d, (S^\lambda)^*) = 0$ for all $d$ and $\lambda \in \Lambda$. In fact, we can take $C = i + 1$.

Proof. Let $i > 0$. Set $C = i + 1$. If $p > i + 1$, or in other words $i < p - 1$ then $H^i(\Sigma_d, (S^\lambda)^*) = 0$ by Corollary 6.1b(iii) for all $d$ and $\lambda$. \qed

6.3. Stability of Extensions II. We can improve the range of the stability of extensions as long as one of the modules involved is $\Delta(d) \cong S^d(V)^\tau$, the contravariant dual of the $d$th symmetric power of the natural module $V$.

Theorem. Let $M \in M(n, d)$ with $n \geq d$.

(a) For $0 \leq i \leq 2(p - 2) + 1$ we have $\text{Ext}^i_{S(n,d)}(M, \Delta(d)) \cong \text{Ext}^i_{k\Sigma_d}(eM, k)$.
(b) In particular for any $\mu \in \Lambda$ and $0 \leq i \leq 2(p - 2) + 1$ we have:

(i) $\text{Ext}^i_{S(n,d)}(L(\mu), \Delta(d)) \cong \begin{cases} \text{Ext}^i_{k\Sigma_d}(D_\mu, k) & \text{if } \mu \in \Lambda_{\text{res}}; \\ 0 & \text{otherwise.} \end{cases}$

(ii) $\text{Ext}^i_{S(n,d)}(\nabla(\mu), \Delta(d)) \cong \text{Ext}^i_{k\Sigma_d}(S^\mu, k)$.

(iii) $\text{Ext}^i_{S(n,d)}(\Delta(\mu), \Delta(d)) \cong \text{Ext}^i_{k\Sigma_d}((S^\mu)^*, k)$.

(iv) $\text{Ext}^i_{S(n,d)}(T(\mu), \Delta(d)) \cong \text{Ext}^i_{k\Sigma_d}(Y^\mu, \text{sgn})$. 
will now be used to obtain stability results

Proof. (a) We have \( G(k) = \Delta(d) \) by [7, 5.5] and \( R^jG(k) = 0 \) for \( 1 \leq j \leq 2(p-2) \) by Theorem 5.4. Now use Proposition 2.3.

(b) Parts (i)-(v) follows from (a) by setting \( M = \nabla(\mu) \) (resp. \( \Delta(\mu), T(\mu), P(\mu) \)) as in the proof of Theorem 6.1b. For (v) we also use the fact that \( \Delta(d) \) is multiplicity free. \( \square \)

**Corollary.** Let \( M \in M(n,d) \) with \( n \geq d \).

(a) For \( 0 \leq i \leq 2(p-2) + 1 \) we have \( H^i(\Sigma_d, eM) \cong Ext^i_{S(n,d)}(\nabla(d), M) \).

(b) In particular for any \( \mu \in \Lambda \) and \( 0 \leq i \leq 2(p-2) + 1 \) we have:

(i) \( H^i(\Sigma_d, D_\mu) \cong Ext^i_{S(n,d)}(\nabla(d), L(\mu)) \) for \( \mu \in \Lambda_{res} \).

(ii) \( H^i(\Sigma_d, (S^\mu)^*) \cong Ext^i_{S(n,d)}(\nabla(d), \Delta(\mu)) \).

(iii) \( H^i(\Sigma_d, S^\mu) \cong Ext^i_{S(n,d)}(\nabla(d), \nabla(\mu)) \).

(iv) \( H^i(\Sigma_d, Y^\mu \otimes \text{sgn}) \cong Ext^i_{S(n,d)}(\nabla(d), T(\mu)) \).

(v) \( H^i(\Sigma_d, Y^\mu) \cong \begin{cases} k & \text{if } i = 0 \text{ and } [\nabla(d) : L(\mu)] \neq 0; \\ 0 & \text{otherwise.} \end{cases} \)

Proof. (a) By Theorem 6.3a, one has for \( 0 \leq i \leq 2(p-2) + 1 \)

\[
H^i(\Sigma_d, eM) \cong Ext^i_{k\Sigma_d}(k, eM) \cong Ext^i_{k\Sigma_d}((eM)^*, k) \cong Ext^i_{k\Sigma_d}(eM^\tau, k) \\
\cong Ext^i_{S(n,d)}(M^\tau, \Delta(d)) \cong Ext^i_{S(n,d)}(\nabla(d), M).
\]

(b) For parts (i)-(iv) make the appropriate substitutions for \( M, (v) \) follows from Theorem 6.3b(v) above by dualizing. \( \square \)

**6.4. Stability of Extensions III.** A conjecture of Burichenko, Kleshchev and Martin verified in Section 6.2 will now be used to obtain stability results involving Specht modules.

**Theorem.** Let \( p > 3, 0 \leq i \leq p-2, M \in M(n,d) \), and \( \lambda, \mu \in \Lambda \). Then:

(a) \( Ext^i_{S(n,d)}(M, \Delta(\lambda)) \cong Ext^i_{k\Sigma_d}(eM, S^\lambda \otimes \text{sgn}) \).

(b) In particular,

(i) \( Ext^i_{S(n,d)}(L(\mu), \Delta(\lambda)) \cong Ext^i_{k\Sigma_d}(D^\mu, S^\lambda) \).

(ii) \( Ext^i_{S(n,d)}(\Delta(\mu), \Delta(\lambda)) \cong Ext^i_{k\Sigma_d}(S^\mu, S^\lambda) \).

(iii) \( Ext^i_{S(n,d)}(\nabla(\mu), \Delta(\lambda)) \cong Ext^i_{k\Sigma_d}(S^\mu, S^\lambda \otimes \text{sgn}) \).

(iv) \( Ext^i_{k\Sigma_d}(Y^\mu, S^\lambda) = 0 \).

(v) \( Ext^i_{\Sigma_d}(S^\lambda, Y^\mu) = 0 \).

Proof. (a) By Theorem 3.2, \( G(S^\lambda \otimes \text{sgn}) = \Delta(\lambda) \). Moreover, by Frobenius reciprocity,

\[
R^jG(S^\lambda \otimes \text{sgn}) = Ext^j_{k\Sigma_d}(V^\otimes_d, S^\lambda \otimes \text{sgn}) = \oplus Ext^j_{k\mu}(k, (S^\lambda)^* \downarrow_\mu)
\]
where the last sum is over all compositions \( \mu \) of \( d \). By [13], \( (S^\lambda)^* \mid_{\Sigma_d} \) admits a filtration with sections of the form \( (S^{\lambda(1)})^* \otimes (S^{\lambda(2)})^* \otimes \cdots \otimes (S^{\lambda(a)})^* \). By the Künneth formula and Corollary 6.1b(iii), it follows that \( \text{Ext}^j_{k\Sigma_d}(k, (S^{\lambda(1)})^* \otimes (S^{\lambda(2)})^* \otimes \cdots \otimes (S^{\lambda(a)})^*) = 0 \) for \( 1 \leq j \leq p - 3 \). Consequently, by using induction on the length of the filtration on \( (S^\lambda)^* \), we have \( R^iG(S^\lambda \otimes \text{sgn}) = 0 \) for \( 1 \leq j \leq p - 3 \). Part (a) now holds by applying Proposition 2.3.

(b) For (i)-(v) substitute \( M = L(\mu), \Delta(\mu), \nabla(\mu), T(\mu) \) and \( P(\mu) \), respectively. \( \square \)

6.5. Stability of Extensions IV. We now obtain stability results on extensions involving Young modules.

**Theorem.** Let \( p > 3 \), \( 0 \leq i \leq p - 2 \), \( M \in M(n, d) \), and \( \lambda, \mu \in \Lambda \).

(a) \( \text{Ext}^i_{S(n, d)}(M, T(\lambda)) \cong \text{Ext}^i_{k\Sigma_d}(eM, Y^\lambda \otimes \text{sgn}) \).

(b) In particular,

(i) \( \text{Ext}^i_{S(n, d)}(L(\mu), T(\lambda)) \cong \text{Ext}^i_{k\Sigma_d}(D^\mu, Y^\lambda) \).

(ii) \( \text{Ext}^i_{S(n, d)}(\nabla(\mu), T(\lambda)) \cong \text{Ext}^i_{k\Sigma_d}(S^\mu \otimes \text{sgn}, Y^\lambda) = 0 \).

(iii) \( \text{Ext}^i_{\Sigma_d}(Y^\mu, Y^\lambda) = 0 \).

(iv) \( \text{Ext}^i_{\Sigma_d}(Y^\mu \otimes \text{sgn}, Y^\lambda) = 0 \).

**Proof.** (a) By Theorem 3.1, \( T(\lambda) = G(Y^\lambda \otimes \text{sgn}) \). Furthermore, if \( \mu \) is a composition of \( d \) then the restriction of \( Y^\lambda \otimes \text{sgn} \) to \( \Sigma_d \) admits a filtration with sections of the form \( (Y^{\lambda(1)} \otimes \text{sgn}) \otimes \cdots \otimes (Y^{\lambda(a)} \otimes \text{sgn}) \). Now use the same argument as in Theorem 6.4 along with Corollary 6.1b(iv) to show that \( R^iG(Y^\lambda \otimes \text{sgn}) = 0 \) for \( 1 \leq j \leq p - 3 \). Finally apply Proposition 2.3.

(b) Substitute \( M = L(\mu), \Delta(\mu), T(\mu) \) and \( P(\mu) \), respectively. \( \square \)

We remark that the results given in Theorem 6.4b(iv)-(v) and Theorem 6.5b(iii) can be viewed as natural generalizations involving higher extension groups of the \( \text{Ext}^1 \)-results given in the work of Cline, Parshall and Scott [3, §3.8, §4.6].

6.6. Alternating group cohomology. Let \( A_d \) be the alternating group on \( d \) letters. Nakaoaka computed the structure of the cohomology of the symmetric group \( H^\ast(\Sigma_d, k) \) [18]. For the prime \( p = 2 \), there is a method to compute \( H^\ast(A_d, k) \) by using the calculation of \( H^\ast(\Sigma_d, k) \) [1]. The first proposition shows that one can compute the cohomology of the alternating group by knowing the cohomology of the symmetric group with coefficients in the trivial and sign representations.

**Proposition.** Let \( p \geq 3 \). Then \( H^\ast(A_d, k) \cong H^\ast(\Sigma_d, k) \oplus H^\ast(\Sigma_d, \text{sgn}) \) as \( k \)-vector spaces.

**Proof.** Note that the module induced from the trivial \( kA_d \)-module to \( k\Sigma_d \) is isomorphic to \( k \oplus \text{sgn} \). The result now follows from the Eckmann-Shapiro lemma. \( \square \)
Corollary. Let \( p \geq 3 \). Then \( H^i(A_d, k) = 0 \) for \( 1 \leq i \leq p - 3 \).

Proof. This follows from the proposition above, Lemma 5.3 and Remark 5.4. \( \square \)

Remark. Consider the spectral sequence (2.2.1) with \( M = S^d(V) \) and \( N = \text{sgn} \):
\[
E_2^{i,j} = \text{Ext}_{S(n,d)}^i(S^d(V), R^j\mathcal{G}(\text{sgn})) \Rightarrow H^{i+j}(\Sigma_d, \text{sgn}).
\]
This indicates that there may be some hope to compute the cohomology of the alternating group by determining \( R^\bullet\mathcal{G}(\text{sgn}) = \text{Ext}_{k\Sigma_d}^\bullet(V^{\otimes d}, \text{sgn}) \) as a \( GL_n(k) \)-module.

References


Received November 18, 1999 and revised May 30, 2000. Research of both authors was supported in part by NSF (grants DMS-9600124 and DMS-9800960).

**Department of Mathematics**  
**University of Oregon**  
**Eugene OR 97403**  
*E-mail address:* klesh@math.uoregon.edu

**Department of Mathematics**  
**University of Georgia**  
**Athens, GA 30602**  
*E-mail address:* nakano@math.uga.edu
We define a new local invariant (called \textit{degeneracy}) associated to a triple \((M, M', H)\), where \(M \subset \mathbb{C}^N\) and \(M' \subset \mathbb{C}^{N'}\) are real submanifolds of \(\mathbb{C}^N\) and \(\mathbb{C}^{N'}\), respectively, and \(H : M \rightarrow M'\) is either a holomorphic map, a formal holomorphic map, or a smooth CR-map. We use this invariant to find sufficient conditions under which finite jet dependence, convergence and algebraicity results hold.

1. Introduction and statement of results.

In this paper, we discuss mappings of generic real submanifolds in complex spaces of different dimensions. We address the following specific problems:

- Give conditions which ensure that a mapping depends on its finite jet.
- Give conditions under which a formal mapping between real-analytic generic submanifolds is convergent.
- Give conditions under which a map between algebraic submanifolds is algebraic.

The first two questions have attracted considerable attention in the equidimensional case, and quite complete results have been obtained for the class of finitely nondegenerate manifolds (see [2], [4]), and more recently, for target manifolds of finite type in the sense of D’Angelo ([9]) in [3]. Whether similar results hold for mappings of generic submanifolds of spaces of different dimension is an intriguing problem which leads to some new geometric notions. The third question has also been answered in terms of characterizing the algebraic manifolds on which every holomorphic map is algebraic (see especially [18] and [1], [6], [7], [11], [17], [13], [16], [14]). We give a new sufficient condition, which can be checked using finitely many derivatives. For the definitions of a generic and a CR-manifold as well as other basic definitions, we refer the reader to e.g., [5].

Our starting point is the notion of \((k_0, s)\)-degeneracy. This is a local invariant associated to the triple \((M, M', H)\), where \(M \subset \mathbb{C}^N\) and \(M' \subset \mathbb{C}^{N'}\) are generic \(C^\infty\)-submanifolds of \(\mathbb{C}^N\) and \(\mathbb{C}^{N'}\), respectively, through 0, and \(H : M \rightarrow M'\) is a map (for example, \(C^\infty\)-CR) which in loose terms measures how “flat” \(H(M)\) is as a submanifold of \(M' \subset \mathbb{C}^{N'}\). The numbers \(k_0\) and \(s\)
can be defined (at 0) as follows: If \( \rho'_1, \ldots, \rho'_{d'} \) are defining functions for \( M' \), \( L_1, \ldots, L_n \) is a local basis for the CR-vector fields on \( M \), and \( H(0) = 0 \), then

\[
N' - s = \max_k \dim \text{span}_C \left\{ L^\alpha \rho'_{j,Z}(H(Z), \overline{H(Z)}) \big| Z = 0 : |\alpha| \leq k, 1 \leq j \leq d' \right\},
\]

where for a multi-index \( \alpha \in \mathbb{N}^n \) we write \( L^\alpha = L_1^{\alpha_1} \cdots L_n^{\alpha_n} \), and \( k_0 \) is the least integer \( k \) for which the maximum dimension on the right hand side of (1) is realized. Here we write \( N = n + d \), where \( d \) is the codimension of \( M \); similarly, we shall write \( N' = n' + d' \), where \( d' \) is the codimension of \( M' \). An extension of this definition is given in Section 2 in the context of formal submanifolds and formal maps, which allows us a unified treatment of real-analytic and smooth manifolds. This new notion is related to the concept of finite nondegeneracy of a real submanifold (which was introduced for hypersurfaces in [6]), and we explore this relationship further in 2.4.

Particularly satisfying is the situation for mappings for which \( s = 0 \). We call such mappings “nondegenerate”, or more specifically, \( k_0 \)-nondegenerate. These maps fulfill a sufficient condition to give a positive answer to all three points above; for example, every CR-diffeomorphism of class \( C^{k_0} \) of \((k_0)\)-finitely nondegenerate submanifolds of \( \mathbb{C}^N \), as introduced for hypersurfaces by Baouendi, Huang and Rothschild [6] (we will define those in Section 2) is a \( k_0 \)-nondegenerate map. The other maps allowing for a further treatment are the ones which are of constant degeneracy (to be defined in Section 2 as well).

Let us recall that a formal holomorphic map \( H = (H_1, \ldots, H_{N'}) : \mathbb{C}^N \rightarrow \mathbb{C}^{N'} \) at a point \( p_0 \) is an \( N' \)-tuple of formal power series \( H_j(Z) = \sum_\alpha c^0_{\alpha j}(Z - p_0)^\alpha \), and if \( H(p_0) = p'_0 \in \mathbb{C}^{N'} \), we write \( H : (\mathbb{C}^N, p_0) \rightarrow (\mathbb{C}^{N'}, p'_0) \) for such a map. If \( p_0 \in M \), \( p'_0 \in M' \) then we say that \( H : (\mathbb{C}^N, p_0) \rightarrow (\mathbb{C}^{N'}, p'_0) \) maps \( M \) into \( M' \) if the following property is satisfied: If \( \rho' = (\rho'_1, \ldots, \rho'_{d'}) \) is a defining function of \( M' \) and \( \rho = (\rho_1, \ldots, \rho_d) \) is a defining function of \( M \) (where \( d \) and \( d' \) are the codimensions of \( M \) and \( M' \), respectively), then there is a \( d' \times d \) matrix \( A \) of formal power series such that \( \rho' (H(Z), \overline{H(Z)}) = A(Z, \zeta) \rho (Z, \zeta) \).

(Here we are abusing notation: This equation shall hold in the sense of Taylor series.)

Let us recall that we say that \( M \) is of finite type at \( p \) (in the sense of Kohn-Bloom-Graham) if the CR and the anti-CR vectors together with their commutators of all length span the complexified tangent space of \( M \) at \( p \). We prove the following theorems. If not stated explicitly otherwise, all submanifolds are assumed to be smooth and connected.

**Theorem 1.** Let \( M, M' \) be generic real-analytic submanifolds of \( \mathbb{C}^N \) and \( \mathbb{C}^{N'} \), respectively, \( p_0 \in M \), \( M \) of finite type at \( p_0 \), \( p'_0 \in M' \), and assume that \( H : (\mathbb{C}^N, p_0) \rightarrow (\mathbb{C}^{N'}, p'_0) \) is a formal holomorphic map which maps \( M \) into
$M'$ and is $k_0$-nondegenerate at $p_0$. Then there exists a neighbourhood $U$ of $p_0$ in $\mathbb{C}^N$ on which $H$ is convergent.

For the next theorem, we denote by $j^k_{p_0} f$ the $k$-jet of $f$ at $p_0$.

**Theorem 2.** Let $M, M'$ be generic real submanifolds of $\mathbb{C}^N$ and $\mathbb{C}^{N'}$, respectively, $p_0 \in M$, and $M$ of finite type at $p_0$. There exists an integer $K$ such that if $H : M \to M'$ and $H' : M \to M'$ are $C^\infty$-CR mappings which are $k_0$-nondegenerate at $p_0 \in M$ and $j^K_{p_0} H = j^K_{p_0} H'$, then $j^l_{p_0} H = j^l_{p_0} H'$ for all $l$.

**Theorem 3.** Let $M, M'$ be generic real submanifolds of $\mathbb{C}^N$ and $\mathbb{C}^{N'}$, respectively, $p_0 \in M$, such that $M$ is of finite type at $p_0$. There exists an integer $K$ such that if $H : U \to \mathbb{C}^{N'}$ is a holomorphic map defined on some neighbourhood $U$ of $p_0$ with $H(U \cap M) \subset M'$ and such that $H$ is $k_0$-nondegenerate at $p_0$, and $H'$ is another holomorphic map defined on some neighbourhood $U'$ of $p_0$ with $H'(U' \cap M) \subset M'$ with

\[
\frac{\partial^\alpha H}{\partial Z^\alpha}(p_0) = \frac{\partial^\alpha H'}{\partial Z^\alpha}(p_0), \quad |\alpha| \leq K k_0,
\]

then $H = H'$.

Theorem 3 is an immediate consequence of Theorem 2. The proof of Theorem 1 and Theorem 2 is given in Section 3. Theorem 3 together with the reflection principle in [12] yields the following.

**Corollary 4.** Let $M, M'$ be generic real-analytic submanifolds of $\mathbb{C}^N$ and $\mathbb{C}^{N'}$, respectively, $p_0 \in M$, and $M$ of finite type at $p_0$. There exists an integer $K$ such that if $H : M \to M'$ and $H' : M \to M'$ are $C^{K k_0}$-CR mappings which are $k_0$-nondegenerate at $p_0 \in M$ and $j^K_{p_0} H = j^K_{p_0} H'$, then both extend to holomorphic mappings and $H = H'$.

Note that the notion of nondegeneracy makes sense even for maps which are a priori only smooth up to a certain finite order, so that the statement of this corollary makes sense. The last result we prove about nondegenerate maps is an algebraicity theorem.

**Theorem 5.** Let $M$ and $M'$ be algebraic generic submanifolds of $\mathbb{C}^N$ and $\mathbb{C}^{N'}$, respectively, $H$ a holomorphic map defined on some connected neighbourhood $U$ of $M$ with $H(M) \subset M'$ and such that $H$ is $k_0$-nondegenerate at some point of $M$. Then $H$ is algebraic.

Our next results are for hypersurfaces. They are valid either in the setting where $N' = N + 1$, and the hypersurfaces are assumed to be Levi-nondegenerate, or, where the target hypersurface is strictly pseudoconvex and the source hypersurface is of finite type (and there are no restrictions on $N'$). In the case of Levi-nondegenerate hypersurfaces, we will consider
maps $H$ which are (CR) transversal (the formal definition of this property is given in Definition 19). We will refer to the following properties in the theorems below:

(P1) $M'$ is strictly pseudoconvex at $p'_0$.
(P2) $N' = N + 1$ and $M$ and $M'$ are Levi-nondegenerate at $p_0$ and $p'_0$, respectively, and $H$ is transversal at $p_0$.

**Theorem 6.** Let $M$, $M'$ be real-analytic hypersurfaces in $\mathbb{C}^N$ and $\mathbb{C}^{N'}$, respectively, $p_0 \in M$, $p'_0 \in M'$, $M$ of finite type at $p_0$, and let $H : (\mathbb{C}^N, p_0) \to (\mathbb{C}^{N'}, p'_0)$ be a formal holomorphic map of constant degeneracy which maps $M$ into $M'$. Then there exists a neighbourhood $U$ of $p_0$ in $\mathbb{C}^N$ on which $H$ is convergent given that either (P1) or (P2) holds.

Note that the case $N' = N + 1$ is very special, as the following example shows.

**Example 1.** Let $M \subset \mathbb{C}^N$ be given by $\text{Im } w = \sum_{j=1}^n |z_j|^2$, and $M' \subset \mathbb{C}^{N+2}$ be given by $\text{Im } w' = |z_{n+2}|^2 - \sum_{j=1}^{n+1} |z'_j|^2$ (“adding a black hole”). Then the map $(z_1, \ldots, z_n, w) \mapsto (z_1, \ldots, z_n, f(z, w), f(z, w), w)$ maps $M$ into $M'$ for every (formal) holomorphic map $f : \mathbb{C}^N \to \mathbb{C}$.

This example also shows that in general, algebraicity and dependence on jets of finite order for Levi-nondegenerate hypersurfaces can only be expected in the case $N' = N + 1$ (without further restrictions on the mappings, as for example nondegeneracy as introduced above).

**Theorem 7.** Let $M$, $M'$ be real hypersurfaces in $\mathbb{C}^N$ and $\mathbb{C}^{N'}$, respectively, $p_0 \in M$, and $M$ of finite type at $p_0$. If $H : M \to M'$ and $H' : M \to M'$ are $C^\infty$-CR mappings which are constantly $(k_0, s)$-degenerate near $p_0 \in M$ with $j^{2k_0} H = j^{2k_0} H'$, then $j^l H = j^l H'$ for all $l$, provided that either (P1) or (P2) holds.

**Theorem 8.** Let $M$, $M'$ be real hypersurfaces in $\mathbb{C}^N$ and $\mathbb{C}^{N'}$, respectively, $p_0 \in M$, and $M$ of finite type at $p_0$. If $H : U \to \mathbb{C}^{N'}$ is a holomorphic map defined on some neighbourhood $U$ of $p_0$ with $H(U \cap M) \subset M'$ which is constantly $(k_0, s)$-degenerate at $p_0$, and $H'$ is another holomorphic map defined on some neighbourhood $U'$ of $p_0$ with $H'(U' \cap M) \subset M'$ with

$$\sum_{\alpha} \frac{\partial^{\alpha} H}{\partial Z^{\alpha}} (p_0) = \frac{\partial^{\alpha} H'}{\partial Z^{\alpha}} (p_0),$$

then $H = H'$, provided that either (P1) or (P2) holds.

We also have the following algebraicity result. Case (i) below is actually contained in the results in [18].

**Theorem 9.** Let $M$ and $M'$ be algebraic hypersurfaces in $\mathbb{C}^N$ and $\mathbb{C}^{N'}$, respectively, $H$ a holomorphic map defined on some connected neighbourhood...
U of M with \(H(U \cap M) \subset M'\). Then \(H\) is algebraic, provided that either of the following additional properties hold:

(i) There exists a point \(p_0\) in \(M\) where \(M\) is of finite type and \(M'\) is strictly pseudoconvex at \(H(p_0)\);

(ii) \(N' = N + 1\), and there exists a point \(p_0 \in M\) at which \(H\) is transversal, and \(M\) and \(M'\) are Levi-nondegenerate at \(p_0\) and \(H(p_0)\), respectively.

Theorem 8 is again an immediate consequence of Theorem 7. In the case \(N' = N + 1\), with the assumptions of the theorem, \(s = 0\) or \(s = 1\) (see Lemma 20); the case \(s = 0\) is covered by Theorems 2 and 3. The proofs of Theorems 6, 9 and 7 in the Levi-nondegenerate case are given in Section 4. The proof for strictly pseudoconvex hypersurfaces is given in Section 5.

Acknowledgements. The author thanks the referee for many helpful comments and suggestions as well as for pointing out many typographical errors. Furthermore, the author would like to thank S. Baouendi and L. Rothschild for their help and their interest in this work.

2. Formal holomorphic maps of constant degeneracy.

2.1. Some definitions. In this section we want to give a short review of some basic definitions. We will be very brief, focusing mostly on the facts which we shall need later on. The purpose of this section is mainly for reference. A thorough discussion of the definitions given here can be found in e.g., [3].

2.1.1. Formal submanifolds and formal maps. Consider the ring of formal power series \(\mathbb{C}[Z,\zeta]\) in the \(2N\) indeterminates \((Z,\zeta) = (Z_1, \ldots, Z_N, \zeta_1, \ldots, \zeta_N)\). A formal generic submanifold \(M \subset \mathbb{C}^N\) of codimension \(d\) is an ideal \(I \subset \mathbb{C}[Z,\zeta]\) which can be generated by \(d\) formal power series \(\rho_1(Z,\zeta), \ldots, \rho_d(Z,\zeta)\) such that \(\rho_j(\zeta,Z) \in I\), \(1 \leq j \leq d\) and such that the vectors \(\rho_1,Z(0), \ldots, \rho_d,Z(0)\) are linearly independent (where for \(\phi \in \mathbb{C}[Z,\zeta]\), \(\phi_Z = (\phi_{Z_1}, \ldots, \phi_{Z_N})\), and \(\overline{\phi}\) denotes the power series with complex conjugated coefficients). A formal holomorphic change of coordinates \(\tilde{Z} = F(Z)\) is the ring isomorphism between \(\mathbb{C}[\tilde{Z},\tilde{\zeta}]\) and \(\mathbb{C}[Z,\zeta]\) induced by \(\tilde{Z}_j = F_j(Z)\), \(\tilde{\zeta}_j = F_j(\zeta)\), where \(F = (F_1, \ldots, F_N)\) with \(F_j \in \mathbb{C}[Z]\), \(1 \leq j \leq N\), satisfying

\[
\det \left( \frac{\partial F_j}{\partial Z_k}(0) \right)_{1 \leq j \leq N, 1 \leq k \leq N} \neq 0.
\]

A formal holomorphic map \(H: \mathbb{C}^N \to \mathbb{C}^{N'}\) is an \(N'\)-tuple of formal power series in \(Z\), \(H = (H_1, \ldots, H_{N'})\), \(H_j \in \mathbb{C}[Z]\) for \(1 \leq j \leq N'\). A formal holomorphic map induces a ring homomorphism \(H^* : \mathbb{C}[Z',\zeta'] \to \mathbb{C}[Z,\zeta]\) by \(H^*(Z'_j) = H_j(Z), H^*(\zeta'_j) = \overline{H}_j(\zeta)\). Given two formal submanifolds \(M \subset \mathbb{C}^N\)
and $M' \subset \mathbb{C}^N$, of codimension $d$ and $d'$ respectively, represented by their ideals $I \subset \mathbb{C}[Z, \zeta]$ and $I' \subset \mathbb{C}[Z', \zeta']$ we say that $H$ maps $M$ into $M'$ if $H^\sharp(I') \subset I$. (From now on we will write $H : M \rightarrow M'$ to indicate that we are in this situation.) This is the case if and only if for any generators $\rho = (\rho_1, \ldots, \rho_d)$ of $I$ and generators $\rho' = (\rho'_1, \ldots, \rho'_{d'})$ we have that

$$\rho'(H(Z), \Pi(\zeta)) = A(Z, \zeta)\rho(Z, \zeta),$$

where $A(Z, \zeta)$ is a $d' \times d$ matrix of formal power series in $(Z, \zeta)$. If in addition $d = d'$ and $H$ is a formal holomorphic change of coordinates, then $\det A(0, 0) \neq 0$, so that $A$ is an invertible matrix of formal power series.

We work with formal power series since if we want to handle $C^\infty$-submanifolds, then they are a convenient way of keeping track of all equations which we arrive from by repeated differentiation. This brings us to the subject of formal vector fields. A formal vector field $X$ is an operator of the form

$$X = \sum_{j=1}^N a_j(Z, \zeta) \frac{\partial}{\partial Z_j} + \sum_{j=1}^N b_j(Z, \zeta) \frac{\partial}{\partial \zeta_j},$$

where $a_j(Z, \zeta)$ and $b_j(Z, \zeta)$ are formal power series. One checks that the formal vector fields are exactly the derivations of the $\mathbb{C}$-algebra $\mathbb{C}[Z, \zeta]$. $X$ is of type $(1, 0)$ if $b_j(Z, \zeta) = 0$, $1 \leq j \leq N$, and of type $(0, 1)$ if $a_j(Z, \zeta) = 0$, $1 \leq j \leq N$, and this distinction is invariant under formal holomorphic changes of coordinates. A formal vector field is tangent to $M$ if $XI \subset I$ (as above, $I$ denotes the ideal representing $M$).

A formal CR-vector field tangent to $M$ is a formal vector field of type $(0, 1)$ tangent to $M$. We write $\mathcal{D}^{0,1}_M$ for the formal CR-vector fields tangent to $M$. This is a $\mathbb{C}[Z, \zeta]$-module. Finally, we say that $L_1, \ldots, L_n$ (where $n = N - d$) is a basis of the formal CR-vector fields tangent to $M$ if they generate the quotient module $\mathcal{D}^{0,1}_M / I \mathcal{D}^{0,1}_M$. For the anti-CR-vector fields tangent to $M$ (that is, the formal vector fields of type $(1, 0)$ tangent to $M$) we write $\mathcal{D}^{1,0}_M$, and set $\mathcal{D}_M = \mathcal{D}^{0,1}_M \oplus \mathcal{D}^{1,0}_M$. For the Lie algebra generated by $\mathcal{D}_M$, we write $\mathfrak{g}$. We say that $M$ is of finite type (at 0) if $\text{dim}_\mathbb{C} \mathfrak{g}(0) = 2n + d$.

2.1.2. Formal normal coordinates. Let $M$ be a generic formal submanifold of codimension $d$. Then after a formal holomorphic change of coordinates we can assume $Z = (z, w) = (z_1, \ldots, z_n, w_1, \ldots, w_d) \in \mathbb{C}^n \times \mathbb{C}^d$ (for the corresponding $\zeta$ we write $\zeta = (\chi, \tau)$) and that $I$ is generated by $d$ functions $w_j - Q_j(z, \chi, \tau)$, $j = 1, \ldots, d$, where $Q_j \in \mathbb{C}[z, \chi, \tau]$ fulfills

$$Q_j(z, 0, \tau) = Q_j(0, \chi, \tau) = \tau_j, \quad j = 1, \ldots, d.$$

In that case, another useful set of generators is given by $\tau_j = Q_j(\chi, z, w)$, $j = 1, \ldots, d$. We shall call such coordinates formal normal coordinates. We
can use them in order to parametrize $M$: Under this, we will understand the ring isomorphism

$$\mathbb{C}[z, w, \chi, \tau]/I \rightarrow \mathbb{C}[z, \chi, \tau]$$

or

$$\mathbb{C}[z, w, \chi, \tau]/I \rightarrow \mathbb{C}[\chi, z, w], \quad z \mapsto z, \ w \mapsto w, \ \chi \mapsto \chi, \ \tau \mapsto \tau.$$  

Note that a basis of CR-vector fields tangent to $M$ is given by

$$L_k = \frac{\partial}{\partial \chi^k} + \sum_{j=1}^{d} Q_{j\chi^k}(\chi, z, w) \frac{\partial}{\partial \tau_j}, \quad k = 1, \ldots, n.$$  

For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ we write $L^{\alpha} = L_1^{\alpha_1} \cdots L_n^{\alpha_n}$. Let now $\phi(z, w, \chi, \tau)$ be a formal power series. We want to relate the image of $\phi$ in the parametrization of $M$ given by (9) with its derivatives along CR-directions. Expand $\phi$ as a series in $\mathbb{C}[z, w][[\chi]]$:

$$\phi(z, w, \chi, Q(\chi, z, w)) = \sum_{\alpha \in \mathbb{N}^n} s_{\alpha}(z, w) \frac{\phi(z, w)}{\alpha!} \chi^\alpha.$$  

Now the $s_{\alpha}$ are obtained by partial differentiation:

$$s_{\alpha}(z, w) = \left. \frac{\partial |\alpha|}{\partial \chi^\alpha} \phi(z, w, \chi, Q(\chi, z, w)) \right|_{\chi=0} = L^{\alpha} \phi(z, w, 0, w).$$  

The last equality is proved by induction on $|\alpha|$ and its proof is left to the reader.

### 2.1.3. Segre-mappings and a finite type criterion.

Again, we are considering a generic formal submanifold $M \subset \mathbb{C}^N$ of codimension $d$. Assume that formal normal coordinates $(z, w)$ as in 2.1.2 have been chosen, along with the corresponding (vector valued) function $Q(z, \chi, \tau) \in \mathbb{C}[z, \chi, \tau]^d$, fulfilling (7), such that $I$ (the ideal associated to $M$) is generated by $w_j - Q_j(z, \chi, \tau)$, $j = 1, \ldots, d$ (or, equivalently, by $\tau_j - Q_j(\chi, z, w)$). The **Segre mappings** are the formal mappings $v^k : \mathbb{C}^k \rightarrow \mathbb{C}^N$, for any integer $k$, defined by

$$v^0 = (0, 0), \quad v^1(z) = (z, 0),$$  

$$(v^{2j}(z, \chi^1, \ldots, z^{j-1}, \chi^j) = (z, Q(z, \chi^1, Q(\chi^1, z^1, \ldots, Q(z^{j-1}, \chi^j, 0) \ldots)))),$$  

$$(v^{2j+1}(z, \chi^1, \ldots, \chi^j, z^j) = (z, Q(z, \chi^1, Q(\chi^1, z^1, \ldots, Q(\chi^j, z^j, 0) \ldots))).$$
We will write \((z, \chi^1, z^1, \ldots) = (z, \xi), \) where \(\xi = (\chi^1, z^1, \ldots) \in \mathbb{C}^{(k-1)n}.\) These mappings have the property that for every \(k \geq 0,\) and for every \(f \in I,\)
\[
(14) \quad f(v^{k+1}(z, \xi), \varpi^k(\xi)) = 0.
\]
We shall need the finite type criterion of Baouendi, Ebenfelt and Rothschild, which we state here for reference; see e.g., [3] or [5].

Let \(F : \mathbb{C}^p \to \mathbb{C}^r\) be a formal mapping, that is, an \(r\)-tuple of formal power series in \(p\) variables \((x_1, \ldots, x_p).\) We denote by \(\text{rk} F\) the rank of the Jacobian matrix of \(F\) over the quotient field of the ring of formal power series \(\mathbb{C}[x_1, \ldots, x_p].\)

**Theorem 10.** Let \(M\) be a formal generic submanifold of \(\mathbb{C}^N\) of codimension \(d.\) Then, \(M\) is of finite type at 0 if and only if there exists a \(k_1 \leq d+1\) such that \(\text{rk}(v^k) = N\) for \(k \geq k_1.\) Moreover, if \(M\) is real-analytic and of finite type at 0, then there exists \((z_0, \xi_0) \in \mathbb{C}^n \times \mathbb{C}^{(2k_1-1)n}\) arbitrarily close to the origin such that \(v^{2k_1}(z_0, \xi_0) = 0\) and the rank of the Jacobian matrix of \(v^{2k_1}\) at \((z_0, \xi_0)\) is \(N.\)

Note the following consequence for real-analytic submanifolds: There exist points \((z_0, \xi_0) \in \mathbb{C}^n \times \mathbb{C}^{2k_1},\) arbitrarily close to 0, such that the function \(v^{2k_1}\) has a holomorphic right inverse \(\psi : \mathbb{C}^N \to \mathbb{C}^n \times \mathbb{C}^{2k_1}\) defined in a neighbourhood of \(0 \in \mathbb{C}^N,\) such that \(\psi(0) = (z_0, \xi_0)\) and \(v^{2k}(\psi(Z)) = Z.\) This follows from the inverse function theorem and Theorem 10.

### 2.2. Constant-rank submodules

We are now considering a free module of rank \(k\) over \(\mathbb{C}[z, \zeta].\) Let \(E \subset \mathbb{C}[z, \zeta]^k\) be a submodule and \(I \subset \mathbb{C}[z, \zeta]\) an ideal.

**Definition 11.** The submodule \(E \subset \mathbb{C}[z, \zeta]^k\) is of constant rank \(l = \dim_{\mathbb{C}} E(0)\) over \(I\) if for all \(e_1, \ldots, e_l, e_{l+1} \in E,\) where \(e_j = (e_j^1, \ldots, e_j^k),\) every subminor of length \(l + 1\) of the matrix \((e_{m}^{i})_{1 \leq m \leq l+1, 1 \leq i \leq k}\) is an element of \(I.\)

Definition 11 is motivated by the following observation. Consider a smooth submanifold \(M \subset \mathbb{C}^N.\) By taking the Taylor series of the defining functions of \(M\) at a point, one obtains a formal submanifold represented by some ideal \(I.\) Consider a vector bundle of rank \(l\) over \(M,\) embedded in \(\mathbb{C}^k.\) Its sections will be a submodule of \(\Gamma(M, \mathbb{C}^k) \cong C^\infty(M)^k\) (the isomorphism is determined by a choice of coordinates). Taking the Taylor expansion of the components, we get a submodule \(E \subset \mathbb{C}[z, \zeta]^k\) which fulfills the condition of Definition 11 over \(I.\)

We may refer to a submodule of constant rank \(l\) as a formal vector bundle over \(M.\) This is also highlighted by the following Lemma, which can be thought of as a characterization of the bases of sections of a formal vector bundle.
Lemma 12. Suppose that $E$ is a submodule of $\mathbb{C}[Z,\zeta]^k$, with $\dim_{\mathbb{C}} E(0) = l$. Then $E$ is of constant rank $l$ over $I \subset \mathbb{C}[Z,\zeta]$ if and only if $E$ has the following property: If $v_1,\ldots,v_l \in E$ are vectors such that $v_1(0),\ldots,v_l(0)$ form a basis of $E(0)$, then $v_1,\ldots,v_l$ generate $E$ up to vectors all of whose components are elements of $I$.

Proof. The “if” direction is trivial; so assume that $E$ is of constant rank $l$. After reordering, we can assume that if we write $v_j = (v_{1j},\ldots,v_{kj})$, the matrix $(v_{mj})_{1 \leq m,n \leq l}$ is invertible. Hence, given any $e = (e^1,\ldots,e^k) \in E$, we can find $a_1,\ldots,a_l \in \mathbb{C}[Z,\zeta]$ such that $\sum_{m=1}^l a_nv_m^j = e^j$, $j = 1,\ldots,l$. Now consider $e' = e - \sum_j a_jv_j \in E$. We want to show that the components of this vector are elements of $I$; this is clear for the first $l$ components (which are 0, after all). Taking a subminor of length $l+1$ of the matrix

$$
\begin{pmatrix}
  v_{11} & \cdots & v_{1l} & \cdots & v_{1k} \\
  \vdots & \ddots & \vdots & & \vdots \\
  v_{l1} & \cdots & v_{ll} & \cdots & v_{lk} \\
  e_{11} & \cdots & e_{1l} & \cdots & e_{1k}
\end{pmatrix}
$$

which contains the first $l$ columns and developing it along the last row, by assumption we have that

$$
\pm \begin{vmatrix}
  v_{11} & \cdots & v_{1l} \\
  \vdots & & \vdots \\
  v_{11} & \cdots & v_{1l}
\end{vmatrix} e^{lj} \in I, \quad l+1 \leq j \leq k
$$

which implies $e^{lj} \in I$ since the determinant in (16) is a unit in $\mathbb{C}[Z,\zeta]$. \hfill \Box

2.3. Degeneracy of a formal holomorphic map. Let $H : M \to M'$ be a formal holomorphic map between the formal submanifolds $M \subset \mathbb{C}^N$ and $M' \subset \mathbb{C}^{N'}$. Choose a basis of formal CR-vector fields $L_1,\ldots,L_n$ tangent to $M$ and generators $\rho^l_1(Z',\zeta'),\ldots,\rho^d_1(Z',\zeta')$ of the ideal $I'$ representing $M'$. The formal series $\rho^l_1(H(Z),\overline{\Pi}(\zeta)),\ldots,\rho^d_1(H(Z),\overline{\Pi}(\zeta))$ are then elements of $I$ (since $H : M \to M'$). For $\rho^j_1(Z',\zeta')$, we denote by $\rho^j_{1,Z'}(Z',\zeta')$ the complex gradient

$$
\rho^j_{1,Z'}(Z',\zeta') = \left( \frac{\partial \rho^j_1(Z',\zeta')}{\partial Z'_1}, \ldots, \frac{\partial \rho^j_1(Z',\zeta')}{\partial Z'_{N'}} \right)
$$

and we usually think of it as a row vector. So $\rho^j_{1,Z'}(Z',\zeta') \in \mathbb{C}[Z',\zeta']^{N'}$, and $\rho^j_{1,Z'}(H(Z),\overline{\Pi}(\zeta)) \in \mathbb{C}[Z,\zeta][N']$. We define an ascending chain of submodules $E_k \subset \mathbb{C}[Z,\zeta][N']$ by

$$
E_k = \text{span}_{\mathbb{C}[Z,\zeta]}\{L_1 \cdots L_r \rho^j_{1,Z'}(H(Z),\overline{\Pi}(\zeta)) : \quad L_1,\ldots,L_r \text{ formal CR-vector fields tangent to } M, 0 \leq r \leq k, 1 \leq j \leq d' \}.
$$
The chain \( E_k(0) \) of subspaces of \( \mathbb{C}^{N'} \) will stabilize, say at the index \( k_0 \), that is, \( E_k(0) = E_{k_0}(0) \) for \( k \geq k_0 \), and \( E_{k_0-1}(0) \neq E_{k_0}(0) \). So will the chain of submodules \( E_k \subset \mathbb{C}[Z,\zeta]^{N'} \), since \( \mathbb{C}[Z,\zeta]^{N'} \) as a free module over a Noetherian ring is Noetherian itself; say \( E = \bigcup_k E_k \), and there is some \( k'_0 \geq k_0 \) such that \( E_k = E \) for \( k \geq k'_0 \).

**Definition 13.** With \( E_{k_0} \) defined above, let \( s = N' - \dim_{\mathbb{C}} E_{k_0}(0) \). We then say that \( H \) has formal degeneracy \( (k_0, s) \) (at 0) or shortly, that \( H \) is \((k_0, s)\)-degenerate (at 0). If \( E \) is of constant rank \( N' - s \) over \( I \), we say that \( H \) has constant degeneracy \( s \) and that \( H \) is constantly \((k_0, s)\)-degenerate. Furthermore, if \( s = 0 \), we say that \( H \) is \( k_0 \)-nondegenerate. Since in that case, \( E = \mathbb{C}[Z,\zeta]^{N'} \), \( H \) has automatically constant degeneracy 0.

Note that if \( H \) is of constant degeneracy, then the submodule \( E_{k_0} \) will actually generate \( E \) up to vectors whose components are in \( I \) (this follows from matrix manipulations, see Lemma 12). However, in general, we do not know whether \( k_0 = k'_0 \). Clearly, the degeneracy \( s \) fulfills the inequality \( 0 \leq s \leq N' - d' \). Without further restrictions on \( M, M' \) and \( H \), this is the best we can hope for, as the example of a Levi-flat submanifold as the target shows: If \( M' \) is defined by the equations \( \text{Im} w'_1 = \cdots = \text{Im} w'_d = 0 \), where \( Z' = (z'_1, w') = (z'_1, \ldots, z'_n, w'_1, \ldots, w'_d) \) are coordinates in \( \mathbb{C}^{N'} \), then every map \( H \) which maps into this manifold has constant degeneracy \( N' - d' \).

We will now show that Definition 13 is actually independent of choices of formal coordinates and generators. First consider a different set of generators \( \tilde{\rho}' = (\tilde{\rho}'_1, \ldots, \tilde{\rho}'_{d'}) \). Then there is an invertible \( d' \times d' \) matrix \( A = (a_{jk}) \) of formal power series in \( \mathbb{C}[Z,\zeta] \) such that \( \tilde{\rho}'(Z', \zeta') = A(Z', \zeta') \rho(Z', \zeta') \).

Taking the complex gradient, we obtain

\[
(19) \quad \tilde{\rho}'_{j, Z'} = \sum_{k=1}^{d'} \rho'_k a_{j,k, Z'} + \sum_{k=1}^{d'} a_{j,k} \rho'_{k, Z'}, \quad 1 \leq j \leq d'.
\]

Now the first sum in (19) is a vector whose entries are elements of \( I' \). We write \( \phi_j(Z, \zeta) = \rho'_j Z \langle H(Z), \overline{H}(\zeta) \rangle, \phi = (\phi_1, \ldots, \phi_{d'}) \), use the same notation for \( \tilde{\rho}' Z', \) and set \( B(Z, \zeta) = A(H(Z), \overline{H}(\zeta)); \) note that \( B \) is an invertible \( d' \times d' \) matrix of formal power series in \( \mathbb{C}[Z,\zeta] \). Then pulling (19) to \( M \) via \( H \), we see that

\[
(20) \quad \tilde{\phi}_j = v_j + B \phi, \quad 1 \leq j \leq d',
\]

where \( v_j \) is a vector in \( I \mathbb{C}[Z,\zeta]^{N'} \). (20) implies that \( \tilde{E}_k = E_k \) modulo \( I \mathbb{C}[Z,\zeta]^{N'} \) for each \( k \), which implies that \( \dim \tilde{E}_k(0) = \dim E_k(0) = e \) and that \( \tilde{E}_k \) is of constant rank \( e \) if and only if \( E_k \) is. This shows that the choice of defining function does not matter.
Note that Definition 13 is independent of the choice of formal holomorphic coordinates in \( \mathbb{C}^N \); that follows easily from the fact that such a biholomorphic change of coordinates \( F \) pushes formal CR-vector fields tangent to \( M \) to formal CR-vector fields tangent to \( F(M) \). The independence from choice of formal holomorphic coordinates in \( \mathbb{C}^{N'} \) is proved in the next Lemma.

**Lemma 14.** Suppose \( M \subset \mathbb{C}^N \), \( M' \subset \mathbb{C}^{N'} \) are generic formal submanifolds of \( \mathbb{C}^N \) and \( \mathbb{C}^{N'} \), respectively, and that \( H : M \to M' \) is a formal holomorphic map. Fix a set of generators of \( I' \subset \mathbb{C}[Z, \zeta] \), and let \( \bar{Z}' = F(Z') \) be a formal holomorphic change of coordinates in \( \mathbb{C}^{N'} \). Fix a set of generators of the ideal \( \bar{I}' \) representing \( M' \) in the \( \bar{Z}' \)-variables. Denote the space defined by (18) in the \( \bar{Z}' \)-variables by \( \bar{E}_k \), then

\[
\bar{E}_k \left( \frac{\partial F}{\partial Z'}(H(Z)) \right) = E_k
\]

modulo \( I\mathbb{C}[Z, \zeta]^{N'} \).

**Proof.** Let \( \bar{\rho}' \) be the fixed generators for the ideal \( \bar{I}' \) representing \( M' \) in the coordinates \( \bar{Z}' \). Then we can choose \( \rho' = F^* \bar{\rho}' \) as generators for \( I' \subset \mathbb{C}[Z, \zeta] \). We now take the complex gradient, and use the chain rule to obtain

\[
\rho'_{j,Z}(Z', \zeta') = (\bar{\rho}_j(F(Z'), \bar{F}(\zeta')) \big|_{Z'} = \bar{\rho}_{j,Z'}(F(Z'), \bar{F}(\zeta')) \frac{\partial F}{\partial Z'}(Z').
\]

Pulling (22) to \( M \) and applying CR-vector fields tangent to \( M \), we obtain (21), since all (0,1)-vector fields annihilate the entries of the matrix \( \frac{\partial F}{\partial Z'}(H(Z)) \). We have already shown above that if we choose different generators for \( I' \) in the same variables, the spaces \( E_k \) are equal modulo \( I\mathbb{C}[Z, \zeta]^{N'} \). \( \square \)

Now the transformation \( v = (v_1, \ldots, v_{N'}) \mapsto vA \), where \( A \) is an invertible \( N' \times N' \) matrix of formal power series in \( (Z, \zeta) \), is a module isomorphism of \( \mathbb{C}[Z, \zeta]^{N'} \) which maps \( I\mathbb{C}[Z, \zeta]^{N'} \) onto itself. So Definition 13 is in fact independent of all the choices made there. The next Lemma, the proof of which we leave to the reader, gives a means of actually computing with Definition 13.

**Lemma 15.** Suppose \( M \subset \mathbb{C}^N \), \( M' \subset \mathbb{C}^{N'} \) are generic formal submanifolds of \( \mathbb{C}^N \) and \( \mathbb{C}^{N'} \), respectively, and that \( H : M \to M' \) is a formal holomorphic map. Let \( L_1, \ldots, L_n \) be a basis of the CR-vector fields tangent to \( M \) (that is, a set of generators of \( \mathcal{D}_M^{0,1}/I\mathcal{D}_M^{0,1} \), see 2.1.1). For a multi-index \( \alpha \in \mathbb{N}^n \), let \( L^\alpha = L_1^{\alpha_1} \cdots L_n^{\alpha_n} \). Furthermore, let generators \( \rho' = (\rho'_1, \ldots, \rho'_{d'}) \) of \( I' \) be chosen. Define the submodules

\[
F_k = \text{span}_{\mathbb{C}[Z, \zeta]} \{ L^\alpha \rho'_{j,Z}(H(Z), \bar{H}(\zeta)) : |\alpha| \leq k, 1 \leq j \leq d' \}
\]
of $\mathbb{C}[Z,\zeta]^N$. Let $E_k$ be defined by (18). Then $E_k(0) = F_k(0)$, and $E_k$ is of constant rank over $I$ if and only if $F_k$ is; hence, in order to determine the degeneracy of $H$, it suffices to consider the $F_k$.

We now want to give a different characterization of the degeneracy $s$ in the case of constant degeneracy. For that, we will formulate the conditions of Definition 13 in terms of formal normal coordinates (see 2.1.2). So $I'$ is generated by the functions $\rho'_j = w'_j - Q'_j(z', \chi', \tau')$, $j = 1, \ldots, d'$, and we will write $H = (f, g) = (f_1, \ldots, f_n', g_1, \ldots, g_{d'})$ for $H$ in these normal coordinates. The complex gradient is easily computed to be

$$\rho'_{j, z'} = (-Q'_{j, z'_1}, \ldots, -Q'_{j, z'_{d'}}, e_j), \quad 1 \leq j \leq d', \tag{24}$$

where $e_j$ is the $j$th unit vector in $\mathbb{C}^{d'}$. In particular, the last $d'$ entries of any CR-derivative of length bigger than 0 of any $\rho'_{j, z'}(H(Z), \Pi(\zeta))$ will be 0. We will write $Q'_{j, x'_k}(f(z, w), \bar{f}(\chi, \tau), \bar{g}(\chi, \tau)) = \phi^k_j(Z, \zeta)$. So $H$ is of constant degeneracy $s$ if and only if (after possibly reordering the $z'$ variables) there exist $t = n' - s$ multi-indices $\alpha^1, \ldots, \alpha^t \in \mathbb{N}^n$ and integers $l_1, \ldots, l_t$, $1 \leq l_j \leq d'$, such that the vectors $(L^{\alpha^j_1} \phi^1_{l_1}, \ldots, L^{\alpha^j_t} \phi^t_{l_t})$, $1 \leq j \leq t$, evaluated at 0 form a basis of $\mathbb{C}^t$, and for all multi-indices $\beta \in \mathbb{N}^n$, all $k$, $t + 1 \leq k \leq n'$, and all $l$, $1 \leq l \leq d'$, the determinant

$$\begin{vmatrix}
L^{\alpha^1_t} \phi^1_{l_1} & \ldots & L^{\alpha^1_t} \phi^t_{l_t} & L^{\alpha^1_t} \phi^t_{l_t} \\
\vdots & \ddots & \vdots & \vdots \\
L^{\alpha^t_t} \phi^1_{l_1} & \ldots & L^{\alpha^t_t} \phi^t_{l_t} & L^{\alpha^t_t} \phi^t_{l_t} \\
L^{\beta} \phi^1_{l_1} & \ldots & L^{\beta} \phi^t_{l_t} & L^{\beta} \phi^t_{l_t}
\end{vmatrix} \in I. \tag{25}$$

More specifically, if $H$ is constantly $(k_0, s)$-degenerate, one of the $\alpha_t$ must have length $k_0$. We can use this to formulate the following technical result:

**Lemma 16.** Assume that normal coordinates $(z, w)$ and $(z', w')$ have been chosen for $M$ and $M'$, respectively, that $L_1, \ldots, L_n$ is a basis of the (formal) CR-vector fields tangent to $M$ and that $w'_j - Q'_j(z', \chi', \tau')$ are generators of $I'$ as in 2.1.2. Let $H : M \to M'$ be of constant degeneracy $s$, let $t = n' - s$, and write $H^2 Q'_{j, x'_k} = \phi^k_j$. We can choose $t$ multi-indices $\alpha^1, \ldots, \alpha^t$, and integers $l_1, \ldots, l_t$, $1 \leq l_j \leq d'$, such that (after possibly reordering the $z'$ variables) the vectors $(L^{\alpha^j_1} \phi^1_{l_1}, \ldots, L^{\alpha^j_t} \phi^t_{l_t})$, $1 \leq j \leq t$, evaluated at 0, form a basis of $\mathbb{C}^t$. Then

$$\Delta \phi^k_t - \sum_{m=1}^t \Delta_{m_k} \phi^m_t \in I, \quad 1 \leq l \leq d', \quad t + 1 \leq k \leq n'. \tag{26}$$
where

\[ \Delta(z, w) = \begin{vmatrix} L^{\alpha_1^1} \phi^1_{l_1} & \cdots & L^{\alpha_1^t} \phi^t_{l_1} \\ \vdots & \ddots & \vdots \\ L^{\alpha_t^1} \phi^1_{l_t} & \cdots & L^{\alpha_t^t} \phi^t_{l_t} \end{vmatrix} (z, w, 0, w), \quad \Delta(0) \neq 0, \]

and

\[ \Delta_{mk}(z, w) = (-1)^{t+m} \begin{vmatrix} L^{\alpha_1^i} \phi^1_{l_1} & \cdots & L^{\alpha_1^m} \phi^m_{l_1} & L^{\alpha_1^1} = \phi^1_{l_1} \\ \vdots & \ddots & \vdots & \vdots \\ L^{\alpha_t^i} \phi^1_{l_t} & \cdots & L^{\alpha_t^m} \phi^m_{l_t} & L^{\alpha_t^1} = \phi^1_{l_t} \end{vmatrix} (z, w, 0, w), \]

and where the \(^\sim\) means that this column has been dropped. More specifically, if \(H\) is constantly \((k_0, s)\)-degenerate, then the \(\alpha_j\) can be chosen to fulfill \(1 \leq |\alpha_j| \leq k_0\), and the same choice of \(\alpha_j\) is possible for every map \(H' : M \to M'\) (of constant degeneracy) agreeing with \(H\) up to order \(k_0\).

**Proof.** We will be using the parametrization of \(M\) as in 2.1.2. Note that by (7), for a formal series \(\phi(z, w, \chi, \tau) \in I, \phi(z, w, 0, w) = 0\). We use this in (25). Developing the resulting determinant along the last row, we see that for every \(\beta \in \mathbb{N}^n\), for every \(l, 1 \leq l \leq d'\), and for every \(k, t + 1 \leq k \leq n'\)

\[ \Delta(z, w)L^\beta \phi^k(z, w, 0, w) - \sum_{m=1}^{t} \Delta_{mk}(z, w)L^\beta \phi^m(z, w, 0, w) = 0, \]

where \(\Delta\) and \(\Delta_{mk}\) are defined by (27) and (28), respectively. Recalling (12) we conclude that

\[ \Delta(z, w)\phi^k(z, w, \chi, \overline{Q}(\chi, z, w)) - \sum_{m=1}^{t} \Delta_{mk}(z, w)\phi^m(z, w, \chi, \overline{Q}(\chi, z, w)) = 0. \]

This immediately implies (26). The last statement follows from the construction. \(\square\)

We are now going to characterize the degeneracy \(s\) of a mapping of constant degeneracy in terms of certain formal holomorphic vector fields. These results generalize some results about holomorphic nondegeneracy (defined in [15]) which can be found in e.g., [5].

**Definition 17.** Let \(M \subset \mathbb{C}^N, M' \subset \mathbb{C}^{N'}\) be formal submanifolds of \(\mathbb{C}^N\) and \(\mathbb{C}^{N'}\), respectively, \(H : M \to M'\) a formal holomorphic map. A formal
holomorphic vector field $X$ in $\mathbb{C}^{N'}$ tangent to $M'$ along $H(M)$ is an operator of the form

\begin{equation}
X = \sum_{j=1}^{N'} a_j(Z) \frac{\partial}{\partial Z_j'},
\end{equation}

where $a_j \in \mathbb{C}[Z]$ (called the coefficients of $X$), with the property that for every $\phi \in \mathcal{I}'$, $\sum_{j=1}^{N'} a_j(Z) \phi Z_j' (H(Z), \overline{H}(\zeta)) \in \mathcal{I}$. We say that a set $\{X_1, \ldots, X_l\}$ of such formal holomorphic vector fields is linearly independent if their coefficients evaluated at 0 form a linearly independent set of vectors in $\mathbb{C}^{N'}$.

If $H$ is an immersion, one can associate to $X$ as in Definition 17 a formal holomorphic vector field in $\mathbb{C}^{N'}$ in the following way. $H(M)$ can be regarded as a formal submanifold (not generic) of $\mathbb{C}^{N'}$ by taking as its ideal $(H^2)^{-1}(I) \supset \mathcal{I}'$. Since $H$ is immersive, $H$ has a right inverse $G$, so $H^\sharp$ has a left inverse $G^\flat$. Hence, an expression of the form $X = \sum_{j=1}^{N'} a_j(Z) \frac{\partial}{\partial Z_j}$ gives rise to a formal holomorphic vector field $X' = \sum_{j=1}^{N'} G^\flat a_j(Z') \frac{\partial}{\partial Z_j'}$ in $\mathbb{C}^{N'}$, and $X$ is tangent to $M'$ along $H(M)$ if and only if $H^\sharp(X'f) \in \mathcal{I}$ for all $f \in \mathcal{I}'$.

Also note that it is enough to check the condition in Definition 17 on a set of generators of $\mathcal{I}$, and that the space of all holomorphic vector fields tangent to $M'$ along $H(M)$ can be identified with a submodule of the free module $\mathbb{C}[Z]^{N'}$.

**Example 2.** Consider the standard linear injection of the boundary of the $N$-ball $|Z_1|^2 + \cdots + |Z_N|^2 = 1$ into the boundary of the $N'$-ball $|Z_1|^2 + \cdots + |Z_{N'}|^2 = 1$, $(Z_1, \ldots, Z_N) \mapsto (Z_1, \ldots, Z_N, 0, \ldots, 0)$. This map is of constant degeneracy $N' - N$ everywhere and there are $N' - N$ linearly independent holomorphic vector fields tangent to $M'$ along $H(M)$ given by $\frac{\partial}{\partial Z_{N+1}'}, \ldots, \frac{\partial}{\partial Z_{N'}'}$.

The situation in Example 2 is typical in the following sense:

**Proposition 18.** Assume that $M \subset \mathbb{C}^N$, $M' \subset \mathbb{C}^{N'}$ are generic formal submanifolds of $\mathbb{C}^N$ and $\mathbb{C}^{N'}$, respectively, and $H : M \to M'$ is of constant degeneracy $s$. Then

\begin{equation}
s = \dim_{\mathbb{C}} \{X(0) : X \text{ formal holomorphic vector field tangent to } M' \text{ along } H(M)\}.
\end{equation}

**Proof.** Let $\tilde{s}$ denote the dimension of the space on the right hand side of (32). From Lemma 16 we see that there are at least $s$ linearly independent holomorphic vector fields along $H$ tangent to $M'$ along $H(M)$, so that $s \leq \tilde{s}$. In fact, choosing local holomorphic coordinates $(z, w)$ for $M$, $(z', w')$ for $M'$
and a defining function $Q' = (Q'_1, \ldots, Q'_{n'})$ for $M'$ as in Lemma 16, we have that the formal holomorphic vector fields

$$X_k = \frac{\partial}{\partial z'_k} - \sum_{m=1}^{t} a_{m,k}(z,w) \frac{\partial}{\partial z'_m}, \quad k = t + 1, \ldots, n',$$

are tangent to $M'$ along $H(M)$ by (26) where $a_{m,k}(z,w) = \Delta_{mk}(z,w)/\Delta(z,w)$ and $\Delta_{mk}$ and $\Delta$ are defined by (28) and (27), respectively.

On the other hand, assume that $\{X_1, \ldots, X_{d'}\}$ are linearly independent holomorphic vector fields tangent to $M'$ along $H(M)$; say $X_k = \sum_{j=1}^{d'} a_{j,k}(Z) \frac{\partial}{\partial Z'_j}$, and let $\rho' = (\rho'_1, \ldots, \rho'_{d'})$ be a set of generators for $I'$. We have that

$$\sum_{j=1}^{d'} a_{j,k}(Z) \rho'_{l,Z'_j}(H(Z),\overline{H}(\zeta)) \in I, \quad 1 \leq l \leq d'.$$

Applying CR-vector fields $L_1, \ldots, L_r$ tangent to $M$ to (34), we see that

$$\sum_{j=1}^{d'} a_{j,k}(Z) L_1 \cdots L_r \rho'_{l,Z'_j}(H(Z),\overline{H}(\zeta)) \in I, \quad 1 \leq l \leq d'.$$

Evaluating (35) at 0, we conclude that $\dim_C E_k(0) \leq N' - \tilde{s}$ for all $k$. Hence, $\tilde{s} \leq s$, and the proof is complete.

Note that the second part of the proof of Proposition 18 shows that if we denote the dimension of the space on the right hand side of (32) by $\tilde{s}$, then the degeneracy $s$ of a formal map $H$ always satisfies $\tilde{s} \leq s$, whether the degeneracy is constant or not.

We now want to relate our notion of nondegeneracy of a map with the notion of finite nondegeneracy of manifolds. In particular, we give a bound on the degeneracy for a certain class of maps between finitely nondegenerate manifolds.

2.4. Finitely nondegenerate manifolds. The notion of finite nondegeneracy was introduced for hypersurfaces in [6], and has been used extensively in the study of mapping problems. We say that a generic submanifold is finitely nondegenerate (or, more specifically, $\ell_0$-nondegenerate) if its identity map is $\ell_0$-nondegenerate in the sense of Definition 13. For the original definition, see e.g., [5], Chapter IX. By the chain rule we see that if there is a $k_0$-nondegenerate map into some generic formal submanifold $M' \subset \mathbb{C}^{N'}$, then $M'$ is $\ell_0$-nondegenerate for some $\ell_0 \leq k_0$. In fact, we also see that every formal biholomorphism between generic formal submanifolds $M \subset \mathbb{C}^{N}$ and $M' \subset \mathbb{C}^{N'}$ of the same codimension which are $\ell_0$-nondegenerate is in fact $\ell_0$-nondegenerate.
In order to use finite nondegeneracy of submanifolds to put bounds on the
degeneracy, we need the mapping to fulfill another crucial property, which
we will introduce next.

**Definition 19.** Let \( M \subset \mathbb{C}^N, M' \subset \mathbb{C}^{N'} \) be formal generic submanifolds
in \( \mathbb{C}^N \) and \( \mathbb{C}^{N'} \), respectively, \( H : M \to M' \) a formal holomorphic map.
We say that \( H \) is transversal if for one (and hence every) set of generators
\( \rho' = (\rho'_1, \ldots, \rho'_{d'}) \) of \( I' \), \( H^2\rho' \) generates \( I \).

Note that in particular, if \( H : M \to M' \) is transversal, then \( d \leq d' \).

**Lemma 20.** Let \( M \subset \mathbb{C}^N, M' \subset \mathbb{C}^{N'} \) be formal generic submanifolds of \( \mathbb{C}^N \)
and \( \mathbb{C}^{N'} \), respectively, with \( M \) being \( \ell_0 \)-nondegenerate, and \( H : M \to M' \) a
transversal mapping. Then the degeneracy \( s \) of \( H \) fulfills \( 0 \leq s \leq N' - N \).

*Proof.* Assume that \( \rho' = (\rho'_1, \ldots, \rho'_{d'}) \) generates \( I' \). Without loss of general-
ity, assume that \( (H^2\rho'_1, \ldots, H^2\rho'_{d'}) \) generate \( I \). Now using the chain rule it
follows that

\[
(\rho'_j(H(Z), \overline{\Pi}(\zeta)))_Z = \rho'_{j,Z'}(H(Z), \overline{\Pi}(\zeta)) \frac{\partial H}{\partial Z_j}(Z), \quad j = 1, \ldots, d.
\]

Applying CR-vector fields \( L_1, \ldots, L_r \) tangent to \( M \) to (36), we see that

\[
L_1 \cdots L_r (\rho'_j(H(Z), \overline{\Pi}(\zeta)))_Z =
(L_1 \cdots L_r \rho'_{j,Z'}(H(Z), \overline{\Pi}(\zeta))) \frac{\partial H}{\partial Z_j}(Z), \quad j = 1, \ldots, d.
\]

By hypothesis, if evaluated at 0, the dimension of the space spanned by the
vectors on the right hand side of (37) is \( N \). On the other hand, the span of
the vectors \( L_1 \cdots L_r \rho'_{j,Z'}(H(Z), \overline{\Pi}(\zeta)) \) evaluated at 0 has dimension \( N' - s \),
where \( s \) is the degeneracy of \( H \). Hence, (37) implies that \( N \leq N' - s \), which
is the inequality claimed. \( \square \)

### 2.5. Real-analytic and smooth submanifolds

We now want to apply the theory developed above to smooth submanifolds of \( \mathbb{C}^N \) and \( \mathbb{C}^{N'} \). First,
let \( M \subset \mathbb{C}^N \) and \( M' \subset \mathbb{C}^{N'} \) be generic \( C^\infty \)-submanifolds of \( \mathbb{C}^N \) and \( \mathbb{C}^{N'} \)
of codimension \( d \) and \( d' \), respectively. Assume that \( p_0 \in M, p'_0 \in M' \),
and \( H \) is a holomorphic mapping (or, more generally, a \( C^\infty \)-CR-mapping)
defined in a neighbourhood \( U \) of \( p_0 \), with \( H(p_0) = p'_0 \) and \( H(U \cap M) \subset M' \).
We write \( \mathcal{V}(M) \) for the CR-bundle of \( M \), i.e., the bundle with \( \mathcal{V}(M)_p = \mathbb{C}T_p M \cap \mathbb{C}T^0 p(1) \mathbb{C}^N \).
To $M$ and $M'$ we associate formal submanifolds of $\mathbb{C}^N$ and $\mathbb{C}^{N'}$, respectively, by choosing holomorphic coordinates $Z$ and $Z'$ in $\mathbb{C}^N$ and $\mathbb{C}^{N'}$, respectively, in which $p_0 = 0$ and $p'_0 = 0$ and assigning them the ideals $I \subset \mathbb{C}[Z, \zeta]$ and $I' \subset \mathbb{C}[Z', \zeta']$ which are generated by the Taylor series of their defining functions. $H$ corresponds to a formal holomorphic map—by its Taylor expansion, if it is holomorphic, and by its formal holomorphic power series (see [5], §1.7.) if it is $C^\infty$-CR. Also, a local basis $L_1, \ldots, L_n$ of the CR-vector fields tangent to $M$ gives rise (by taking the Taylor expansion of the coefficients) to a basis for the formal CR-vector fields tangent to the formal manifold $M$.

Abusing notation, we shall always use the same letters to denote the formal object associated to a concrete object; this will cause no confusion, since the operations done on them clearly distinguish the two classes.

Choose defining functions $\rho' = (\rho'_1, \ldots, \rho'_{d'})$ for $M'$ and a local basis $L_1, \ldots, L_n$ for $C^\infty(M, \mathcal{V}(M))$. As above, for a multi-index $\alpha \in \mathbb{N}^n$, we write $L_\alpha = L_\alpha^1 \cdots L_\alpha^n$. After possibly shrinking $U$, we can define the vector subspaces

$$E'_k(p) = \text{span}_\mathbb{C}\{L_\alpha \rho'_j, Z'((H(Z), \overline{H(Z)})|_{Z=p}: |\alpha| \leq k, 1 \leq j \leq d'\} \subset \mathbb{C}^{N'}$$

for $p \in U$. Let $s(p) = N' - \max_k \dim_\mathbb{C} E'_k(p)$. We can then say that $H$ is of degeneracy $s(0)$ at $0$, and that $H$ is of constant degeneracy $s$ at $0$ if $s(p)$ is constant on a neighbourhood of $0$ in $M$. By taking $k_0$ to be the least integer for which $E'_k(0) = E'_{k_0}(0)$ for $k \geq k_0$, we can also define the finer invariant of $(k_0, s)$-degeneracy, like in Definition 13. Just as in the case of formal degeneracy, one sees that this definition is in fact independent of the choices made, and invariant under biholomorphic changes of coordinates in both $\mathbb{C}^N$ and $\mathbb{C}^{N'}$.

Finally, as noted above, the notion of $k_0$-nondegeneracy makes sense for mappings which are a priori only assumed to be $C^{k_0}$. This was used in [12] to prove a reflection principle, and is used in the statement of Corollary 4.

In the case of real-analytic submanifolds, we can give generic bounds on both $k_0$ and $s$ (under some additional assumptions), which we want to do now.

**Definition 21.** Let $M$ and $M'$ be connected, real-analytic, generic submanifolds of $\mathbb{C}^N$ and $\mathbb{C}^{N'}$, respectively, and $H$ a holomorphic mapping defined on an open set $U \subset \mathbb{C}^N$ containing $M$ with $H(M) \subset M'$. For all $p \in M$, let $s(p)$ be the degeneracy of $H$ at $p$. The generic degeneracy $s(H)$ is defined as $s(H) = \min_{p \in M} s(p)$.

The following Lemma implies that the set of points where $H$ is of constant degeneracy $s(H)$ is an open, dense subset of $M$.
Lemma 22. Let $M$, $M'$, $H$ be as in Definition 21. The set $\{ p \in M : s(p) > s(H) \}$ is real-analytic.

Proof. After choosing local defining functions, the points where the degeneracy of $H$ is strictly bigger than $s(H)$ is given by the vanishing of determinants with real-analytic entries; see the arguments before (25). \hfill \Box

The inequality $0 \leq s(H) \leq N' - d'$ holds trivially. The upper bound corresponding to Lemma 20 is sharper:

Lemma 23. Let $M$, $M'$ and $H$ be as in Definition 21 and assume in addition that there exists a point $p_0' \in H(M)$ at which $M'$ is finitely nondegenerate and that $H$ is transversal. Then $0 \leq s(H) \leq N' - N$.

The proof is immediate from Lemma 20.

We are now going to derive a bound on $k_0$. Assume for simplicity that $0 \in M$, $0 \in M'$, and that $H(M) \subset M'$ with $H(0) = 0$. Also let normal coordinates $(z, w)$ for $M$ and $(z', w')$ for $M'$ with corresponding real-analytic functions $Q : \mathbb{C}^{2n+d} \to \mathbb{C}^d$ and $Q' : \mathbb{C}^{2n'+d'} \to \mathbb{C}^{d'}$ (each defined and convergent in a neighbourhood of $0 \in \mathbb{C}^{2n+d}$ and $0 \in \mathbb{C}^{2n'+d'}$) be chosen. That is, both $Q$ and $Q'$ fulfill (7), $M$ is given by $w = Q(z, z, w)$ in a neighbourhood of 0, and $M'$ is given by $w' = Q'(z', z', w')$. As in 2.3, we write $H = (f, g)$ and set $Q'_{j, z_k}(f(z, w), f(z, w), g(z, w)) = \phi^k_j(z, w, z, w)$. If $H$ is of constant degeneracy $s$ at 0, say $H$ is $(s, k_0)$-degenerate at 0, then after reordering we may assume that (writing $e = n' - s$) the vector valued functions

$$\phi_j = (\phi^1_j, \ldots, \phi^e_j, e_j), \quad 1 \leq j \leq d', \quad (39)$$

where $e_j$ is the $j$th unit vector in $\mathbb{C}^{d'}$, are real-analytic at $0 \in \mathbb{C}^N$; they are clearly linearly independent at 0, and furthermore, if we choose the basis of CR-vector fields tangent to $M$

$$L_k = \frac{\partial}{\partial z_k} + \sum_{j=1}^{d} Q'_{j, z_k}(z, z, w) \frac{\partial}{\partial z_j}, \quad 1 \leq k \leq d, \quad (40)$$

and let $L^\alpha = L^\alpha_1 \cdots L^\alpha_n$, then the set $\{ L^\alpha \phi_j \}_{0 : \alpha \in \mathbb{N}^n, 1 \leq j \leq d'}$ spans $\mathbb{C}^{N'-s}$. Now we can complexify all of these statements. So we let $\mathcal{M} = \{ (z, w, \chi, \tau) \in U \subset \mathbb{C}^n \times \mathbb{C}^d \times \mathbb{C}^n \times \mathbb{C}^d : \tau_j = Q'_{j, z_k}(z, z, w) \}$ where $U$ is a neighbourhood of $0$ on which $Q$ is convergent be the complexification of $M$; $\mathcal{M}$ is a holomorphic submanifold of codimension $d$ in $\mathbb{C}^{2N}$. We also need the submanifold $\mathcal{M}_0$ of dimension $n$ defined by $\mathcal{M}_0 = \{ (\chi, \tau) \in \mathbb{C}^n \times \mathbb{C}^d : \tau_j = Q'_{j, z_k}(\chi, 0, 0) \} = \{ (\chi, 0) : \chi \in \mathbb{C}^n \}$. The complexifications of the $L_k$ are

$$L_k = \frac{\partial}{\partial \chi_k} + \sum_{j=1}^{d} Q'_{j, \chi_k}(\chi, z, w) \frac{\partial}{\partial \tau_j}, \quad 1 \leq k \leq d, \quad (41)$$
and if we denote the complexification of \( \phi_j \) again by \( \phi_j \), then we have that the set \( \{ L^a \phi_j |_0 : \alpha \in \mathbb{N}^n, 1 \leq j \leq d' \} \), spans \( \mathbb{C}^{N'-s} \). Now note that we can restrict our attention to \( \mathcal{M}_0 \) in this statement, since none of the \( L_k \) differentiates in \( z \) or \( w \); hence, we have that

\[
\text{span}_{\mathbb{C}} \left\{ \partial^{\alpha} \phi_j(0, 0, \chi, 0) \bigg| \chi = 0, 1 \leq j \leq d' \right\} = \mathbb{C}^{N'-s}.
\]

We now apply e.g., Lemma 11.5.4. in [5] to conclude that generically, the derivatives of \( \phi_j(0, 0, \chi, 0) \) up to order \( N'-s-d' \) span \( \mathbb{C}^{N'-s} \); which in turn implies that generically, \( k_0 \leq N'-d'-s \). We summarize:

**Lemma 24.** Let \( M \) and \( M' \) be connected, real-analytic, generic submanifolds of \( \mathbb{C}^N \) and \( \mathbb{C}^{N'} \), respectively, and \( H \) a holomorphic mapping defined on an open set \( U \subset \mathbb{C}^{N} \) containing \( M \) with \( H(M) \subset M' \). Then there exist numbers \( s \leq N'-d' \) and \( k_0 \leq N'-d'-s \) such that outside some proper real analytic subvariety of \( M \), \( H \) is \((k_0, s)\)-degenerate.

### 3. Nondegenerate mappings.

In this section we shall discuss nondegenerate mappings. We start with the “basic identity”, and in the next subsection, prove Theorems 1, 2, and 5.

#### 3.1. The basic identity.

We write \( K(t) = \{|\alpha| : \alpha \in \mathbb{N}^N, |\alpha| \leq t \} \) for the number of all multi-indices of length less than \( t \). For a multi-index \( \alpha \), \( \partial^\alpha \) denotes the operator \( \partial^{\alpha} \). The following proposition is our starting point.

**Proposition 25** (Basic identity for nondegenerate maps). Let \( M \subset \mathbb{C}^N \), \( M' \subset \mathbb{C}^{N'} \) be generic formal submanifolds, \( H : M \to M' \) a formal holomorphic map which is \( k_0\)-nondegenerate. Then there exists a formal function \( \Psi : \mathbb{C}^N \times \mathbb{C}^N \times \mathbb{C}^{K(k_0)N'} \to \mathbb{C}^{N'} \) (that is, if we write \( W \) for the coordinates in \( \mathbb{C}^{K(k_0)N'} \), \( \Psi \in \mathbb{C}[Z, \zeta, W]^{N'} \)) with the property that

\[
H(Z) - \Psi(Z, \zeta, (\partial^\beta \overline{H}(\zeta) - \partial^\beta \overline{H}(0))_{0 \leq |\beta| \leq k_0} ) \in I^{N'} = I \times \cdots \times I, \quad \text{N' times}
\]

furthermore, \( \Psi \) depends only on \( M, M' \), and on the values of \( \partial^\beta H(0) \) for \( \beta \leq |k_0| \), such that if \( H' : M \to M' \) is another formal map with \( \partial^\beta H(0) = \partial^\beta H'(0) \) for \( |\beta| \leq k_0 \), then (43) holds with \( H' \) in place of \( H \). If \( M \) and \( M' \) are real-analytic, \( \Psi \) is convergent on a neighbourhood of the origin. If \( M \) and \( M' \) are algebraic, so is \( \Psi \).

**Proof.** Choose a basis \( L_1, \ldots, L_n \) of the CR-vector fields tangent to \( M \) and defining functions \( \rho' = (\rho'_1, \ldots, \rho'_{d'}) \). By Lemma 15, we can choose \( N' \) multi-indices \( \alpha^1, \ldots, \alpha^{N'} \) and integers \( l^1, \ldots, l^{N'} \) with \( 0 \leq |\alpha^j| \leq k_0, 1 \leq l^j \leq d' \) for
all \( j = 1, \ldots, N' \) such that
\[
\det \left( \Lambda^j \rho_j \left| (H(Z), \Pi(\zeta)) \right|_0 \right)_{1 \leq j \leq N', \ 1 \leq k \leq N'} \neq 0.
\]

We write \( \Phi_j(Z, \zeta, H(Z), \Pi(\zeta)), (\partial^\beta \Pi(\zeta))_{1 \leq |\beta| \leq k_0} = \Lambda^j \rho_j(H(Z), \Pi(\zeta)) \in I \) for \( 1 \leq j \leq N' \); using the chain rule, we see that \( \Lambda^j \Phi_j \in \mathbb{C}[Z, \zeta, Z', \zeta'][[W]] \) where \( W \) are variables in \( \mathbb{C}^{(K(k_0)-1)N'} \). We make a change of variables by replacing \( W \) by \( W + \partial^\beta \Pi(0)_{1 \leq |\beta| \leq k_0} \) and write again \( \Phi_j \) in these new variables; hence, \( \Phi_j(Z, \zeta, H(Z), \Pi(\zeta)), (\partial^\beta \Pi(\zeta) - \partial^\beta \Pi(0))_{1 \leq |\beta| \leq k_0} \) \( \in I \), and \( \Phi_j \in \mathbb{C}[Z, \zeta, Z', \zeta'][W]; \) also, \( \Phi_j \) depends only on \( M, M' \), and on the values of \( \partial^\beta H(0) \) for \( |\beta| \leq k_0 \).

Now consider the equations
\[
\Phi_j(Z, \zeta, Z', \zeta', W) = 0, \quad 1 \leq j \leq N'.
\]
We claim that this family of equations has a unique solution in \( Z' \). In fact, if we compute the Jacobian of (45) with respect to \( Z' \) at 0, by the definition of \( \Phi_j \) and using (44), we see that the Jacobian matrix \( \left( \frac{\partial \Phi_j}{\partial Z_k}(0) \right) \) is nonsingular. It follows by the formal implicit function theorem that there exist \( N' \) unique formal power series \( \Psi_j \in \mathbb{C}[Z, \zeta, \zeta', W], j = 1, \ldots, N' \), with the property that \( \Phi_j(Z, \zeta, \Psi_1(Z, \zeta, \zeta', W), \ldots, \Psi_{N'}(Z, \zeta, \zeta', W), \zeta', W) = 0 \).

We recall that \( \Phi_j(Z, \zeta, H(Z), \Pi(\zeta)), (\partial^\beta \Pi(\zeta) - \partial^\beta \Pi(0))_{1 \leq |\beta| \leq k_0} \in I \); if we replace \( Z \) and \( \zeta \) by a parametrization (as, for example, in 2.1.2) of \( I \), say \( Z(x) \) and \( \zeta(x) \), we conclude that \( \Phi_j(Z(x), \zeta(x), H(Z(x)), (\partial^\beta \Pi(\zeta(x)) - \partial^\beta \Pi(0))_{1 \leq |\beta| \leq k_0} = 0 \). It follows that \( H_j(Z(x)) = \Psi_j(Z(x), \zeta(x), H(z(x)), (\partial^\beta \Pi(\zeta(x)) - \partial^\beta \Pi(0))_{1 \leq |\beta| \leq k_0}, 1 \leq j \leq N' \). Passing back to the ring \( \mathbb{C}[Z, \zeta] \), we conclude that \( \Psi = (\Psi_1, \ldots, \Psi_{N'}) \) fulfills (43).

By construction, the map \( \Phi \) depends only on \( M, M' \), and \( \partial^\beta H(0), 0 \leq |\beta| \leq k_0 \). The same choice of \( \alpha^1, \ldots, \alpha^{N'} \) and \( l^1, \ldots, l^{N'} \) works for every other map \( H' \) with \( \partial^\beta H'(0) = \partial^\beta H(0), |\beta| \leq k_0 \). Finally, if \( M \) and \( M' \) are real analytic or algebraic, we can choose the defining functions and the basis of CR-vector fields to be real-analytic (or algebraic, respectively) and the last two claims of Proposition 25 follow since those classes of maps are closed under application of the implicit function theorem.

We shall need some formal vector fields tangent to \( M \), which will help us to exploit (43). Let \( \rho = (\rho_1, \ldots, \rho_d) \) be a real-analytic defining function for \( M \). After renumbering, we may assume that \( \rho_\zeta = \left( \frac{\partial \rho_j}{\partial \zeta_k} \right)_{1 \leq j, k \leq d} \) is invertible; set
\[
S_j = \frac{\partial}{\partial Z_j} - \rho_{Z_j}(\rho_\zeta)^{-1} \frac{\partial}{\partial \zeta}, \quad j = 1, \ldots, N,
\]
where \( \frac{\partial}{\partial \zeta} = \left( \frac{\partial}{\partial z_1}, \ldots, \frac{\partial}{\partial z_d} \right)^t \). Then \( S_j \) is a (formal) vector field tangent to \( M \), and its coefficients are convergent, if \( M \) is assumed to be real-analytic, and algebraic functions if \( M \) is assumed to be algebraic. If for \( \alpha \in \mathbb{N}^N \) we write \( S^\alpha = S_1^{\alpha_1} \cdots S_d^{\alpha_d} \), then for \( H \in \mathbb{C}[Z] \), \( S^\alpha H = \partial^\alpha H \). Applying these vector fields repeatedly to (43) and using the chain rule we get the following Corollary to Proposition 25.

**Corollary 26.** Under the assumptions of Proposition 25, the following holds: For all \( \alpha \in \mathbb{N}^N \), there exists a formal function \( \Psi_\alpha : \mathbb{C}^N \times \mathbb{C}^N \times \mathbb{C}^{K(k_0+|\alpha|)N'} \rightarrow \mathbb{C}^{N'} \) which is polynomial in its last \( (K(k_0+|\alpha|) - K(k_0))N' \) entries such that

\[
(47) \quad \partial^\alpha H(Z) - \Psi_\alpha(Z, \zeta, (\partial^\beta \overline{\Pi}(\zeta) - \partial^\beta \overline{\Pi}(0))_{0 \leq |\beta| \leq k_0}, (\partial^\beta \overline{\Pi}(\zeta))_{k_0 < |\beta| \leq k_0 + |\alpha|}) \in I^{N'};
\]

\( \Psi_\alpha \) depends only on \( M, M' \), and on the values of \( \partial^\beta H(0) \) for \( |\beta| \leq k_0 \), such that if \( H' : M \rightarrow M' \) is another formal map with \( \partial^\beta H(0) = \partial^\beta H'(0) \) for \( |\beta| \leq k_0 \), then (47) holds with \( H' \) in place of \( H \). If \( M \) and \( M' \) are real-analytic, \( \Psi_\alpha \) is convergent. If \( M \) and \( M' \) are algebraic, so is \( \Psi_\alpha \).

The next step is to repeatedly use (47) on the Segre sets. Recall (13a) and (14). Hence, choosing normal coordinates \( Z = (z, w) \), \( \zeta = (\chi, \tau) \), we have that \( f(z, 0, 0, 0) = 0 \) for all \( f \in I \). Applying this fact to (47), we conclude that

\[
(48) \quad \partial^\alpha H(z, 0) = \Psi_\alpha(z, 0, 0, 0, (\partial^\beta \overline{\Pi}(0))_{k_0 < |\beta| \leq k_0 + |\alpha|}).
\]

Note that the evaluation occurring causes no problems, since by Corollary 26, \( \Psi_\alpha \) is a polynomial with respect to these variables. Hence the right hand side of (48) defines a formal map \( \mathbb{C}^n \rightarrow \mathbb{C}^{N'} \), is convergent if \( M \) and \( M' \) are real-analytic, and algebraic, if \( M \) and \( M' \) are algebraic. This is the case \( k = 0 \) of the following Corollary (we are using the notation introduced before (14)):

**Corollary 27.** For all \( \alpha \in \mathbb{N}^N \), there exists a formal function \( \Upsilon_{k, \alpha} : \mathbb{C}^{kn} \rightarrow \mathbb{C}^{N'} \) which depends only on \( M, M' \), and the derivatives \( \partial^\beta H(0) \) for \( |\beta| \leq (k+1)k_0 + |\alpha| \) such that

\[
(49) \quad \partial^\alpha H(v^{k+1}(z, \xi)) = \Upsilon_{k, \alpha}(z, \xi).
\]

The dependence on the derivatives is as in Proposition 25: If \( H' : M \rightarrow M' \) is another formal mapping with \( \partial^\beta H(0) = \partial^\beta H'(0) \) for \( |\beta| \leq (k+1)k_0 + |\alpha| \), then (49) holds with \( H' \) instead of \( H \). If \( M \) and \( M' \) are real-analytic, \( \Upsilon_{k, \alpha} \) is convergent on a neighbourhood of \( 0 \in \mathbb{C}^{kn} \). If \( M \) and \( M' \) are algebraic, so is \( \Upsilon_{k, \alpha} \).
Proof. We note that (48) is just the case $k = 0$. We are doing induction on $k$. Assume the Corollary holds for $k < k'$. By (47),

\begin{equation}
\partial^\alpha H(v^{k'+1}(z', \xi')) = \Psi_\alpha(v^{k'+1}(z', \xi'), \partial^\beta \bar{H}(\bar{v}^{k'}(\xi'))
- \partial^\beta \bar{H}(0))_{0 < |\beta| \leq k_0}, \quad (\partial^\beta(\bar{H}(\bar{v}^{k'}(\xi'))))_{k_0 < |\beta| \leq k_0 + |\alpha|}.
\end{equation}

Note that the compositions occurring on the right hand side are all well-defined. We now plug the induction hypothesis (49) for $k = k' - 1$ into (50).

In fact, conjugating (49) and replacing $(z, \xi)$ by $(\xi')$, we get that

\begin{equation}
\partial^\beta \bar{H}(\bar{v}^{k'}(\xi')) = \Upsilon_{k' - 1, \beta}(\xi').
\end{equation}

Now the highest order derivative we need is $|\beta| = k_0 + |\alpha|$, which by assumption depends on the derivatives of $H$ of order up to $(k' + 2)k_0 + |\alpha|$. This finishes the induction.

$\square$

3.2. Proof of Theorems 1, 2, and 5. We start with Theorem 1. We use Corollary 27 for $k = 2k_1 - 1$, where $k_1$ is the integer given by Theorem 10. Since the manifolds are assumed to be real-analytic, $\Upsilon_{2k_1, 0}$ is convergent in a neighbourhood of the origin. By Theorem 10, we can choose $(z_0, \xi_0)$ in this neighbourhood with $v^{2k_1}(z_0, \xi_0) = 0$ and such that the rank of $v^{2k_1}$ is $N$ at $(z_0, \xi_0)$. As in the remark after Theorem 10, this implies that there is a holomorphic function $\psi$ defined in a neighbourhood of $0 \in \mathbb{C}^N$ such that $\psi(0) = (z_0, \xi_0)$ and $v^{2k}(\psi(Z)) = Z$. Hence,

\begin{equation}
H(Z) = H(v^{2k}(\psi(Z))) = \Upsilon_{2k_1 - 1, 0}(\psi(Z)).
\end{equation}

Since the right hand side of (52) is convergent, so is the left hand side. This completes the proof of Theorem 1.

Now assume that $H$ is $C^\infty$-CR and that $M$ and $M'$ are smooth. Its associated formal holomorphic power series is then a formal holomorphic map between the formal submanifolds associated to $M$ and $M'$ (see the remarks in 2.5). We use Corollary 27 for $k = k_1 - 1$, where $k_1$ is the integer given by Theorem 10. Now set $K = k_1$. Then Corollary 27 implies that

\begin{equation}
H(v^{k_1}(z, \xi)) = \Upsilon_{k_1 - 1, 0}(z, \xi) = H'(v^{k_1}(z, \xi)).
\end{equation}

But $\text{rk}(v^k) = N$, which by e.g., Proposition 5.3.5. in [5] implies that $H = H'$ in the sense of equality of formal power series, which finishes the proof of Theorem 2.

Theorem 5 follows from Corollary 27 exactly like Theorem 1; we just note that it is enough to check that $H$ is algebraic on some open set $U$ containing the point $p_0$ where $H$ is assumed to be $k_0$-nondegenerate.

The case \( N' = N + 1 \).

In this section, we will assume that \( M \) and \( M' \) are hypersurfaces, i.e., \( d = d' = 1 \). In addition, we assume that they are Levi-nondegenerate (at our distinguished points). We start with a couple of general facts.

4.1. Levi-nondegeneracy. In normal coordinates, which we choose at our distinguished points \( p_0 \) and \( p'_0 \), \( M \) being Levi-nondegenerate means that we can assume

\[
Q_{z_j \chi_k}(0, 0, 0) = \delta_{jk} \epsilon_k, \quad 1 \leq k \leq n,
\]

(54)

where every \( \epsilon_k \) is either \( +i \) or \( -i \), and likewise for \( M' \). Here is an easy technical result about the pullback of the Levi form by a map \( H \) in normal coordinates, which we will use in the proof of Proposition 30.

**Lemma 28.** Let \( M \subset \mathbb{C}^N \), \( M' \subset \mathbb{C}^{N'} \) be given in normal coordinates by \( w = Q(z, z, w) \) and \( w' = Q(z', z', w') \), respectively. Assume that \( H = (f, g) : (\mathbb{C}^N, 0) \to (\mathbb{C}^{N'}, 0) \) is a formal holomorphic map, and \( H(M) \subset M' \). Then

\[
g_w(0)Q_{z_j z_k}(0) = \sum_{r,s=1}^{n'} Q'_{z'_r z'_s}(0) f_{r z_j}(0) f_{s z_k}(0), \quad 1 \leq j, k \leq n.
\]

(55)

To prove this, set in \( g(z, w) = Q'(f(z, w), f(\chi, \tau), g(\chi, \tau)) \) \( \tau = 0 \), \( w = Q(z, \chi, 0) \) to obtain \( g(z, Q(z, \chi, 0)) = Q'(f(z, Q(z, \chi, 0)), f(\chi, 0), g(\chi, 0)) \).

Differentiation with respect to \( z_j \) and \( \chi_k \) and evaluating at \( z = \chi = 0 \) yields (55). This has the following consequence:

**Corollary 29.** Suppose that \( M \subset \mathbb{C}^N \) and \( M' \subset \mathbb{C}^{N'} \) are (formal) real hypersurfaces which are Levi-nondegenerate at \( p_0 \) and \( p'_0 \), respectively, and \( H : (\mathbb{C}^N, p_0) \to (\mathbb{C}^{N'}, p'_0) \) is a (formal) holomorphic map which takes \( M \) into \( M' \) and is transversal at \( p_0 \). Then \( H \) is immersive.

This is easy to see using normal coordinates (which we shall choose in a way such that \( p = 0 \) and \( p' = 0 \)). Since \( g_{z \nu}(0) = 0 \), the differential of \( H \) has the following form:

\[
\partial H(0) = \begin{bmatrix}
    f_{1 z_1}(0) & \cdots & f_{1 z_n}(0) & f_{1 w}(0) \\
    \vdots & \ddots & \vdots \\
    f_{n' z_1}(0) & \cdots & f_{n' z_n}(0) & f_{n' w}(0) \\
    0 & \cdots & 0 & g_w(0)
\end{bmatrix}
\]

(56)

and \( H \) is immersive if this matrix has rank \( N \). Hence, if \( H \) is transversal, \( H \) is immersive if and only if the matrix

\[
\partial f(0) = \begin{bmatrix}
    f_{1 z_1}(0) & \cdots & f_{1 z_n}(0) \\
    \vdots & \ddots & \vdots \\
    f_{n' z_1}(0) & \cdots & f_{n' z_n}(0)
\end{bmatrix}
\]

(57)
has rank \( n \). But by (55),

\[
g_w(0) \left( Q_{z_j z_k}(0) \right)_{1 \leq j, k \leq n} = \partial f'(0)^{\xi} \left( Q'_{z'_s z'_r}(0) \right)_{1 \leq r, s \leq n'} \partial f(0).
\]

This implies that if \( H \) is transversal, the rank of \( \partial f(0) \) is at least \( n \), which proves the corollary.

4.2. The basic identity for 1-degenerate maps. From now on we shall assume that \( N' = N + 1 \). Note that in the Levi-nondegenerate case,

\[
L_k Q'_{z'_j} (f(z, w), f(z, w), g(z, w))(0) = \epsilon'_j f_{jz}(0), \quad 1 \leq j \leq n, \quad 1 \leq k \leq n + 1.
\]

By Lemma 20, if \( N' = N + 1 \) and \( H \) is transversal, the degeneracy \( s \) of \( H \) at \( p_0 \) is either 0 or 1. In the case \( s = 0 \), we can apply Theorem 1 and 2 to obtain Theorem 6 and 7, since by Theorem 10 we see that \( K \leq 2 \) if the source manifold is a hypersurface. Hence, from now on we will assume that \( s = 1 \). In this subsection, we will develop a basic identity for 1-degenerate maps between Levi-nondegenerate hypersurfaces. From (59) we see that in Lemma 16 we can choose \( \alpha^j \) to be the multi-index with a 1 in the \( i \)-th spot and 0 elsewhere and reorder the \( z' \)'s, to get that after barring (26),

\[
\Delta(\chi, \tau) Q'_{\chi_{n+1}} (\bar{f}(\chi, \tau), f(z, w), g(z, w)) - \sum_{m=1}^{n} \Delta_m(\chi, \tau) Q'_{\chi_m} (\bar{f}(\chi, \tau), f(z, w), g(z, w)) \in I,
\]

where now

\[
\Delta(z, w) = \begin{vmatrix}
L_1 Q'_{z'_1} (f, \bar{f}, \bar{g}) & \ldots & L_1 Q'_{z'_n} (f, \bar{f}, \bar{g}) \\
\ldots & \ldots & \ldots \\
L_n Q'_{z'_1} (f, \bar{f}, \bar{g}) & \ldots & L_n Q'_{z'_n} (f, \bar{f}, \bar{g})
\end{vmatrix}(z, w, 0, w),
\]

\[
\Delta(0) = \begin{vmatrix}
\epsilon'_1 f_{1z_1}(0) & \ldots & \epsilon'_n f_{nz}(0) \\
\vdots & \ddots & \vdots \\
\epsilon'_1 f_{1z_n}(0) & \ldots & \epsilon'_n f_{nz_n}(0)
\end{vmatrix} \neq 0,
\]

\[
\Delta_m(z, w) = \begin{vmatrix}
L_1 Q'_{z'_1} (f, \bar{f}, \bar{g}) & \ldots & L_1 Q'_{z'_m} (f, \bar{f}, \bar{g}) & \ldots & L_1 Q'_{z'_n} (f, \bar{f}, \bar{g}) \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
L_n Q'_{z'_1} (f, \bar{f}, \bar{g}) & \ldots & L_n Q'_{z'_m} (f, \bar{f}, \bar{g}) & \ldots & L_n Q'_{z'_n} (f, \bar{f}, \bar{g})
\end{vmatrix},
\]
this matrix evaluated at \((z, w, 0, w)\),

\[
\Delta_m(0) = (-1)^{n+m} \begin{vmatrix}
\epsilon_1 f_{1z_1}(0) & \cdots & \epsilon_m f_{mz_1}(0) & \cdots & \epsilon_n f_{nz_1}(0) \\
\vdots & & \vdots & & \vdots \\
\epsilon_1 f_{1z_n}(0) & \cdots & \epsilon_m f_{mz_n}(0) & \cdots & \epsilon_n f_{nz_n}(0)
\end{vmatrix},
\]

Since \(H\) maps \(M\) into \(M'\), the chain rule implies that we have formal functions \(\Phi_1, \ldots, \Phi_n\) such that \(\Phi_j(Z, \zeta, H(Z), H(\zeta), \partial H(\zeta)) = L_j Q'(f(z, w), \vec{f}(\chi, \tau), \vec{g}(\chi, \tau))\) which are convergent if \(M\) and \(M'\) are, and are polynomial in the derivatives of \(H\). As in the proof of Proposition 25 we obtain functions \(\Phi_j \in \mathbb{C}[Z, \zeta, Z', \zeta', Z, \zeta]\) where \(W \in \mathbb{C}^{(N+1)^2}\), such that

\[
\Phi_j(Z, \zeta, H(Z), H(\zeta), H(0, \tau), \partial H(0, \tau) - \partial H(0)) \in I.
\]

From (60) we conclude that after a change of variables, we can write the function given there as \(\Upsilon \in \mathbb{C}[Z, \zeta, Z', \zeta', T, U, W']\) such that

\[
\Upsilon(Z, \zeta, H(Z), H(\zeta), H(0, \tau), \partial H(0, \tau) - \partial H(0)) \in I.
\]

\(\Phi_1, \ldots, \Phi_n, \Upsilon\) only depend on \(M, M'\) and the derivative of \(H\) evaluated at 0, will agree for \(H\) and \(H'\) with \(\partial H(0) = \partial H'(0)\), are convergent if \(M\) and \(M'\) are real-analytic, and algebraic if \(M\) and \(M'\) are algebraic. Consider the system of equations

\[
\begin{align*}
\Phi_j(Z, \zeta, Z', \zeta', W) &= 0, \quad 1 \leq j \leq n, \\
\Upsilon(Z, \zeta, Z', \zeta', T, U, W') &= 0, \\
Z_{N+1}' &= Q'(Z_1', \ldots, Z_N', = \zeta'),
\end{align*}
\]

in \(\mathbb{C}[Z, \zeta, Z', \zeta', T, U, W, W']\). We claim that we can apply the implicit function theorem to (67) to see that this system admits a unique solution in \(Z'\).

In order to compute the Jacobian of (67) with respect to \(Z'\), first note that since \(\Upsilon_{Z_{N+1}'}(0, 0, 0, 0, 0) = 0\) and \(Z_{N+1}'\) does not appear in any of the \(\Phi_j\), it is enough to show that the determinant

\[
D = \begin{vmatrix}
\Phi_{1, Z_1'}(0) & \cdots & \Phi_{1, Z_{N'}'}(0) \\
\vdots & & \vdots \\
\Phi_{n, Z_1'}(0) & \cdots & \Phi_{n, Z_{N'}'}(0) \\
\Upsilon_{Z_1'}(0) & \cdots & \Upsilon_{Z_{N'}'}(0)
\end{vmatrix}
\]

is nonzero. Note that \(\Phi_{k, Z_j'} = \epsilon_j f_{jz_k}(0)\), that \(\Upsilon_{Z_j'}(0) = -\epsilon_j \Delta_j(0)\) for \(1 \leq j \leq n\) and \(\Upsilon_{Z_N'}(0) = \epsilon_N \Delta_N(0)\). To simplify notation in the following argument, we write \(\Delta(0) = -\Delta_N(0)\). Developing \(D\) along the last row and
using (62) and (64) we see that

\[
D = \pm \begin{pmatrix}
\epsilon'_1 & \ldots & 0 & \ldots & 0 \\
0 & \epsilon'_2 & \ldots & 0 & \ldots \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \ldots & \epsilon'_{n'} & \ldots \\
0 & 0 & \ldots & 0 & \epsilon'_{n'}
\end{pmatrix}
\begin{pmatrix}
f_{2z_1} & \ldots & f'_{nz_1} \\
\vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots \\
f_{2z_n} & \ldots & f'_{nz_n}
\end{pmatrix}
\begin{pmatrix}
f_{2z_1} & \ldots & f'_{nz_1} \\
\vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots \\
f_{2z_n} & \ldots & f'_{nz_n}
\end{pmatrix}
\begin{pmatrix}
f_{1z_1} & \ldots & f_{1z_1} \\
\vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots \\
f_{1z_n} & \ldots & f_{1z_n}
\end{pmatrix}
\begin{pmatrix}
f_{1z_1} & \ldots & f_{1z_1} \\
\vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots \\
f_{1z_n} & \ldots & f_{1z_n}
\end{pmatrix}
= ABC.
\]

We apply the Cauchy-Binet Formula to (69) to see that \( \pm D \) is equal to the determinant of

\[
\begin{pmatrix}
f_{1z_1} & f_{2z_1} & \ldots & f_{nz_1} \\
\vdots & \vdots & \ddots & \vdots \\
f_{1z_n} & f_{2z_n} & \ldots & f_{nz_n}
\end{pmatrix}
\begin{pmatrix}
\epsilon'_1 & 0 & \ldots & 0 \\
0 & \epsilon'_2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \epsilon'_{n'}
\end{pmatrix}
\begin{pmatrix}
f_{1z_1} & \ldots & f_{1z_1} \\
\vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots \\
f_{1z_n} & \ldots & f_{1z_n}
\end{pmatrix}
\begin{pmatrix}
f_{1z_1} & \ldots & f_{1z_1} \\
\vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots \\
f_{1z_n} & \ldots & f_{1z_n}
\end{pmatrix}
= ABC.
\]

The Cauchy-Binet formula tells us that in order to compute the determinant of this product, we just need to take the sum of the products of the determinants of square matrices obtained from \( A \) and \( BC \) by deleting a column in \( A \) and the corresponding row in \( BC \); but this sum is just the sum in (69). Now apply Lemma 28 to see that the determinant of (70) is just \( \pm ig_w(0) \) which we assume to be nonzero. Hence, the claim is proved, and summarizing, we have proved the following.

**Proposition 30.** Let \( M \subset \mathbb{C}^N \) and \( M' \subset \mathbb{C}^{N+1} \) be Levi-nondegenerate formal real hypersurfaces. Let \( H : M \to M' \) be a formal holomorphic map which is constantly 1-degenerate and transversal. Let \( Z = (z, \tau) \) be normal coordinates for \( M \). Then there exists a formal function \( \Psi : \mathbb{C}^N \times \mathbb{C}^N \times \mathbb{C}^{N+1} \times \mathbb{C}^{N+1} \times \mathbb{C}^{N(N+1)} \times \mathbb{C}^{N+1} \times \mathbb{C}^{N(N+1)} \to \mathbb{C}^{N+1} \) such that

\[
H(Z) - \Psi(Z, \zeta, \overline{\zeta}, H(\zeta), \partial \overline{H}(\zeta) - \partial H(0), H(0, \tau), \partial H(0, \tau) - \partial H(0)) \in \mathcal{I}^{n+1}.
\]

Furthermore, \( \Psi \) depends only on \( M, M' \) and the first derivative of \( H \) at 0, such that if \( H' \) is another map fulfilling the assumptions of the proposition with \( \partial H(0) = \partial H'(0) \) then (71) holds with \( H \) replaced by \( H' \). If \( M \) and \( M' \) are real-analytic, \( \Psi \) is convergent on a neighbourhood of the origin. If \( M \) and \( M' \) are algebraic, so is \( \Psi \).
Differentiating this identity as in the proof of Corollary 26 we obtain the following.

**Corollary 31.** Under the assumptions of Proposition 30, the following holds: For all multi-indices \( \alpha \in \mathbb{N}^n \), there is a formal function \( \Psi_\alpha : \mathbb{C}^N \times \mathbb{C}^N \times \mathbb{C}^{N+1} \times \mathbb{C}^{(K(1+|\alpha|)-1)(N+1)} \times \mathbb{C}^{K(|\alpha|)} \times \mathbb{C}^{N+1} \rightarrow \mathbb{C}^{N+1} \) which is polynomial in its 6th, 7th and 10th variable such that

\[
\begin{align*}
(72) \quad & \partial^\alpha H(Z) - \Psi_\alpha(Z, \zeta, H(\zeta), \bar{H}(\zeta), \bar{H}(\zeta), \\
& \partial \bar{\Pi}(\zeta) - \partial \bar{\Pi}(0), (\partial^\beta \bar{\Pi}(\zeta))_{2 \leq |\beta| \leq 1 + |\alpha|}, (\partial^\beta H(\zeta))_{1 \leq |\beta| \leq |\alpha|}, \\
& H(0, \tau), \partial H(0, \tau) - \partial H(0), (\partial^\beta H(0, \tau))_{2 \leq |\beta| \leq 1 + |\alpha|} \in I^{N+1};
\end{align*}
\]

\( \Psi_\alpha \) depends only on \( M, M' \), and on the first derivative of \( H \) at 0 such that if \( H' : M \rightarrow M' \) is another formal map fulfilling the assumptions of Proposition 30 with \( \partial H(0) = \partial H'(0) \) then (72) holds with \( H \) replaced by \( H' \). If \( M \) and \( M' \) are real analytic, \( \Psi_\alpha \) is convergent. If \( M \) and \( M' \) are algebraic, so is \( \Psi_\alpha \).

The main difference between (47) and (72) is that in (72) the argument \((0, \tau)\) appears. This means that we can only iterate (71) once, and hence we can determine \( H \) from its 2-jet at 0 only on the 2nd Segre set. This is the main reason why we have to restrict to hypersurfaces here.

**Corollary 32.** Under the assumptions of Proposition 30, for all \( \alpha \in \mathbb{N}^N \) there is a formal function \( \Upsilon_\alpha : \mathbb{C}^n \times \mathbb{C}^n \rightarrow \mathbb{C}^{N+1} \) such that

\[
(73) \quad \partial^\alpha H(z, Q(z, \chi, 0)) = \Upsilon_\alpha(z, \chi).
\]

\( \Upsilon_\alpha \) depends only on \( M, M' \) and \( \partial^\beta H(0) \) for \( |\beta| \leq 2 + |\alpha| \) such that if \( H' \) is another map fulfilling the hypotheses of Proposition 30 with \( \partial^\beta H(0) = \partial^\beta H'(0) \) for \( |\beta| \leq 2 + |\alpha| \), then (73) holds with \( H \) replaced by \( H' \). If \( M \) and \( M' \) are real-analytic, then \( \Upsilon_\alpha \) is convergent on a neighbourhood of 0 in \( \mathbb{C}^n \times \mathbb{C}^n \). If \( M \) and \( M' \) are algebraic, so is \( \Upsilon_\alpha \).

The proof of this corollary is by induction just as in Corollary 26 and left to the reader. Theorem 7 in the case \( s = 1 \) follows from Corollary 32 just as Theorem 2 follows from Corollary 26. Theorem 9 also follows easily from Corollary 32 since by Lemma 22 we can always pass to a point where \( H \) is of constant degeneracy. However, since we can only work on the second Segre set, we have to work a little harder for Theorem 6. We are basically following an argument given in [2].

**4.3. Proof of Theorem 6.** From Corollary 32 we conclude that

\[
(74) \quad H(z, w) = \sum_{j=1}^{\infty} H^j(z) w^j
\]
where \( H^j(z) = \frac{1}{j!} H_{w^j}(z, 0) \) is convergent. We now want to solve the equation \( w = Q(z, \chi, 0) \) for \( \chi \) in \( w \). Choose a \( \chi_0 \) such that the function \( \phi(z, t) = Q(z, \chi_0 t, 0) \), which is defined in a neighbourhood of the origin in \( \mathbb{C}^n \times \mathbb{C} \), has a derivative in \( t \) which is not constantly vanishing. We write

\[
\phi(z, t) = \sum_{j=1}^{\infty} \phi_j(z) t^j
\]

and define a convergent power series \( \psi(z, t) = t + \sum_{j=2}^{\infty} C_j(z) t^j \)

where \( C_j(z) = \phi_j(z) \phi_1(z)^{j-2} \) for \( j \geq 2 \). By the implicit function theorem, the equation \( w = \psi(z, t) \) has a solution \( t = \theta(z, w) \) which is convergent in a neighbourhood of the origin in \( \mathbb{C}^n \times \mathbb{C} \). Then \( t = \phi_1(z) \theta(z, \frac{w}{\phi_1(z)^2}) \) solves \( w = Q(z, \chi_0 t, 0) \). By changing \( \theta \), we can assume that \( \phi_1(z) = A(z) \) is a Weierstrass polynomial. Hence we conclude that

\[
H(z, w) = F(z, \frac{w}{A(z)^2})
\]

where \( F \) is now a function which is convergent in a neighbourhood of the origin in \( \mathbb{C}^N \). We expand \( F \) in the following way: \( F(z, t) = \sum_{j=1}^{\infty} F_j(z) t^j \).

Comparing coefficients in (76), we conclude that \( H_j(z) = F_j(z) A(z)^{-2} \). We now apply the division theorem to see that

\[
F_j(z) = B_j(z) A(z)^{2j} + r_j(z)
\]

where \( r_j(z) \) is a \( (\mathbb{C}^{N+1}, \text{valued}) \) Weierstrass polynomial of degree less than \( 2jp \) where \( p \) is the degree of the Weierstrass polynomial \( A(z) \). Furthermore, we have the inequality

\[
\|B_j(z)\| \leq C_j \|F_j(z)\|
\]

which holds for \( z \) in a neighbourhood of the origin, with some constant \( C \) (see e.g., [10], Theorem 6.1.1.). Since \( H_j \) is convergent, we conclude that \( r_j \) is the zero polynomial. So \( H_j(z) = B_j(z) \) and from (78) we finally conclude that \( H(z, w) \) is convergent in a neighbourhood of the origin.

5. Strictly pseudoconvex targets.

We will just indicate how to derive a basic identity in this case; the proof is then finished by exactly the same arguments as in the Levi-nondegenerate case. By the Chern-Moser normal form ([8]), we can in particular assume that the target hypersurface is given in normal coordinates \( (z', \tau') \) by \( w' = Q'(z', \chi', \tau') \), where

\[
Q'(z', \chi', \tau') = \tau + i(z', \chi') + \sum_{\alpha, \beta, \gamma \geq 2} c_{\alpha, \beta, \gamma} z'^\alpha \chi'^\beta \tau'^\gamma.
\]
In this equation, \( \langle z', \chi' \rangle = \sum_{j=1}^{n'} z_j' \chi_j' = (z')^t \chi' \). It follows that

\[
L^\alpha Q'_Z(f(z,w), \bar{f}(\chi,\tau), \bar{g}(\chi,\tau))|_0 = L^\alpha \bar{f}(0).
\]

We only needed the Chern-Moser normal form in order to get rid of terms of the form \( z_j' \chi_j^\beta \) in the power series expansion for \( Q' \) where \(|\beta| > 1\) in order to arrive at this easy formula for \( L^\alpha Q'_Z(f(z,w), \bar{f}(\chi,\tau), \bar{g}(\chi,\tau))|_0 \). Hence, if \( H \) is constantly \((k_0, s)\) degenerate at 0, we can choose \( t = n' - s \) multi-indices \( \alpha^1, \ldots, \alpha^t, \alpha^j \in \mathbb{N}^n, |\alpha^j| \leq k_0 \), such that the vectors \( \xi_j = L^\alpha \bar{f}(0), 1 \leq j \leq t \) are linearly independent in \( \mathbb{C}^{n'} \). We extend this set to a basis \( \xi_1, \ldots, \xi_{n'} \) of \( \mathbb{C}^{n'} \). Since \( M' \) is strictly pseudoconvex, we can use the Gram-Schmidt orthogonalization process to obtain vectors \( v_1, \ldots, v_{n'} \) which are orthonormal with respect to the standard hermitian product on \( \mathbb{C}^{n'} \), and a lower triangular invertible matrix \( C \) such that \( V = CE \), where \( V \) denotes the unitary matrix with rows \( v_1, \ldots, v_{n'} \), and \( E \) denotes the matrix with rows \( \xi_1, \ldots, \xi_{n'} \). We change coordinates by \( \tilde{z} = Vz, \tilde{w} = w \). Since \( V \) is unitary, the defining function still has the form (79); in particular, (80) holds with \( f \) replaced by \( \bar{f} \), and \( L^\alpha \bar{f}(0) = \nabla L^\alpha \bar{f}(0) \). Since \( \nabla E = (E')^{-1} \) is upper triangular, it follows that we can assume \( \xi^k_j = (L^\alpha \bar{f}(0))_k = 0 \) for \( j > k \). Note that the same change of coordinates works for every other map whose \( k_0 \)-jet at the origin agrees with the \( k_0 \)-jet of \( H \), if \( k_0 > 1 \). Note that we have used the strict pseudoconvexity only to reduce to the case where the matrix \( (L^\alpha \bar{f}(0)) \) is triangular.

We now start as in the proof of Proposition 25 and obtain formal functions \( \Phi_j(Z, \zeta, Z', \zeta', W) \in \mathbb{C}[Z, \zeta, Z', \zeta', W] \), defined by \( \Phi_j(Z, \zeta, H(Z), \bar{H}(\zeta), (\partial^\beta \bar{H}(\zeta) - \partial^\beta \bar{H}(0))|_{|\beta| \leq k_0}) = (L^\alpha Q'(f, \bar{f}, \bar{g}))(Z, \zeta), \) which are convergent if \( M \) and \( M' \) are real-analytic and algebraic if \( M \) and \( M' \) are algebraic. Also note that \( \Phi_j \) does not depend on \( Z_{j'} \).

The missing equations are supplied by Lemma 16. So we define formal functions \( \Upsilon_k(Z, \zeta, Z', \zeta', S, T, W') \), \( 1 \leq k \leq s \), by

\[
\Upsilon_{j-t}(Z, \zeta, H(Z), \bar{H}(\zeta), (\partial^\beta H(0,\tau) - \partial^\beta \bar{H}(0))|_{1 \leq |\beta| \leq k_0}) = \Delta(\chi, \tau) \bar{Q}_{\chi_j'}(\bar{f}(\chi, \tau), f(z,w), g(z,w))
- \sum_{m=1}^t \Delta_m(\chi, \tau) \bar{Q}_{\chi_m'}(\bar{f}(\chi, \tau), f(z,w), g(z,w))
\]

for \( t + 1 \leq j \leq n' \), where \( \Delta \) is defined by (27) and \( \Delta_m \) is given by (28). Recall that the functions \( \Upsilon_k \) are convergent if \( M \) and \( M' \) are real-analytic and algebraic if \( M \) and \( M' \) are algebraic. We now claim that we can apply
the implicit function theorem to solve the system
\[
\Phi_j(Z, \zeta, Z', \zeta', W) = 0, \quad 1 \leq j \leq t,
\]
\[
(82) \quad \Upsilon_k(Z, \zeta, Z', \zeta', S, T, W') = 0, \quad 1 \leq k \leq s,
\]
\[
Z'_{N'} = Q'(Z'_1, \ldots, Z'_{n'}, \zeta'),
\]
uniquely in $Z'$. First note that $\Phi_j(Z, \zeta, Z', \zeta', W) = 0$ and $\Upsilon_k(Z, \zeta, Z', \zeta', S, T, W') = 0$ for $1 \leq j \leq t$, and $1 \leq k \leq s$. So we only need to consider the Jacobian of $(\Phi_1, \ldots, \Phi_t, \Upsilon_1, \ldots, \Upsilon_s)$ with respect to $(Z'_1, \ldots, Z'_{n'})$. Now $\Phi_j(Z', \zeta', 0) = \xi_j$, and a little computation shows that $\Upsilon_k(Z', \zeta', 0) = (\Delta_1(0), \ldots, \Delta_t(0), 0, \ldots, \Delta(0), \ldots, 0)$, where the $\Delta(0)$ appears in the $(t+k)$-th spot, $1 \leq k \leq s$. Recall that $\xi_j^k = 0$, $j > k$, by our choice of coordinates. Writing out the determinant we see that indeed the implicit function theorem applies. This gives the desired basic identity.

References


Received November 9, 1999 and revised June 16, 2000.

DEPARTMENT OF MATHEMATICS, 0112
UNIVERSITY OF CALIFORNIA, SAN DIEGO
SAN DIEGO, CA 92092-0112

E-mail address: blamel@euclid.ucsd.edu
REGULARITY OF SOLUTIONS OF OBSTACLE PROBLEMS FOR ELLIPTIC EQUATIONS WITH OBLIQUE BOUNDARY CONDITIONS

GARY M. LIEBERMAN

Much has been written about various obstacle problems in the context of variational inequalities. In particular, if the obstacle is smooth enough and if the coefficients of associated elliptic operator satisfy appropriate conditions, then the solution of the obstacle problem has continuous first derivatives. For a general class of obstacle problems, we show here that this regularity is attained under minimal smoothness hypotheses on the data and with a one-sided analog of the usual modulus of continuity assumption for the gradient of the obstacle. Our results apply to linear elliptic operators with Hölder continuous coefficients and, more generally, to a large class of fully nonlinear operators and boundary conditions.

Introduction.

For a smooth bounded domain $\Omega \subset \mathbb{R}^n$ with unit inner normal $\gamma$, we are concerned with generalizations of the simple obstacle problem of finding a function $u \in W^{1,2}(\Omega)$ which minimizes the functional $F$ defined on $W^{1,2}$ by

$$F(v) = \int_{\Omega} |Dv|^2 \, dx + \int_{\partial\Omega} v^2 \, d\sigma$$

over the set of all $v \in W^{1,2}$ with $v \geq \psi$ for a given function $\psi$. From standard results in the theory of variational inequalities and the arguments in [12], it follows that this minimizer has bounded second derivatives if $\psi$ has bounded second derivatives and satisfies the inequality $\partial \psi / \partial \gamma - \psi \geq 0$ on $\partial \Omega$, which is assumed sufficiently smooth. To state our generalization of this problem, we note that the minimizer $u$ will be superharmonic in $\Omega$ and harmonic on the set where $u > \psi$; in addition $\partial u / \partial \gamma - u = 0$ on $\partial \Omega$. It is this formulation of the minimization problem that we wish to generalize.

We write $S^n$ for the set of all $n \times n$ symmetric matrices, and we set $\Gamma = \Omega \times \mathbb{R} \times \mathbb{R}^n \times S^n$ and $\Gamma' = \partial \Omega \times \mathbb{R} \times \mathbb{R}^n$. For real-valued, differential functions $F$ and $G$ defined on $\Gamma$ and $\Gamma'$, respectively, we consider the problem

(0.1) $\min \{-F(x, u, Du, D^2u), u - \psi\} = 0$ in $\Omega$, $G(x, u, Du) = 0$ on $\partial \Omega$. 

389
(We justify this somewhat nonstandard way of writing the problem by pointing out that, in the special case that $F$ is the Laplace operator, our solution $u$ will be superharmonic with $u \geq \psi$ and $u$ is harmonic on the set \{ $u > \psi$ \}. Using subscripts to denote partial derivatives with respect to the variables $z \in \mathbb{R}$, $p \in \mathbb{R}^n$ and $r \in S^n$, we assume at least that the matrix $F_r(x,z,p,r)$ is positive definite for all $(x,z,p,r) \in \Gamma$ (so that the equation $F(x,u,Du,D^2u) = 0$ is elliptic) and that $G_p(x,z,p) \cdot \gamma(x) > 0$ for all $(x,z,p) \in \Gamma'$ for the unit inner normal $\gamma$ to $\partial \Omega$ (so the boundary condition is an oblique derivative condition). If $\psi \in C^1$, then the analog of the condition $\partial \psi / \partial \gamma - \psi \geq 0$ would be $G(x,\psi, D\psi) \geq 0$ on $\partial \Omega$ (compare with [12, (0.7)]); however, we shall assume a weaker condition than continuity of $D\psi$ which still implies the continuity of $Du$, so we shall modify this condition appropriately (see conditions (2.1) and (3.9) below). Under suitable regularity hypotheses on $F$, $G$, $\psi$, and $\Omega$, we shall show that a modulus of continuity for the first derivatives of $u$ can be estimated in terms of known data. In conjunction with known first derivative estimates, our results give a complete description of the regularity of solutions for several problems. As particular examples, we mention here the capillarity obstacle problem from [12] and the Bellman equation problem with linear boundary condition from [23] (strictly speaking, we refer to the problem which the authors of that paper defer to a sequel, listed there as reference [27], which has never appeared in print). In [12], $F$ has the special form

\[(0.2a) \quad F(x,z,p,r) = (1 + |p|^2)^{-1/2} \left( \frac{\delta^{ij} - \frac{P_iP_j}{1 + |p|^2}}{1 + |p|^2} \right) r_{ij} + a(x,z),\]

for a suitable, Lipschitz function $a$ and $G$ has the form

\[(0.2b) \quad G(x,z,p) = \frac{p}{(1 + |p|^2)^{1/2}} \cdot \gamma + \varphi(x,z)\]

for a suitable, smooth function $\varphi$ such that $\sup |\varphi(x,z)| < 1$. Because of the bound on the gradient of the solution of (0.1) in [12], it follows that our estimates apply to this problem assuming that $a$ is Lipschitz and $\varphi$ is $C^{1,\alpha}$ for some $\alpha \in (0,1)$. In [23], $F$ has the form

\[(0.3a) \quad F(x,z,p,r) = \inf_{k \in J} \left( a_k^{ij}(x) r_{ij} + b_k^i(x)p_i + c_k(x)z + f_k(x) \right),\]

where $J$ is some index set (assumed to be countable in [23]) and there are uniform (with respect to $k$) bounds on the $C^2$ norms of the coefficients $a_k^{ij}$, $b_k^i$, $c_k$, and $f_k$ as well as a positive lower bound (independent of $k$) on the minimum eigenvalue of the matrix $[a_k^{ij}]$; $G$ has the form

\[(0.3b) \quad G(x,z,p) = \beta(x) \cdot p + b(x)z + g(x)\]
for some vector $\beta$ such that $\beta \cdot \gamma$ is bounded from below by a positive constant, and the $C^2$ norms of $\beta$, $b$ and $g$ are assumed to be bounded. Again, from the gradient bounds proved in [23], it can be shown that our results apply to such problems if we only assume bounds on the H"older norms of $a_k^i j$, $b_k^i$, $c_k$, and $f_k$ (see Theorem 2.2) and on the H"older norms of $\beta$, $b$ and $g$ (see Theorem 3.2); for second derivative bounds, we need to assume that $\beta$, $b$, and $g$ have H"older continuous derivatives (see Theorem 3.3). Of course, the uniform lower bounds on the minimum eigenvalue of $[a_k^i j]$ and on $\beta \cdot \gamma$ cannot be relaxed for our techniques to work.

In addition to the one-sided condition on $\psi$, our hypotheses are weaker than those in [11, Section 2], [12], [16, Section 4], [1], and [2] because we relax the smoothness hypotheses on $F$, $G$, and $\Omega$.

A basic interpolation inequality appears in Section 1, which allows us to use a weak Harnack inequality rather than the usual Harnack inequality. An interior regularity result is proved in Section 2 using a modification of the technique pioneered by Caffarelli and Kinderlehrer [4]. Specifically, we show (via the weak Harnack inequality) that our one-sided condition on $\psi$ implies a two-sided integral bound for $u - L$ with $L$ a suitable linear function, and then the interpolation inequality from Section 1 gives a two-sided estimate on the first derivatives of $u$. The corresponding estimates at the boundary are proved in Section 3. Most of our work is to analyze the hypotheses on the obstacle; only some simple elements of the theory of differential equations enters into this analysis. Some similar results, with a Dirichlet boundary condition replacing the oblique derivative boundary condition, appear in a preprint by Jensen [13]. The analysis of the obstacle also provides a straightforward extension to the two obstacle problem, which we present in Section 4, and Section 5 discusses applications of our methods to some degenerate variational inequalities; in particular, problems with the $p$-Laplacian operators are considered. We close in Section 6 with an outline of the existence theory in a special case.

Our notation follows that in [10]. In addition, we write $F^{i j}$ for the components of the matrix $F$, and $F^i$ for the components of the vector $F$. Similarly, $G^i$ denotes the components of the vector $G$. We always assume here that $\psi$ is Lipschitz with

\[(0.4) \quad |\psi| + |D\psi| \leq \Psi_1,\]

and we define

$\Gamma'_1 = \{(x, z, p) \in \Gamma' : |z| + |p| \leq \Psi_1\}$. 
1. An interpolation lemma.

Our first lemma is an improvement of results on second derivative estimates in terms of estimates on lower order derivatives. For brevity, if $\Sigma \subset \Omega$, we use $|u|^{(b)}_{a;\Sigma}$ to denote the norms weighted in terms of distance to $\partial \Sigma \cap \Omega$.

**Lemma 1.1.** Let $\Omega$ be a bounded Lipschitz domain, let $\Sigma$ be a subset of $\overline{\Omega}$, and suppose $u \in C^{2+\alpha}(\Sigma)$ for some $\alpha \in (0, 1)$. Suppose that there are positive constants $C_1$ and $C_2$ such that

\[
|D^2u|^{(2)}_{\alpha;\Sigma \cap B(R)} \leq C_1 R^{-\alpha} |D^2u|_{0;\Sigma \cap B(2R)} + C_2
\]

for any two concentric balls $B(R)$ and $B(2R)$, with radii $R$ and $2R$, respectively, such that the boundary of $\Sigma \cap B(2R)$ is disjoint from $\Omega \setminus \Sigma$. Then there is a constant $C$ determined only by $C_1$, $\alpha$, and $\Omega$ such that

\[
|u|^{(0)}_{2+\alpha;\Sigma} \leq C \left( |u|_{0;\Sigma} + C_2 (\text{diam } \Sigma)^{2+\alpha} \right).
\]

In addition, if $\Sigma = \overline{\Omega} \cap B(2R)$ for some ball $B(2R)$, and if $\kappa > 0$, then

\[
|D^2u|_{0;\Sigma'} \leq C(C_1, \alpha, \kappa, \Omega) \left( R^{-2-(n/\kappa)} \|u\|_{\kappa;\Sigma} + C_2 R^\alpha \right),
\]

where $\Sigma' = \overline{\Omega} \cap B(R)$.

**Proof.** The proof of (1.2) is a simple combination of the interpolation inequality

\[
|D^2u|^{(2)}_{\alpha} \leq C \left( |D^2u|^{(2+\alpha)}_{\alpha} + |u|_0 \right)^{2/(\alpha+2)} \left( |u|_0 \right)^{\alpha/(2+\alpha)}
\]

and the observation that (1.1) implies that

\[
|D^2u|^{(2+\alpha)}_{\alpha} \leq C \left( C_1 |D^2u|^{(2)}_{0} + C_2 (\text{diam } \Sigma)^{2+\alpha} \right).
\]

To prove (1.3), we imitate the proof of [19, Lemma 4.5]. From (1.2), we infer that

\[
\rho \sup_{S(\rho)} |Du| \leq C \left( \sup_{S(2\rho)} |u| + C_2 \rho^{2+\alpha} \right),
\]

where $S(\rho) = \Sigma \cap B(\rho)$ and the boundary of $S(2\rho)$ is disjoint from $\Omega \setminus \Sigma$. It follows that there is a constant $C_0$ determined only by $C_1$, $\alpha$, $\kappa$, and $\Omega$ such that

\[
\text{osc}_{S(\theta \rho)} u \leq C_0 \theta \left( \sup_{S(2\rho)} |u| + C_2 \rho^{2+\alpha} \right)
\]

for any $\theta \in (0, 1)$. Now we take $x_1$ so that $d(x_1)^{n/\kappa} |u(x_1)| \geq \frac{1}{3} |u|_0^{(n/\kappa)}$ and we choose our balls to be centered at $x_1$ with $\rho = \frac{1}{4} d(x_1)$. Then

\[
|u(x)| \geq |u(x_1)| \left[ 1 - C_0 \theta 2^{n/\kappa} \right] - C_0 \theta C_2 \rho^{2+\alpha}
\]
for $x \in S(\theta \rho)$. If we take $\theta$ so small that $C_0 \theta 2^{n/\kappa} \leq 1/2$ and $C_0 \theta \leq 1$, then rearranging the resulting inequality and integrating over $S(\rho)$ yields

$$
\rho^n |u(x_1)|^\kappa \leq C \left( \int_{S(\theta \rho)} |u|^\kappa \, dx + C_0^\kappa \rho^{(2+\alpha)\kappa + n} \right),
$$

and therefore

$$
[u]^{(n/\kappa)}_{0} \leq C \left( ||u||_\kappa + C_2 R^{2+\alpha + (n/\kappa)} \right).
$$

Hence

$$
|u|_{0;\Omega \cap B(3R/2)} \leq C \left( ||u||_\kappa + C_2 R^{2+\alpha} \right)
$$

for any ball $B(3R/2)$. The desired result follows from this one after applying (1.2) with $\Sigma$ replaced by $\Omega \cap B(3R/2)$. □

2. Interior derivative estimates.

Our main ingredient is a pointwise estimate of how fast $u$ moves away from the obstacle near a contact point. In this section, we prove this estimate at an interior point. The argument is a straightforward modification of that in [4], but because of Lemma 1.1, we only need to estimate the $L^\kappa$ norm of a function related to $u$; this estimate is proved quite simply. Our basic assumption on the obstacle $\psi$ is that there are functions $Y$ defined on $\Omega$ and $\zeta$ defined on $[0, \text{diam} \Omega]$ with $\zeta$ continuous and increasing such that

$$
\psi(x_1) \geq \psi(x_2) + Y(x_2) \cdot (x_1 - x_2) - \zeta(|x_1 - x_2|)|x_1 - x_2|
$$

for all $x_1$ and $x_2$ in $\Omega$. We have not assumed that $\zeta(0) = 0$, even though this assumption is needed to conclude that $Du$ is actually continuous, because it does not affect the form of our estimates. Note that the usual assumption (from [1, 4, 11, 12, 16, 18]) is that $|D\psi(x_1) - D\psi(x_2)| \leq \zeta(|x_1 - x_2|)$, which is equivalent to the combination of (2.1) and the companion inequality

$$
\psi(x_1) \leq \psi(x_2) + Y(x_2) \cdot (x_1 - x_2) + \zeta(|x_1 - x_2|)|x_1 - x_2|.
$$

Our condition includes functions which are not continuously differentiable even if $\zeta(0) = 0$. For example, if $(\psi_\alpha)_{\alpha \in I}$ is a family of functions (with arbitrary index set $I$) satisfying (2.1), then a simple calculation shows that $\psi$ defined by $\psi(x) = \sup_{\alpha \in I} \psi_\alpha(x)$ also satisfies this condition provided we have a uniform $L^\infty$ bound on $\psi_\alpha$ and $D\psi_\alpha$. In particular, condition (2.1) includes the obstacles studied by Troianiello in [30, 31].

**Lemma 2.1.** Suppose that $u \in W^{2,n}_{\text{loc}}$ satisfies

$$
\min\{-F(x,u,Du,D^2u), u - \psi\} = 0 \text{ in } \Omega
$$

and that there are positive constants $\lambda$, $\Lambda$, and $\mu_0$ such that

$$
\lambda |\xi|^2 \leq F^{ij}(x,u,Du,D^2u)\xi_i \xi_j
$$

(2.3)
for all $x \in \Omega$ and $\xi \in \mathbb{R}^n$,

\begin{align}
\tag{2.4a}
F(x, u, Du, 0) & \geq -\mu_0 \lambda, \\
\tag{2.4b}
|F_r(x, u, Du, tD^2 u)| & \leq \Lambda,
\end{align}

for all $x \in \Omega$ and all $t \in [0,1]$. Suppose also that (2.1) holds and that $x_0$ is a point such that $u(x_0) = \psi(x_0)$. Then there are constants $\kappa$ and $C$ determined only by $n$ and $\Lambda/\lambda$ such that the function $\bar{u}$ defined by

\begin{equation}
\bar{u}(x) = u(x) - u(x_0) - Y(x_0) \cdot (x - x_0)
\end{equation}

satisfies the estimate

\begin{equation}
\left( R^{-n} \int_{B(x_0, R/2)} |\bar{u}|^\kappa \, dx \right)^{1/\kappa} \leq C[\mu_0 R^2 + \zeta(R) R]
\end{equation}

for all $R \leq d(x_0)/2$.

Proof. Since $u \geq \psi$ and $u(x_0) = \psi(x_0)$, it follows from (2.1) that $\bar{u} \geq -\zeta(R) R$ in $B(x_0, R)$. Next, we note that $a^{ij} D_{ij} u \leq -F(x, u, Du, 0)$ for

\begin{equation}
a^{ij}(x) = \int_0^1 F^{ij}(x, u(x), Du(x), tD^2 u(x)) \, dt.
\end{equation}

It follows that $v = \bar{u} + \zeta(R) R$ satisfies the conditions $a^{ij} D_{ij} v \leq \lambda \mu_0$ and $v \geq 0$ in $B(x_0, R)$. Therefore [10, Theorem 9.22] and the obvious inequality $\inf_{B(R/2)} v \leq v(0) = \zeta(R) R$ yield

\begin{equation}
\left( R^{-n} \int_{B(R/2)} |v|^\kappa \, dx \right)^{1/\kappa} \leq C[\mu_0 R^2 + \zeta(R) R],
\end{equation}

and the triangle inequality gives

\begin{equation}
\left( R^{-n} \int_{B(R/2)} |\bar{u}|^\kappa \, dx \right)^{1/\kappa} \leq C(\kappa) \left[ \left( R^{-n} \int_{B(R/2)} |v|^\kappa \, dx \right)^{1/\kappa} + \zeta(R) R \right].
\end{equation}

We complete the proof by combining these last two inequalities. \hfill \Box

Note Lemma 2.1 continues to hold if we only assume that the minimum in (2.2) is nonnegative; however, our full regularity result will use that the minimum is zero.

The regularity of the derivatives of $u$ at an arbitrary point follows from this estimate and Lemma 1.1 by a simple variation of the argument in [4].

Theorem 2.2. Suppose that $u$, $\psi$, and $F$ satisfy conditions (2.1)-(2.4) with $\zeta$ a continuous increasing function on $[0, \text{diam} \Omega]$ satisfying

\begin{equation}
\frac{\zeta(t_1)}{t_1} \geq \frac{\zeta(t_2)}{t_2} \text{ if } t_1 \leq t_2.
\end{equation}

Suppose also that there are constants $\alpha \in (0, 1)$ and $\mu_1$ such that

$$\left| F(x, z, p, r) - F(y, w, q, r) \right| \leq (\mu_0 + \mu_1 |r|) \lambda |x - y|^\alpha, \tag{2.8}$$

and that $F$ is convex or concave with respect to $r$. Then there is a constant $C$ determined only by $n$, $\alpha$, $\mu_1$, $\Lambda/\lambda$, $|u_1|$, $|\psi_1|$, and $\operatorname{diam} \Omega$ such that

$$|Du(x_1) - Du(x_2)| \leq C \left[ \zeta(|x_1 - x_2|) + \left( \mu_0 + \frac{\sup |Du|}{d(x_1)} \right) |x_1 - x_2| \right] \tag{2.9}$$

for all $x_1$ and $x_2$ in $\Omega$ with $|x_1 - x_2| \leq \frac{1}{4} \min\{d(x_1), d(x_2)\}$.

Proof. Using $I$ to denote the contact set $I = \{ x \in \Omega : \psi(x) = u(x) \}$, we consider three cases:

(i) both points are in $I$,

(ii) one point is in $I$,

(iii) neither point is in $I$.

In all cases, we set $\rho = |x_1 - x_2|$ and $Z = \zeta(\rho) + \mu_0 \rho$.

In the first case, we use (2.6) twice, first with $R = \rho$ and $x_0 = x_1$ and then with $R = 2\rho$ and $x_0 = x_2$ to infer that

$$\left( \rho^{-n} \int_{B(x_1, \rho)} |u(x) - \psi(x_1) - Y(x_1) \cdot (x - x_1)|^\kappa \, dx \right)^{1/\kappa} \leq CZ \rho$$

for $i = 1, 2$ because $B(x_1, \rho) \subset B(x_2, 2\rho) \subset \Omega$. Next, we use the observation that

$$\begin{align*}
\psi(x_1) - \psi(x_2) - Y(x_2) \cdot (x_1 - x_2) &\geq [Y(x_1) - Y(x_2)] \cdot (x - x_1) \\
&= [u(x) - \psi(x_1) - Y(x_1) \cdot (x - x_1)] - [u(x) - \psi(x_2) - Y(x_2) \cdot (x - x_2)]
\end{align*}$$

along with the triangle inequality to infer that

$$\left( \rho^{-n} \int_{B(x_1, \rho)} \left| \psi(x_1) - \psi(x_2) - Y(x_2) \cdot (x_1 - x_2) \right|^{1/\kappa} \, dx \right)^{1/\kappa} \leq CZ \rho,$$

where $V = Y(x_1) - Y(x_2)$. In addition, (2.1) implies that

$$\psi(x_1) - \psi(x_2) - Y(x_2) \cdot (x_1 - x_2) + V \cdot (x - x_1) \geq V \cdot (x - x_1) - \zeta(\rho) \rho$$

in $B(\rho)$. We therefore infer that

$$\left( \rho^{-n} \int_{B(x_1, \rho)} \left| \left( V \cdot (x - x_1) \right)^+ \right|^{1/\kappa} \, dx \right)^{1/\kappa} \leq CZ \rho,$$
and this inequality easily gives $|V| \leq CZ$. It follows that

$$
(2.10) \quad |u(x_1) - u(x_2) - Y(x_2) \cdot (x_1 - x_2)| \leq C[|\zeta(|x_1 - x_2|)| + \mu_0 |x_1 - x_2||x_1 - x_2|]
$$

for any $x_1$ and $x_2$ in $I$, so $u$ is differentiable in the interior of $I$ with $Du = Y$ there.

In the second case, we may assume without loss of generality that $x_1 \in I$, and we write $\xi_2$ for the closest point to $x_2$ in $I$. Note that $u$ is a solution of the equation $F(x,u,Du,D^2u) = 0$ in $\Sigma_0 = B(x_2,|x_2 - \xi_2|)$, so (1.1) holds with $\Sigma$ any subset of $\Sigma_0$. (This estimate is proved in [28], but the precise form used here does not appear in that reference; see [19, Theorem 14.7] for a proof of the corresponding parabolic estimate.) If $x$ is on the line segment between $x_2$ and $\xi_2$, it follows from (1.3) with $R = 2|x - \xi_2|$ (applied to $\bar{u}$ defined with $\xi_2$ in place of $x_0$) that

$$
|Du(x)| \leq C[R^{1-(\alpha/\kappa)}\|u\|_\kappa + \mu_0 R^{1+\alpha}]
$$

and hence

$$
|Y(\xi_2) - Du(x_2)| \leq C|x_2 - \xi_2| \leq CZ.
$$

Since $|x_1 - \xi_2| \leq 2\rho$, it follows from Case (i) that $|Y(x_1) - Y(\xi_2)| \leq CZ$, and hence (2.10) holds if $x_1 \in I$ and $x_2 \notin I$. Therefore $u$ is also differentiable on $\partial I$ with $Du = Y$ there. Now that we know $Du = Y$ on $I$, our estimates imply (2.9) for $x_1 \in I$ and $x_2 \in \Omega$.

In the third case, we set $d^*(x) = \text{dist}(x,I)$ and $m_0 = \min\{d^*(x_1),d^*(x_2)\}$, and we consider three possibilities. If $2\rho \geq m_0$, then, with $\xi_i$ denoting the closest point to $x_i$ in $I$, we have

$$
|Du(x_1) - Du(x_2)| \leq |Du(x_1) - Du(\xi_1)| + |Du(\xi_1) - Du(\xi_2)| + |Du(\xi_2) - Du(x_2)|,
$$

and the three terms on the right-hand side of this inequality are estimated either by Case (i) or Case (ii) along with the observation that

$$
|x_1 - \xi_1| \leq C\rho, \quad |\xi_1 - \xi_2| \leq C\rho, \quad |x_2 - \xi_2| \leq C\rho.
$$

If $2\rho < m_0$ and $d(x_1) \leq m_0$, then we can use Lemma 1.1 as in Case (ii) and (2.7) to infer that

$$
|D^2u| \leq C\left[\frac{\zeta(m_0)}{m_0} + \mu_0 m_0^\alpha\right] \leq C\left[\frac{\zeta(\rho)}{\rho} + \mu_0\right]
$$

on the line segment joining $x_1$ and $x_2$. An easy integration of this inequality yields (2.9) in this case as well. Finally, if $2\rho < m_0$ and $d(x_1) > m_0$, then (1.3) with $\kappa = \infty$ gives the desired result.

\[\square\]
Note that the hypothesis (2.7) really involves no loss of generality. Specifically if \( \zeta \) is a continuous, increasing function, then the function \( \zeta_1 \), defined by
\[
\zeta_1(t) = t \sup_{s \geq t} \frac{\zeta(s)}{s},
\]
satisfies (2.7) and \( \zeta_1 \geq \zeta \), so (2.1) holds with \( \zeta_1 \) in place of \( \zeta \). In addition \( \zeta_1 \) is clearly continuous. To see that \( \zeta_1 \) is increasing, we let \( t_1 < t_2 \) and choose \( s_i \) so that \( \zeta_1(t_i) = (t_i/s_i)\zeta(s_i) \). If \( \zeta(s_1) = \zeta(s_2) \), then \( \zeta_1(t_1)/\zeta_1(t_2) = t_1/s_1 \). If \( \zeta(s_1) < \zeta(s_2) \), then \( s_1 < t_2 \), so \( \zeta_1(t_1) = (t_1/s_1)\zeta(s_1) \leq (t_1/s_1)\zeta(t_2) \leq \zeta(t_2) \leq \zeta_1(t_2) \). Moreover, if \( \zeta(0) = 0 \), then \( \zeta_1(t) \to 0 \) as \( t \to 0 \), as we see by considering two cases. If \( \zeta(s)/s \) is bounded as \( s \to 0 \), say by \( S \), then \( \zeta_1(t) \leq St \to 0 \). If \( \zeta(s)/s \) is unbounded as \( s \to 0 \), let \( (s_j) \) be a sequence tending to zero with \( \zeta(s_j)/s_j \geq j \) and \( \zeta(s_j)/s_j \geq \zeta(s)/s \) if \( s \geq s_j \). Then \( \zeta_1(s_j) = \zeta(s_j) \) so \( \zeta_1(s_j) \to 0 \), and then \( \zeta(t) \to 0 \) as \( t \to 0 \) because \( \zeta_1 \) is increasing. In addition, we note (see [24, Section 3.5] for details) that the modulus of continuity for a function defined on an open set satisfies (2.7).

Condition (2.8) can be weakened, say to
\[
|F(x, z, p, r) - F(y, w, q, r)| \leq (\mu_0 + \mu_1|r|)\lambda|x - y| + \mu_0 \lambda|p - q|,
\]
since this condition is only used to infer the appropriate form of the Hölder for second derivatives of solutions of the equation \( F(x, u, Du, D^2u) = 0 \) (see [28]). In particular, our results apply to the operator \( F \) defined by (0.3a) if we assume uniform Hölder estimates on the functions \( a^{ij}_k \), \( b^i_k \), \( c_k \), and \( f_k \) along with a uniform lower bound on the minimum eigenvalue of \( [a^{ij}_k] \); this structure was considered in [23]. Moreover, we can infer condition (2.8) for more general classes of fully nonlinear, uniformly elliptic operators \( F \) once we have a Hölder gradient estimate for \( u \). Such an estimate follows by virtue of the following variant of Theorem 2.2, which is also important for our study of oblique derivative problems.

**Theorem 2.3.** Suppose \( u, \psi, \) and \( F \) satisfy conditions (2.1)-(2.4) with \( \zeta \) a continuous, increasing function on \([0, diam \Omega]\). Suppose also that \( F \) is concave or convex with respect to \( r \). Suppose finally that there are a positive constant \( \nu_1 \) and a continuous, increasing function \( \zeta_1 \) with \( \zeta_1(0) = 0 \) such that
\[
|F(x, z, p, r) - F(y, w, q, r)| \leq \mu_0 \lambda + \lambda(\nu_1|p - q| + \zeta_1(|x - y|))|r|.
\]
Then there are positive constants \( \alpha(n, \Lambda/\lambda, \nu_1) \) and \( C(n, \zeta_1, \Lambda/\lambda, \nu_1, diam \Omega) \) such that
\[
\frac{\zeta(t_1)}{t_1^{\alpha}} \geq \frac{\zeta(t_2)}{t_2^{\alpha}} \text{ for } t_1 \leq t_2
\]
implies
\begin{equation}
|Du(x_1) - Du(x_2)| \leq C \left[ \zeta(|x_1 - x_2|) + \left( \mu_0 + \frac{\sup |Du|}{d(x_1)^\alpha} \right) |x_1 - x_2|^\alpha \right]
\end{equation}
for all $x_1$ and $x_2$ in $\Omega$ with $|x_1 - x_2| \leq \frac{1}{4} \min\{d(x_1), d(x_2)\}$.

Proof. We basically follow the proof of Theorem 2.2. The main notational change is that we set $Z = \zeta(\rho) + \mu_0 \rho^\alpha$. From the argument in [19, Lemma 12.13] (see also [3, Theorem 2] and [32]), we infer that
\begin{equation}
[Du]_{\alpha; B(R)} \leq C[R^{-\alpha} |Du|_{[0; B(2R)]} + \mu_0]
\end{equation}
if $B(2R) \subset \Omega$ and $F(x, u, Du, D^2u) = 0$ in $B(2R)$. The proof is completed by using this inequality in the obvious modification of Lemma 1.1.

Note that if $\zeta$ satisfies (2.7), then $\zeta_2$ defined by $\zeta_2(t) = (\sup \zeta)^{1-\alpha}(\zeta(t))^\alpha$ satisfies (2.12), so Theorem 2.3 also does not restrict our choice of obstacles.

Condition (2.11) is certainly satisfied for quasilinear operators, that is, $F(x, z, p, r) = a^{ij}(x, z, p) r_{ij} + a(x, z, p)$ provided $[a^{ij}]$ is elliptic, continuous with respect to $x$ and $z$, and Lipschitz with respect to $p$. In particular (after using the gradient bound from [12]), this result applies when $F$ is given by (0.2a). Moreover, we can remove the hypothesis that $F$ be either concave or convex with respect to $r$ in Theorem 2.3 by considering viscosity solutions as in [3, 32] and suitably modifying the arguments. Finally, as noted before, we can replace condition (2.11) by any condition which yields the Hölder gradient estimate
\begin{equation}
\operatorname{osc}_{B(x_0, r)} Du \leq C \left( \frac{r}{R} \right)^\alpha \operatorname{osc}_{B(x_0, R)} Du + \mu_0 r^\alpha
\end{equation}
See [5] for an alternative structure condition which provides such an estimate.


To prove a modulus of continuity estimate for the gradient up to the boundary for the oblique derivative problem, we use a slight variation of the ideas in the proof of Theorem 2.2. We begin with a preliminary estimate which is related to the boundary condition in which we write $v'$ for the first $n - 1$ components of the vector $v$. The connection of this lemma to our original problem will be made clear in Theorem 3.2.

Lemma 3.1. Let $\omega_0$, $\omega_1$, and $r$ be positive constants with $\omega_0 > \omega_1$, and define
\begin{equation}
K = \{ x^n \geq \omega_0 |x'|, \ |x| \leq r \}, \ E = \{ x^n \geq \omega_1 |x'|, \ r/4 < |x| \leq r \}.
\end{equation}
Let $\psi$ be a Lipschitz function defined in $K$ and suppose that there are positive constants $z$ and $\kappa$ along with a vector-valued function $\nabla{\psi}$ such that
\begin{equation}
\psi(x) \geq \psi(x_1) + \nabla{\psi}(x_1) \cdot (x - x_1) - z
\end{equation}
for all $x$ and $x_1$ in $K$ and
\begin{equation}
\left( \int_E |\psi|^\kappa \, dx \right)^{1/\kappa} \leq z r^{n/\kappa}.
\end{equation}

Suppose also that there is a Lipschitz function $g$ defined on $\mathbb{R}^n$ with
\begin{align}
|\frac{\partial g}{\partial p'}| &\leq \mu_2 \chi_0, \\
\frac{\partial g}{\partial p_n} &\geq \chi_0
\end{align}
for some positive constants $\chi_0$ and $\mu_2$ with $\mu_2 \omega_1 < 1$. Then
\begin{equation}
g(0) \geq g(\nabla{\psi}(0)) - C(n, \kappa, \omega_0, \omega_1, \mu_2) \chi_0 z / r.
\end{equation}

**Proof.** The first step is to prove a pointwise upper bound for $\psi$ in $E' = \{x^n \geq |x'| / \mu_2, 3r / 8 \leq |x| \leq 3r / 4\}$.

To prove this estimate, let $x_1$ be a point in $E'$ at which the maximum of $\psi$ is attained and suppose that $\psi(x_1) > 2z$. Then
\begin{equation}
\left( \int_{B(x_1, \rho)} |\psi|^\kappa \, dx \right)^{1/\kappa} \leq \left( \int_E |\psi|^\kappa \, dx \right)^{1/\kappa} \leq z r^{n/\kappa}
\end{equation}
for any $\rho$ such that $B(x_1, \rho) \subset E$. In particular, we can take $\rho = C(\omega_1, \mu_2) r$.

With this choice for $\rho$, we set
\begin{equation}
E^+ = \{x \in B(x_1, \rho) : \nabla{\psi}(x_1) \cdot (x - x_1) \geq 0\},
\end{equation}
and note that $|E^+| \geq \frac{1}{2} |B(x_1, \rho)| \geq C r^n$. In addition, for $x \in E^+$, we have
\begin{equation}
\psi(x) \geq \psi(x_1) + \nabla{\psi}(x_1) \cdot (x - x_1) - z \geq \psi(x_1) - z,
\end{equation}
and therefore
\begin{align}
\left( \int_{B(x_1, \rho)} |\psi(x)|^\kappa \, dx \right)^{1/\kappa} &\geq \left( \int_{E^+} |\psi(x)|^\kappa \, dx \right)^{1/\kappa} \geq (|E^+| (\psi(x_1) - z)^\kappa)^{1/\kappa} \\
&\geq C (\psi(x_1) - z)^{n/\kappa}.
\end{align}

In conjunction with (3.6), this inequality implies that $\psi \leq C z$ on $E'$.

Next, we note that there is a point $x_2$ with $|x_2| \in (7r / 16, 9r / 16)$ and $x_2^n > 2 \mu_2 |x_2'|$ such that $\psi(x_2) \geq -C z$. In addition, if $\nabla{\psi}(x_2) \neq 0$, then
there is a positive constant $c_2$ such that $x_3 = x_2 + c_2 r \bar{Y}(x_2)/|Y(x_2)| \in E'$. Therefore

$$Cz \geq \overline{\psi}(x_3) \geq \overline{\psi}(x_2) + \bar{Y}(x_2) \cdot (x_3 - x_2) - z = \bar{\psi}(x_2) - c_2 r |\bar{Y}(x_2)| - z.$$  

It follows that $|\bar{Y}(x_2)| \leq Cz/r$ and hence

$$\psi(x) \geq \psi(x_2) + \bar{Y}(x_2) \cdot (x - x_2) - z \geq -Cz$$

for any $x \in K$.

To continue, we define $\xi$ to be the unit vector in the direction of

$$\int_0^1 g_p(t \bar{Y}(0)) \, dt,$$

so

$$g(\bar{Y}(0)) - g(0) = \int_0^1 g_p(t \bar{Y}(0)) \cdot \bar{Y}(0) \, dt \leq C \chi_0 \xi \cdot \bar{Y}(0).$$

Now set $\rho = r/(2 \xi^2)$. It is easy to see that $r/2 \leq \rho \leq C r$. In addition, we infer from our estimate $\overline{\psi} \leq Cz$ on $E'$ along with (3.2) and (3.7) that

$$Cz \geq \overline{\psi}(\rho \xi) \geq \overline{\psi}(0) + \bar{Y}(0) \cdot (\rho \xi) - z \geq -Cz + \rho \bar{Y}(0) \cdot \xi.$$

It follows that $\bar{Y}(0) \cdot \xi \leq Cz/r$, which yields (3.5). □

To state our gradient estimate for the oblique derivative problem, we use $\Gamma'_2$ to denote the set of all $(x,z,p) \in \Gamma'$ with $|z| + |p| \leq \max\{|u|_1, \Psi_1\}$. Because of the way that a Hölder gradient estimate is used to prove second derivative estimates for the oblique derivative problem without an obstacle, we first prove our estimate in a situation analogous to that in Theorem 2.3.

**Theorem 3.2.** Let $u \in W^{2,n}_{\text{loc}} \cap C^{1,0}(\overline{\Omega})$ solve (0.1) with $\partial \Omega \in C^{1,\alpha}$ for some $\alpha \in (0,1)$ and $F$ either convex or concave with respect to $r$. Suppose that there are positive constants $\lambda$, $\Lambda$, $\mu_0$, and $\nu_1$ along with a continuous, increasing function $\zeta_1$ with $\zeta_1(0) = 0$ such that conditions (2.3), (2.4), and (2.11) hold. Suppose also that there are positive constants $\chi_0$, $\mu_2$, and $\mu_3$ such that

\begin{align*}
(3.8a) \quad & G_p(x,z,p) \cdot \gamma(x) \geq \chi_0, \\
(3.8b) \quad & |G_p(x,z,p) \cdot \tau(x)| \leq \mu_2 \chi_0, \\
(3.8c) \quad & |G(x,z,p) - G(y,w,p)| \leq \mu_0 \chi_0 (|x - y| + |z - w|)^\alpha
\end{align*}

for all $(x,z,p)$ and $(y,w,p)$ in $\Gamma'_2$ and any vector field $\tau(x)$ with $\tau \cdot \gamma = 0$. Suppose further that there is a continuous increasing function $\zeta$ on $[0, \text{diam} \Omega]$ satisfying (2.12) such that $\psi$ satisfies (2.1) and

$$G(x,\psi, Y) \geq 0$$
for all \( x \in \partial \Omega \). Then there are constants \( \alpha_0(n, \mu_2, \nu_1, \Lambda/\lambda) \) and \( C \) determined only by \( n, \alpha, \Lambda/\lambda, \mu_2, \Psi_1, \zeta_1, \) and \( \Omega \) such that \( \alpha \leq \alpha_0 \) implies

\[
|Du(x_1) - Du(x_2)| \leq C[\zeta(|x_1 - x_2|) + (\mu_0 + \sup |Du|)|x_1 - x_2|^\alpha]
\]

for all \( x_1 \) and \( x_2 \) in \( \Omega \).

**Proof.** We imitate the proof of Theorem 2.2. First, we show (as in Lemma 2.1) that, if \( x_0 \in \Omega \) is a point at which \( u(x_0) = \psi(x_0) \) and if \( \overline{u} \) is defined by (2.5), then

\[
(R^{-n} \int_{B(x_0, R/2) \cap \Omega} |\overline{u}|^\kappa \, dx)^{1/\kappa} \leq C[(\mu_0 + \mu_2)R^{1+\alpha} + \zeta(R)R]
\]

for any sufficiently small \( R \) (that is, \( R \) is smaller than a constant determined only by \( \mu_2 \) and \( \Omega \)). If \( d(x_0) \geq R \), then this inequality is just (2.6). If \( d(x_0) < R \), then we first prove an estimate for \( G(x, u(x), Y(x_0)) \) by appropriate application of Lemma 3.1.

Let \( x^* \) be a closest point to \( x_0 \) in \( \partial \Omega \). By rotation and translation, we may assume that \( x^* \) is the origin and that \( x_0 \) is on the positive \( x^n \)-axis. Then \( K \subset \Omega \) provided \( \nu_0 > 1/\mu_2 \) and \( R \) is sufficiently small (determined only by \( \Omega \) and \( \mu_2 \)), and \( g(p) = G(x^*, \psi(x^*), p + Y(x_0)) \) satisfies (3.4). Next, we define \( \overline{\psi} \) by \( \overline{\psi}(x) = \psi(x) - \psi(x_0) - Y(x_0) \cdot (x - x_0) \) and we set \( \overline{Y}(x) = Y(x) - Y(x_0) \).

For \( z = C[\mu_0 R^\alpha + \zeta(R)]R \) and \( r = 2d(x_0) \), we have (3.2) directly from (2.1) because \( r \leq R \). Now, we note that using a chaining argument in the proof of [10, Theorem 9.22] allows us to replace \( B(x_0, R/2) \) by \( E \) and \( R \) by \( r \) in the proof of (2.6). Thus, we obtain

\[
\left( r^{-n} \int_E |\overline{u}|^\kappa \, dx \right)^{1/\kappa} \leq Cz,
\]

which yields (3.3) because \( \overline{u} \geq \overline{\psi} \geq -C\zeta(r)r \) in \( E \). It then follows from Lemma 3.1 that

\[
G(x^*, \psi(x^*), Y(x_0)) = g(0) \geq -Cz
\]

because \( g(\overline{Y}(0)) = G(x^*, u(x^*), Y(x^*)) \geq 0 \). For \( x \in B(x_0, R) \cap \partial \Omega \), we have

\[
|\psi(x^*) - u(x)| \leq |\psi(x^*) - \psi(x_0)| + |u(x) - u(x_0)| \leq (\Psi_1 + |Du|_0)|\alpha| R^\alpha,
\]

and therefore

\[
G(x, u(x), Y(x_0)) \geq -\mu_0 \chi_0 (\Psi_1 + |Du|_0 + 1)|\alpha| R^\alpha - C\chi_0 z \geq -Cz/R.
\]

It follows that

\[
\beta \cdot D\overline{u} \leq C\chi_0[\zeta(R) + \mu_0 R^\alpha]
\]

on \( B(x_0, R) \cap \partial \Omega \) for

\[
\beta(x) = \int_0^1 G_p(x, u, tDu + (1 - t)Y(x_0)) \, dt.
\]

**Remark.**
We then infer (3.11) by arguing as in Lemma 2.1 but with [20, Theorem 4.2] in place of [10, Lemma 9.22].

To prove the modulus of continuity estimate for \( Du \), we consider the three cases from Theorem 2.2 with \( Z = \zeta(\rho) + \mu_0 \rho^\alpha \). In addition, we set \( \Omega[y, R] = B(y, R) \cap \Omega \). In Case (i), we note that there is a cone \( Q \) with height \( \rho \), opening angle \( \theta \) (determined only by \( \Omega \)), and vertex 0 such that \( x_i + Q \subset \Omega[x_i, \rho] \) for \( i = 1, 2 \). It follows that

\[
\left( \rho^{-n} \int_{\Omega[x_i, \rho]} (|V \cdot (x - x_1)|^\kappa) \, dx \right)^{1/\kappa} \leq CZ \rho,
\]

and similar reasoning gives

\[
\left( \rho^{-n} \int_{\Omega[x_2, \rho]} (|V \cdot (x - x_2)|^\kappa) \, dx \right)^{1/\kappa} \leq CZ \rho.
\]

Combining these two estimates gives

\[
\left( \rho^{-n} \int_{Q} |V \cdot x|^\kappa \, dx \right)^{1/\kappa} \leq CZ \rho,
\]

which again implies \( |V| \leq CZ \). For Cases (ii) and (iii), we proceed as in Theorem 2.3 with [19, Lemma 13.22] in place of [19, Lemma 12.13] to prove (3.10).

The remarks from Section 2 show that this result applies to the examples from [1, 2, 12, 16, 23]. The function \( G \) given by (0.2b) satisfies conditions (3.8) by virtue of the gradient bound in [12] and \( G \) from (0.3b) clearly satisfies these conditions if \( b, g \) and \( g \) are Hölder continuous. Moreover, if \((\psi_\alpha)_{\alpha \in I} \) is a family of \( C^1 \) functions which satisfy conditions (2.1) and (3.9) with \( Y = D\psi_\alpha(x) \), then it is immediate that there is a vector field \( Y \) such that \( \psi = \sup_{\alpha \in I} \psi_\alpha \) satisfies these conditions.

We note that this result is a purely local one. Hence if the hypotheses of the theorem are satisfied only in a neighborhood \( N \) of some point \( x^* \), then we obtain a modulus of continuity estimate for the first derivatives of \( u \) in \( N' \cap \Omega \) for any compact subset \( N' \) of \( N \). The corresponding local result was proved by B. Huisken [11] although she only considered quasilinear equations and her hypotheses are stronger than ours.

In addition, we have the following result which corresponds to Theorem 2.2.

**Theorem 3.3.** Let \( u \in W^{2,n}_{\text{loc}} \cap C^1(\overline{\Omega}) \) solve (0.1) with \( \partial \Omega \in C^{1,\alpha} \) for some \( \alpha > 0 \). Suppose that there are positive constants \( \lambda, \lambda_1, \Lambda, \mu_0, \) and \( \mu_1 \) such that conditions (2.1), (2.3), (2.4), (2.7), (2.8), (3.8) are satisfied. Suppose also that \( G \in C^{1,\alpha}(\overline{\Gamma}_2') \) and that \( \zeta(t) \leq z_0 t^\alpha \) for some \( z_0 \). Then there is
a constant $C$ determined only by $n$, $z_0$, $\alpha$, $\Lambda/\lambda$, $\mu_0$, $\mu_1$, $\mu_2$, $\mu_3$, $\Psi_1$, $\zeta_1$, $\sup |Du|$, $|G|_{1,\alpha}$, and $\Omega$ such that

\begin{equation}
(3.12) \quad |Du(x_1) - Du(x_2)| \leq C(\zeta(|x_1 - x_2|) + |x_1 - x_2|).
\end{equation}

Proof. We observe that our hypotheses imply a Hölder estimate for $Du$. With this estimate, we can follow the proof of Theorem 2.2 with [29] (as modified in [19, Theorem 14.22] to deal with nonlinear boundary conditions; this step uses the Hölder gradient estimate) in place of [28].

In particular, if $\zeta(t) = t$, Theorem 3.3 gives a bound on the second derivatives of $u$.

4. The double obstacle problem.

The crucial new element in our study of double obstacle problems is a Harnack-type inequality for the difference between the upper obstacle and the lower obstacle. The basic ideas for this inequality were used in Lemma 3.1, but, here, we shall use some precise information on how fast the ratio of the maximum of the difference to its minimum goes to one on a ball of shrinking radius, provided the obstacles are defined in a ball of fixed radius. Specifically, we have the following result.

**Lemma 4.1.** Suppose $\psi_1$ and $\psi_2$ are two functions defined in $B(x_0, r)$ with $\psi_1 \leq \psi_2$ and that there are two vector fields $Y_1$ and $Y_2$ such that

\begin{align}
(4.1a) & \quad \psi_1(x_1) \geq \psi_1(x_2) + Y_1(x_2) \cdot (x_1 - x_2) - \zeta(\epsilon|x_1 - x_2|)|x_1 - x_2|, \\
(4.1b) & \quad \psi_2(x_1) \leq \psi_2(x_2) + Y_2(x_2) \cdot (x_1 - x_2) + \zeta(|x_1 - x_2|)|x_1 - x_2|
\end{align}

for all $x_1$ and $x_2$ in $B(x_0, r)$. Then for any $\epsilon \in (0, 1)$, we have

\begin{equation}
(4.2) \quad \sup_{B(x_0, \epsilon r)} (\psi_2 - \psi_1) \leq \frac{1 + \epsilon}{1 - \epsilon} \inf_{B(x_0, \epsilon r)} (\psi_2 - \psi_1) + 4\epsilon(\zeta(2\epsilon r) + \zeta(r))r.
\end{equation}

Proof. Set $\psi = \psi_2 - \psi_1$ and $I = \inf_{B(x_0, \epsilon r)} \psi$. Then choose $x_2$ so that $|x_0 - x_2| \leq \epsilon r$ and $\psi(x_2) = I$. Our first step is to show that

\begin{equation}
(4.3) \quad |Y(x_2)| \leq \frac{I}{(1 - \epsilon)r} + 2\zeta(r),
\end{equation}

so let us assume that $Y(x_2) \neq 0$ and set $\xi = Y(x_2)/|Y(x_2)|$. Then for $R < (1 - \epsilon)r$, we have that $x_2 - R\xi \in B(x_0, r)$, so (4.1) implies that

\[ \psi(x_2 - R\xi) \leq \psi(x_2) - Y(x_2) \cdot (R\xi) + 2\zeta(R)R \]
\[ = I - R|Y(x_2)| + 2\zeta(R)R. \]

We infer (4.3) from this inequality by sending $R \to (1 - \epsilon)r$ and noting that $\psi(x_2 - R\xi) \geq 0$. 


Now let $x \in B(x_0, \varepsilon r)$ and use (4.1) to infer that
\[
\psi(x) \leq \psi(x_2) + Y(x_2) \cdot (x - x_2) + 2(|x - x_2||x - x_2|) \leq I + 2\varepsilon r|Y(x_2)| + 4\zeta(2\varepsilon r)\varepsilon r.
\]
Simple algebra then completes the proof.

We also shall use the following simple variant of (4.2).

**Corollary 4.2.** In addition to the hypotheses of Lemma 4.1, suppose that $\psi_1(x_0) = 0$ and $Y_1(x_0) = 0$. Then

\[
\sup_{B(x_0, \varepsilon r)} \psi_2 \leq \frac{1 + \varepsilon}{1 - \varepsilon} \inf_{B(x_0, \varepsilon r)} \psi_2 + \frac{4\varepsilon}{1 - \varepsilon} \left( \zeta(2\varepsilon r) + \zeta(r) \right) r.
\]

If also $\psi_2 \geq 0$ on $B(x_0, \varepsilon r)$, then

\[
\sup_{B(x_0, \varepsilon r)} \psi_1 \leq \frac{4\varepsilon}{1 - \varepsilon^2} \inf_{B(x_0, \varepsilon r)} \psi_2 + \frac{6\varepsilon}{1 - \varepsilon^2} \left( \zeta(2\varepsilon r) + \zeta(r) \right) r.
\]

**Proof.** Because $\psi_2 \geq -\zeta(r)r$, we can follow the proof of Lemma 4.1 with $\psi$ replaced by $\psi_2 + \zeta(r)r$ to infer that

\[
\sup_{B(x_0, \varepsilon r)} (\psi_2 + \zeta(r)r) \leq \frac{1 + \varepsilon}{1 - \varepsilon} \inf_{B(x_0, \varepsilon r)} (\psi_2 + \zeta(r)r) + 2\varepsilon(\zeta(2\varepsilon r) + \zeta(r))r
\]

\[
\leq \frac{1 + \varepsilon}{1 - \varepsilon} \inf_{B(x_0, \varepsilon r)} (\psi_2 + \zeta(r)r) + 2\varepsilon(\zeta(2\varepsilon r) + \zeta(r))r,
\]

so

\[
\sup_{B(x_0, \varepsilon r)} \psi_2 \leq \frac{1 + \varepsilon}{1 - \varepsilon} \inf_{B(x_0, \varepsilon r)} \psi_2 + \left( \frac{1 + \varepsilon}{1 - \varepsilon} - 1 \right) \zeta(r)r + 2\varepsilon(\zeta(2\varepsilon r) + \zeta(r))r.
\]

Since $(1 + \varepsilon)/(1 - \varepsilon) - 1 = 2\varepsilon/(1 - \varepsilon)$ and $1 < 1/(1 - \varepsilon)$, this inequality gives (4.4).

Next, we set $\psi_1 = \psi_2 - \psi_1$, $I = \inf_{B(x_0, \varepsilon r)} \psi$ and $I_2 = \inf_{B(x_0, \varepsilon r)} \psi_2$ to see that

\[
I_2 \leq \psi_2(x_0) = \psi(x_0) \leq \sup_{B(x_0, \varepsilon r)} \psi \leq \frac{1 + \varepsilon}{1 - \varepsilon} I + 4\varepsilon(\zeta(2\varepsilon r) + \zeta(r))r
\]

and hence

\[
\sup_{B(x_0, \varepsilon r)} \psi_1 \leq \sup_{B(x_0, \varepsilon r)} \psi_2 - I
\]

\[
\leq \left( \frac{1 + \varepsilon}{1 - \varepsilon} - \frac{1 - \varepsilon}{1 + \varepsilon} \right) I_2 + \left( \frac{4\varepsilon}{1 + \varepsilon} + \frac{2\varepsilon}{1 - \varepsilon} \right) (\zeta(2\varepsilon r) + \zeta(r))r.
\]

The desired inequality follows from this one by simple algebra. \qed
These lemmata allow us to imitate the argument in [18, Lemma 1.1] to prove an analog of Lemma 2.1 when \( u \) is a solution of the double obstacle problem:

\[
\begin{align*}
\psi_1 \leq u \leq \psi_2 & \text{ in } \Omega, \\
\min\{-F(x, u, Du, D^2u), u - \psi_1\} = 0 & \text{ if } u < \psi_2, \\
\min\{F(x, u, Du, D^2u), \psi_2 - u\} = 0 & \text{ if } u > \psi_1.
\end{align*}
\]

**Lemma 4.3.** Suppose \( u, \psi_1, \) and \( \psi_2 \) are as above, and suppose that there are positive constants \( \lambda, \Lambda, \) and \( \mu_0 \) such that conditions (2.3), (2.4b), and

\[
|F(x, u, Du, 0)| \leq \mu_0 \lambda \tag{4.7}
\]

are satisfied. If \( x_0 \) is a point such that \( u(x_0) = \psi_1(x_0) \), then there are positive constants \( C, \delta, \) and \( \kappa \) determined only by \( n \) and \( \Lambda/\lambda \) such that \( \pi \), defined by

\[
\pi(x) = u(x) - u(x_0) - Y_1(x_0) \cdot (x - x_0),
\]

satisfies the estimate

\[
\left( R^{-n} \int_{B(x_0, \delta R)} |\pi|^{\kappa} \, dx \right)^{1/\kappa} \leq C[\mu_0 R^2 + \zeta(R)R] \tag{4.9}
\]

for all \( R \leq d(x_0) \).

**Proof.** We first note that the hypotheses of this lemma are unchanged if we subtract the same linear function from \( u, \psi_1 \) and \( \psi_2 \), so we may assume that \( \psi_1(x_0) = 0 \) and \( Y_1(x_0) = 0 \). Then, for \( \varepsilon \in (0, 1) \) to be chosen, we set

\[
I_2 = \inf_{B(x_0, \varepsilon R)} \psi_2.
\]

If \( I_2 \leq 12\zeta(R)R + \mu_0 R^2 \), then (4.4) implies that \( \psi_2 \leq C(\varepsilon)[\mu_0 R^2 + \zeta(R)R] \) in \( B(x_0, \varepsilon R) \), and (4.9) follows for any \( \delta \leq \varepsilon \).

On the other hand, if \( I_2 > 12\zeta(R)R + \mu_0 R^2 \), we set

\[
M = (1 - \varepsilon)I_2,
\]

\[
M_1 = \frac{5\varepsilon}{1 - \varepsilon^2} I_2 + 2\zeta(R)R,
\]

\[
U = \min\{u, M\} + M_1,
\]

and we note that \( U \geq 0 \). Now, for \( \eta > 0 \), define \( f_\eta \) by

\[
f_\eta(t) = (\max\{t, 0\}^3 + \eta^3)^{1/3} - \eta,
\]

and set

\[
U_\eta = M + M_1 - f_\eta(M - u).
\]

Then \( U_\eta \to U \) uniformly as \( \eta \to 0 \) and \( U_\eta \geq 0 \) in \( B(x_0, \varepsilon R) \). Moreover, because \( a^{ij}D_{ij}u \leq \lambda \mu_0 \) wherever \( u \leq M \) and \( f_\eta \) is \( C^2 \) with \( f''_\eta \geq 0 \), it follows
that \( \sigma^{ij}D_{ij}\eta \leq \lambda \mu_0 \). It follows from the weak Harnack inequality [10, Theorem 9.22] that

\[
\left( (\varepsilon R)^{-n} \int_{B(x_0,\varepsilon R/2)} |U\eta|^{1/\kappa} \, dx \right)^{1/\kappa} \leq C_1 \left[ \inf_{B(\varepsilon R/2)} U\eta + \mu_0 (\varepsilon R)^2 + \zeta(\varepsilon R)(\varepsilon R) \right]
\]

for some \( C_1(n, \Lambda/\lambda) \) and \( \kappa(n, \Lambda/\lambda) \). Sending \( \eta \to 0 \), we infer that

\[
(4.10) \quad \left( (\varepsilon R/2)^{-n} \int_{B(x_0,\varepsilon R/2)} |U|^{1/\kappa} \, dx \right)^{1/\kappa} \leq C_1[M_1 + \varepsilon \mu_0 R^2 + \varepsilon \zeta(\varepsilon R)R]
\]

because \( U(x_0) = M_1 \).

Next, we set \( M_2 = \sup_{B(x_0,\varepsilon R)} \psi_1 + \zeta(\varepsilon R)R \) and \( V = \max\{u, M_2\} \). By a similar approximation argument, we infer from the local maximum principle [10, Theorem 9.20] that there is a constant \( C_2(n, \Lambda/\lambda) \) so that

\[
\sup_{B(x_0,\varepsilon R/4)} V \leq C_2 \left[ \left( \frac{(\varepsilon R)^{-n}}{2} \int_{B(x_0,\varepsilon R/2)} V^{\kappa} \, dx \right)^{1/\kappa} + \mu_0 (\varepsilon R)^2 + \zeta(\varepsilon R)\varepsilon R \right].
\]

Now we note that \( u \leq U \) (because \( M_1 \geq 0 \) and \( M + M_1 \geq \sup_{B(x_0,\varepsilon R)} \psi_2 \) and \( M_2 \leq M_1 - \zeta(\varepsilon R)R \) (because \( M_2 \leq 5I_2 \varepsilon/(1 - \varepsilon^2) \)), so \( U_2 \leq U_1 \) provided \( \varepsilon \leq 1/2 \). It follows that

\[
\sup_{B(x_0,\varepsilon R/4)} u \leq C_2(C_1[M_1 + \mu_0 \varepsilon R^2 + \varepsilon \zeta(\varepsilon R)R] + \mu_0 \varepsilon R^2 + \varepsilon \zeta(\varepsilon R)R)
\]

\[
\leq C_1 C_2 \frac{2\varepsilon}{1 - \varepsilon} \psi_2(x_2) + \left( \frac{5C_1}{1 - \varepsilon^2} + C_1 + 1 \right) C_2 \varepsilon I_2.
\]

By taking \( \varepsilon \) sufficiently small, we conclude that \( u < \psi_2 \) on \( B(x_0,\varepsilon R/4) \). Therefore, \( u + \zeta(\varepsilon R)R \) is a positive supersolution on \( B(x_0,\varepsilon R/4) \) and we can use the weak Harnack inequality directly to infer (4.9) with \( \delta = \varepsilon/8 \). \( \square \)

Note that the arguments of Lemmata 3.1 and 4.3 can be combined to prove pointwise decay of \( \bar{\eta} \) near a contact point. Specifically, suppose \( u \) satisfies (2.2) with \( F \) satisfying (2.3), (2.4b), and (4.7). If \( u(x_0) = \psi(x_0) \), then (2.6) and the proof of \( \bar{\eta} \leq Cz \) in \( E' \) (from Lemma 3.1) give a constant \( c_1 \) such that \( \bar{\eta} \leq c_1[\zeta(\varepsilon R) + \mu_0 R]R \) in \( B(R) \). The local maximum principle applied to \( \max\{\bar{\eta}, (1 + c_1)[\zeta(\varepsilon R) + \mu_0 R]R \} \) then yields \( \bar{\eta} \leq C[\zeta(\varepsilon R) + \mu_0 R]R \) in \( B(R/4) \). With this pointwise estimate in hand, we can imitate the proof of [4, Theorem 2.3] to obtain a modulus of continuity estimate for solutions of obstacle problems with linear equations when \( \zeta \) does not necessarily satisfy condition (2.7).
For our purposes, the next important step is to obtain a modulus of continuity for $Du$.

**Theorem 4.4.** Suppose that $u$, $\psi_1$, $\psi_2$, and $F$ satisfy conditions (4.6), (4.7), (2.3), and (2.4b) with $\zeta$ a continuous increasing function on $[0, \text{diam}\Omega]$. Suppose also that there are constants $\alpha$, $\nu_0$ and $\nu_1$ along with a continuous increasing function $\zeta_1$ with $\zeta_1(0) = 0$ such that conditions (2.11) and (2.12) hold. If $F$ is convex or concave with respect to $r$, then there are constants $\alpha_0(n, \Lambda/\lambda, \nu_1)$ and $C$ determined only by $n$, $\alpha$, $\nu_1$, $\Lambda/\lambda$, $|u|_1$, and $\text{diam}\Omega$ such that $\alpha_0 \leq \alpha_0$ implies (3.12) for all $x_1$ and $x_2$ in $\Omega$ with $|x_1 - x_2| \leq \frac{1}{2} \min\{d(x_1), d(x_2)\}$.

Of course, the double obstacle analog of Theorem 2.2 holds with $\alpha = 1$ if condition (2.11) is replaced by (2.8).

We can use the same ideas for oblique derivative problems, but the proofs are more complicated. In place of Lemma 4.1, we have a similar, but more subtle, inequality. To state our results more simply, we let $\omega$ be a $C^1$ function in some $(n - 1)$-dimensional ball $B(0, R)$ with $R > 0$ and $\omega(0) = 0$, and we set $\omega_0 = \sup |D\omega|$. We also define

\[ K[r] = \{ x \in \mathbb{R}^n : x^n < r - (\omega_0 + 1)|x'|, x^n > \omega(x') \}, \]

\[ \Sigma[r] = \{ x \in \mathbb{R}^n : x^n < r - (\omega_0 + 1)|x'|, x^n = \omega(x') \} \]

for $r \in (0, R)$, where here and below we abbreviate $x' = (x^1, \ldots, x^{n-1})$.

**Lemma 4.5.** Let $\psi_1$ and $\psi_2$ be two functions defined in $K[r] \cup \Sigma[r]$ for some $r \in (0, R)$ with $\psi_1 \leq \psi_2$ there. Suppose that there are vector fields $Y_1$ and $Y_2$ such that conditions (4.1) hold for all $x_1$ and $x_2$ in $K[r] \cup \Sigma[r]$. Suppose also that there are positive constants $\alpha \leq 1$, $\mu_3 < 1/\omega_0$, $\mu_3$, and $\chi_0$ along with a function $G$ such that

\[(4.11a) \quad \frac{\partial G}{\partial p'}(x, z, p) \leq \mu_3 \chi_0, \]

\[(4.11b) \quad G^n(x, z, p) \geq \chi_0, \]

\[(4.11c) \quad |G(x, z, p) - G(x, w, p)| \leq \mu_0 \chi_0 |w - z|^\alpha \]

for any $(x, z, p) \in \Sigma[r] \times \mathbb{R} \times \mathbb{R}^n$ and any $w \in \mathbb{R}$, and set

\[(4.12) \quad g_0 = \frac{1}{\chi_0} \inf_{K[r]} (G(x, \psi_2(x), Y_2(x)) - G(x, \psi_1(x), Y_1(x)))^+. \]

Then, for any $\varepsilon \in (0, 1)$, there is a constant $\eta(\varepsilon, \mu_3, \omega_0)$ such that

\[(4.13) \quad \sup_{K[r]} (\psi_2 - \psi_1) \leq (1 + \varepsilon) \inf_{K[r]} (\psi_2 - \psi_1) + 3\varepsilon \zeta(r)\varepsilon + C_1 \varepsilon r^{2/(2-\alpha)} + \varepsilon g_0\]

for $C_1 = 3(\mu_0 \sup |\psi_2 - \psi_1|^{\alpha/2})^{2/(2-\alpha)}$. 
Proof. To simplify the notation, we shall set \( \psi = \psi_2 - \psi_1, Y = Y_2 - Y_1, \) and 
\( K = K[\eta r]. \) In addition, we write \( I \) for the infimum of \( \psi \) over \( K \) and we let \( x_2 \) be a point in the closure of \( K \) at which \( \psi(x_2) = I. \) We now consider several cases.

Suppose first that \( x_2 \in \Sigma[r]. \) Then we infer from (4.11) and (4.12) that

\[
g_0 \chi_0 \geq -\mu_0 \chi_0 |\psi(x_2)|^\alpha + v \cdot Y(x_2)
\]

for some vector \( v \) with \(|v'| \leq \mu_3 v^n \) and \( v^n \geq \chi_0. \) Now we set \( \xi = v/|v| \) and we set

\[
x_1 = x_2 + \frac{\varepsilon}{8} r \xi, \quad x_3 = x_2 + \frac{1}{2} r \xi.
\]

If \( \eta < 1/2, \) it follows that \( x_1 \) and \( x_3 \) are in \( \Omega[r]. \) Setting \( I_1 = \psi(x_1), \) we see that

\[
I_1 \leq I + Y(x_2) \cdot (x_1 - x_2) + \frac{\varepsilon}{8} \zeta(r)r = I + \frac{\varepsilon T}{8} Y(x_2) \cdot \xi + \frac{\varepsilon}{8} \zeta(r)r
\]

\[
\leq I + \frac{\mu_0 \varepsilon}{8} I^\alpha r + \frac{\varepsilon}{8} \zeta(r)r + \frac{\varepsilon}{8} g_0 r
\]

\[
\leq \left(1 + \frac{\varepsilon}{2}\right) I + \frac{C_1}{3} \varepsilon r^{2/(2-\alpha)} + \frac{\varepsilon}{8} \zeta(r)r + \frac{\varepsilon}{8} g_0 r
\]

by virtue of (4.14) and Young’s inequality. Now we obtain two estimates for \( Y(x_1). \) First, there is a constant \( k(\omega_0, \mu_3) \) such that \( B(x_1, k\varepsilon r) \subset K[r] \) and then the proof of (4.3) with \( \varepsilon = 1/2 \) shows that

\[
|Y(x_1)| \leq \frac{2I_1}{k\varepsilon r} + 4 \zeta(r) \leq \frac{3}{2\varepsilon kr} I + \frac{2C_1}{3k} r^{\alpha/(2-\alpha)} + \left(\frac{2}{k + 4}\right) \zeta(r) + \frac{g_0}{4k}.
\]

Moreover,

\[
0 \leq \psi(x_3) \leq I_1 + \left(\frac{1}{2} - \frac{\varepsilon}{8}\right) rY(x_1) \cdot \xi + \zeta(r)r
\]

because \( x_3 - x_1 = (1/2 - \varepsilon/8)r \xi \) and hence

\[
-rY(x_1) \cdot \xi \leq \frac{1 + \varepsilon/2}{1/2 - \varepsilon/8} I + 8 \zeta(r)r + \frac{4}{3} C_1 \varepsilon r^{2/(2-\alpha)} + \frac{\varepsilon}{2} g_0 r.
\]

Now we note that, for any \( x \in K, \) we have \(|x| < 2\eta r, \) and hence

\[
\psi(x) \leq \psi(x_1) + Y(x_1) \cdot (x - x_1) + 4\eta \zeta(4\eta r)r.
\]
To analyze the right hand side of this inequality, we first observe that \( x - x_1 = (x - x_2) + (x_2 - x_1) \) and that \( |x - x_2| \leq 4\eta r \). It follows that

\[
Y(x_1) \cdot (x - x_1) \\
\leq -\frac{\varepsilon}{8} r Y(x_1) \cdot \xi + 4\eta r |Y(x_1)| \\
\leq \left( (1 + \varepsilon/2) \frac{\varepsilon/8}{1/2 - \varepsilon/8} + \frac{6\eta}{2\varepsilon k} \right) I + \left( \left( \frac{2}{k} + 4 \right) 2\eta + \varepsilon \right) \zeta(r)r \\
+ C_1 \left( \frac{\varepsilon^2}{3} + \frac{8\eta}{3k} \right) r^{2/(2-\alpha)} + \left( \frac{\varepsilon}{2} + \frac{2\eta}{2k} \right) g_0 r,
\]

and

\[
\psi(x) \leq \left( (1 + \varepsilon/2) \frac{1/2}{1/2 - \varepsilon/8} + \frac{6\eta}{k\varepsilon} \right) I + \left( \frac{2 \varepsilon}{k} + 6 \right) 2\eta \zeta(r)r + \left( \frac{\varepsilon}{2} + \frac{2\eta}{2k} \right) g_0 r
\]

provided \( 4\eta \leq 1 \). By simple calculation, \((1+\varepsilon/2)(1/2)/((1/2)-\varepsilon/8) < 1+\varepsilon\), so we can take \( \eta \) sufficiently small to infer (4.13) in this case. If \( x_2 \notin \Sigma[r] \), then \( Y^n(x_2) \geq 0 \) and we can imitate the calculations of the preceding case with \( \xi = (0, \ldots, 0, 1) \), to see that

\[
\psi(x) \leq \left( \frac{1/2}{1/2 - \varepsilon/8} + \frac{4\eta}{k\varepsilon} \right) I + \left[ \left( \frac{1}{2k} + 8 \right) 2\eta + \left( \frac{5/8}{1/2 - \varepsilon/8} + 1 \right) \varepsilon \right] \zeta(r)r,
\]

which implies (4.13) if \( \eta \) is sufficiently small.

Our next step is to prove a corresponding estimate for our general geometric situation.

**Lemma 4.6.** Let \( \psi_1 \) and \( \psi_2 \) be two functions defined in \( \overline{\Omega} \) with \( \psi_1 \leq \psi_2 \). Suppose conditions (4.1) and (3.8) are satisfied. Let \( x_0 \in \Omega \) and set

\[
g_1(r) = \frac{1}{\chi_0} \sup_{\partial \Omega[r]} (G(x, \psi_2(x), Y_2(x)) - G(x, \psi_1(x), Y_1(x)))^+.
\]

If \( \partial \Omega \in C^1 \), then for any \( \varepsilon > 0 \), there are constants \( R(\mu_2, \Omega), \delta(\varepsilon, \mu_2, \Omega) \) and \( C(\mu_2, \mu_0, \alpha, \sup(\psi_2 - \psi_1)) \) such that

\[
\sup_{\Omega[\delta r]} (\psi_2 - \psi_1) \leq (1 + \varepsilon) \inf_{\Omega[\delta r]} (\psi_2 - \psi_1) + C \varepsilon \zeta(r)r + r^{2/(2-\alpha)} + \varepsilon g_1(r)r
\]

for any \( x_0 \in \Omega \) and \( r \in (0, R) \).

**Proof.** Let \( x_1 \) be a closest point to \( x_0 \) in \( \partial \Omega \), which we can take to be the origin, and rotate axes so that \( x_0' = 0 \) and \( x_0'' > 0 \). Then there is a
constant $R_1$ determined only by $\mu_2$ and $\Omega$ so that there is a function $\omega$ with $\mu_2 \sup |D\omega| < 1/2$ such that

$$\Omega[R] = \{ x \in \mathbb{R}^n : |x - x_0| < R, x^n > \omega(x') \}$$

and $\omega(0) = 0$ and $D\omega(0) = 0$. By choosing $R < R_1$ sufficiently small, we can also arrange that $|G^n - G_p \cdot \gamma| \leq \frac{1}{2} \chi_0$ and $|D\omega| < 1/2$. It follows that conditions (4.11a,b) hold with $\mu_3 = 2\mu_2$.

Now take $\eta$ to be the constant from Lemma 4.5 and note that there is a constant $\eta_1$ such that $d(x_0) \leq \eta_1 r$ implies that $\Omega[\eta_1 r] \subset K[\eta r]$. Therefore (4.16) holds in this case with any $\delta \leq \eta_1$. On the other hand, if $d(x_0) > \eta_1 r$, then Lemma 4.1 (with $\eta_1 r$ in place of $r$) implies (4.16) in this case with $\delta = \eta_1 \varepsilon/3$ because $(1 + \varepsilon/3)/(1 - \varepsilon/3) \leq 1 + \varepsilon$. Combining this two cases yields the desired result with $\delta = \eta_1 \varepsilon/2$. \hfill \Box

As before, we then have the following estimates.

**Corollary 4.7.** In addition to the hypotheses of Lemma 4.6, suppose that $\psi_1(x_0) = 0$ and $Y_1(x_0) = 0$, and set

$$g_2(r) = \frac{1}{\lambda_0} \sup_{\partial \Omega[r]} (G(x, \psi_2(x), Y_2(x)) - \min\{G(x, 0, 0), G(x, \psi_1(x), Y_1(x))\})^+.$$  

Then

$$\sup_{B(x_0, \delta r)} \psi_2 \leq (1 + \varepsilon) \inf_{\Omega[\delta r]} \psi_2^+ + C\varepsilon[\zeta(r) + r^{\alpha/(2-\alpha)} + g_2(r) + (\zeta(r)r)^\alpha]r.$$  

If also $\psi_2 \geq 0$ in $\Omega[\delta r]$, then

$$\sup_{B(x_0, \delta r)} \psi_1 \leq 2\varepsilon \inf_{\Omega[\delta r]} \psi_2 + C\varepsilon[\zeta(r) + r^{\alpha/(2-\alpha)} + g_2(r) + (\zeta(r)r)^\alpha]r.$$  

**Proof.** We follow the proof of Corollary 4.2, noting that

$$G(x, -\zeta(r)r, 0) \geq G(x, 0, 0) - \mu_0 (\zeta(r)r)^\alpha$$

and that $(1 + \varepsilon) - 1/(1 + \varepsilon) \leq 2\varepsilon$. \hfill \Box

The estimate on the modulus of continuity for the gradient of the solution of the double obstacle problem follows easily.

**Theorem 4.8.** Let $\partial \Omega \in C^{1,\alpha}$ for some $\alpha \in (0, 1)$, and let $\psi_1$ and $\psi_2$ be two functions satisfying condition (4.1) in $\Omega$ for some continuous increasing function $\zeta$. Suppose also that $\psi_1 \leq \psi_2$ in $\Omega$. Let $u \in W^{2,\alpha}_{\text{loc}} \cap C^1(\Omega)$ satisfy (4.6) and $G(x, u, Du) = 0$ on $\partial \Omega$. Suppose there are constants $\lambda$, $\mu_0$, $\mu_2$, and $\nu_1$, along with a continuous increasing function $\zeta_1$ with $\zeta_1(0) = 0$ such
that conditions (2.11), (2.12), (2.3), (2.4b), (4.7), and (3.8) hold. Suppose finally that

\[ G(x, \psi_1(x), Y_1(x)) \geq 0, \quad G(x, \psi_2(x), Y_2(x)) \leq 0, \]

for all \( x \in \partial \Omega \). If \( F \) is concave or convex with respect to \( r \), then there are constants \( \alpha_0(n, \mu_2, \Lambda/\lambda, \nu_1) \) and \( C(n, \alpha, \Lambda/\lambda, \mu_1, \mu_2, |D\psi_1|_1, |D\psi_2|_1, \zeta_1, \Omega) \) such that if \( \alpha \leq \alpha_0 \), then (3.10) holds for all \( x_1 \) and \( x_2 \) in \( \Omega \).

**Proof.** We proceed by combining the proof of Theorem 3.2 with that of Theorem 4.4, taking Corollary 4.7 into account. \( \square \)

We omit the obvious two-obstacle analog of Theorem 3.3.

### 5. Variational inequalities.

Our methods also apply to certain types of variational inequalities. In particular, let \( H \) be a convex, \( C^2 \) function defined on \([0, \infty)\) with \( H(0) = 0 \) and suppose that \( h = H' \) satisfies the conditions

\[ \delta \leq \frac{th'(t)}{h(t)} \leq g_0 \]

for some positive constants \( \delta \) and \( g_0 \), and all \( t > 0 \); we also assume that \( H(1) = 1 \) for simplicity. The model such function is \( H(t) = t^m \) with \( m > 1 \). Let \( W^{1,H} \) denote the set of all functions \( v \in W^{1,1} \) with \( H(|Dv|) \in L^1(\Omega) \), and write \( K \) for a convex subset of \( W^{1,H} \) such that \( v \geq \psi \) for all \( v \in K \).

(For example if \( H(t) = t^m \), then \( W^{1,H} = W^{1,m} \) and we can take \( K \) to be the set of all \( v \in W^{1,m} \) with \( v \geq \psi \).) We then consider the problem of finding a function \( u \in K \) such that

\[ \int_\Omega [A(x, u, Du) \cdot D(u - v) - B(x, u, Du)(u - v)] \, dx \leq 0 \]

for all \( v \in K \), where \( A \) is a vector-valued function (for example \( A(x, z, p) = h(|p|)p/|p| \)) and \( B \) is a scalar-valued function, which we shall assume to be bounded. Such problems have a long history for various choices of \( h \) provided \( A \) and \( B \) satisfy suitable structure conditions; see, for example, [14, Section III.4], [18], [6], [8], [9], [22], [25], [27]. We note, however, that all of these works assume that \( \psi \) has Hölder continuous gradient when trying to prove a modulus of continuity estimate for the gradient of \( u \).

We first observe that, when \( h(t)/t \) is bounded from above and below by positive constants and \( A \) and \( B \) are sufficiently smooth, smooth solutions of this variational inequality are also solutions of (2.2) with

\[ F(x, z, p, r) = \frac{\partial A^i}{\partial p_j}(x, z, p)r_{ij} + \frac{\partial A^i}{\partial z}(x, z, p)p_i + \frac{\partial A^i}{\partial x^j}(x, z, p) + B(x, z, p). \]
More generally, we assume that $A$ is differentiable with respect to $p$ and that there are nonnegative constants $\alpha$, $\Lambda$ and $\Lambda_1$ with $\alpha \in (0, 1)$ such that

\begin{align}
(5.3\text{a}) & \quad \frac{\partial A^i}{\partial p_j} \xi_i \xi_j \geq \frac{h(|p|)}{|p|} |\xi|^2, \\
(5.3\text{b}) & \quad |A_p| \leq \Lambda h(|p|)/|p|, \\
(5.3\text{c}) & \quad |B| \leq \Lambda_1, \\
(5.3\text{d}) & \quad |A(x, z, p) - A(y, w, p)| \leq \Lambda_1(|x - y| + |w - z|)^\alpha.
\end{align}

These conditions were studied extensively in [17]. In fact, we have simplified the conditions there somewhat by assuming a known bound for the gradient of $u$.

We begin by proving an estimate like (2.6). As in [18], we first prove the estimate for a simpler problem.

**Lemma 5.1.** Let $A$ be a vector valued function defined on $\mathbb{R}^n$ and suppose that there are positive constants $\delta$, $g_0$, and $\Lambda$ along with a function $h$ such that conditions (5.1) and (5.3a,b) are satisfied. Let $u$ and $\psi$ be in $C^0_0(B(x_0, r))$ for some ball $B(x_0, r)$ with $u \geq \psi$, and let $K$ be the set of all $v$ with $v - u \in W_0^{1, H}(B(x_0, r))$ and $v \geq \psi$ in $B(x_0, r)$. Then there is a unique solution $U$ of the variational inequality

\begin{equation}
(5.4) \quad \int_{B(x_0, r)} A(DU) \cdot D(U - v) \, dx \leq 0 \quad \text{for all } v \in K,
\end{equation}

and there are constants $C_1(n, \delta, g_0, |Du|_0, |D\psi|_0, \Lambda)$, $C_2(\Lambda, n)$, $\theta(\Lambda, n, \delta, g_0)$, and $\kappa(\Lambda, n)$ such that

\begin{equation}
(5.5) \quad [U]_{\theta; B(x_0, r)} \leq C_1 r^{1-\theta}
\end{equation}

and

\begin{equation}
(5.6) \quad \left( r^{-n} \int_{B(x_0, r/2)} |U - L|^\kappa \, dx \right)^{1/\kappa} \leq C_2 \inf_{B(x_0, r/2)} (U - L)
\end{equation}

for any linear function $L$ such that $U - L \geq 0$ in $B(x_0, r)$.

**Proof.** The standard theory of variational inequalities gives the existence and uniqueness of $U$. In addition, (5.5) follows from the arguments in [18, Lemma 1.3].

To prove (5.6), we proceed by approximation. First, we fix $\alpha \in (0, 1)$ and note (from the proof of [17, Lemma 5.2]) that there is a sequence of $C^{1, \alpha}$ functions $(A_k)$ which converge uniformly to $A$ on compact subsets of $B(x_0, r)$ and which satisfy

\begin{equation}
\frac{\partial A_k^i}{\partial p_j} \xi_i \xi_j \geq \frac{h_k(|p|)}{|p|} |\xi|^2
\end{equation}
and

$$\left| \frac{\partial A_k^i}{\partial p_j} \right| \leq 2\Lambda h_k(|p|)/|p|$$

for functions $h_k$ satisfying (5.1). For $\varepsilon \in (0,1)$, define $\beta_\varepsilon$ by $\beta_\varepsilon(t) = (\min\{t,0\})^2/\varepsilon$, and let $U_k$ solve $\nabla A_k(DU_k) + \beta_1/k(U_k - \psi) = 0$ in $B(x_0,r)$ and $U_k = u$ on $\partial B(x_0,r)$. The existence of a unique solution to this problem is straightforward, and classical regularity theory implies that $U_k \in C^2(B(x_0,r))$. Thus, the weak Harnack inequality [10, Theorem 9.22] implies that

$$\left( r^{-n} \int_{B(x_0,r/2)} |U_k - L|^\kappa \, dx \right)^{1/\kappa} \leq C(\Lambda, n) \inf_{B(x_0,r/2)} (U_k - L)$$

for any linear function $L$ with $U_k - L \geq 0$ in $B(x_0,r)$. It is not hard to show that $U_k$ converges uniformly to $U$ as $k \to \infty$ (see, for example, [14, Theorem IV.5.2]), so the desired result follows immediately.

From this lemma and a suitable choice for the linear function $L$, we infer a version of (2.6).

**Lemma 5.2.** Under the hypotheses given before Lemma 5.1, suppose that $\psi$ satisfies condition (2.1) for some continuous increasing function $\zeta$. Suppose $u(x_0) = \psi(x_0)$ and define $\bar{u}$ by (2.5). If $\kappa$ and $\theta$ are the constants from Lemma 5.1 and if $r \leq 1$, then there is a constant $C$ determined only by $\Lambda$, $n$, $A_1$, $|Du|_0$ such that

$$r^{-n} \int_{B(x_0,r/2)} |\bar{u}|^\kappa \, dx \leq C \left[ \frac{r^{\alpha/2 + 2\theta}}{\Lambda} + \zeta(r) \right] r.$$

**Proof.** Let $U$ be the solution of (5.4) given by Lemma 5.1 with $A(p) = A(x_0,u(x_0),p)$ and set $w = u - U$. Then we can use $v = U$ in (5.2) and $v = u$ in (5.4) to see from (17), (5.8) and Lemma 2.2 that

$$\int_{B(x_0,r)} H(|w|/r) \, dx \leq C \int_{B(x_0,r)} H(|Du|) \, dx \leq Cr^{n+\alpha/2}$$

because $H$ is convex. Then Jensen’s inequality gives

$$\int_{B(x_0,r)} |w| \, dx \leq Cr^{n+1+\alpha/2}$$

because [17, Lemma 1.1(c)] says that $H(r^{\alpha/2})/H(1) \leq r^{\alpha/2}/1$.

To continue, we use a variation of the argument in Lemma 1.1. Choose $x_1$ so that $d(x_1)\{w(x_1)\} \geq (1/2)|w|_0^{(n)}$ and set $\rho = \varepsilon d(x_1)$ with $\varepsilon \in (0,1/2)$. We have from (5.5) that

$$|w(x)| \geq |w(x_1)| - |w(x) - w(x_1)| \geq \frac{1}{2} d(x_1)^{-n} |w|_0^{(n)} - cr^{-\theta} \rho^\theta$$
for $|x - x_1| \leq \rho$, so
\[
\int_{B(x_0,r)} |w| \, dx \geq \int_{B(x_1,\rho)} |w| \, dx \geq \frac{\omega_n}{2} \varepsilon^n |w|_0^{(n)} - c\omega_n r^{1-\theta} \rho^{\alpha + \theta},
\]
where $\omega_n$ is the measure of the $n$-dimensional unit ball. Therefore
\[
|w|_0^{(n)} \leq C\varepsilon^{-n} \int_{B(x_0,r)} w \, dx + C\varepsilon^\theta d_m^\alpha r^{1-\theta} \leq Cr^{n+1}(\varepsilon^{-n} r^{\alpha/2} + \varepsilon^\theta)
\]
from (5.8). Now we take $\varepsilon = r^{\alpha/(2n+2\theta)}/2$ to conclude that
\[
\sup_{B(x_0, r/2)} |w| \leq cr^{-n} |w|_0^{(n)} \leq Cr^{1+\alpha/(2n+2\theta)}.
\]
Thus we can take $L(x) = \psi(x_0) - Y(x_0) \cdot (x - x_0) - \zeta(r)r - Cr^{1+\alpha/(2n+2\theta)}$ in Lemma 5.2, and hence (5.7) holds.

The interior gradient modulus of continuity estimate for such problems follows by using the argument of Theorem 2.3 and the Hölder gradient estimates for weak solutions of divergence structure equations from [17, Section 5]. The correct form of this estimate is an easy consequence of the last inequality on page 346 of [17].

**Theorem 5.3.** Let $A$ and $B$ be, respectively, a vector-valued function and a scalar-valued function on $\Omega \times \mathbb{R} \times \mathbb{R}^n$, and let $H$ be a convex, $C^2$ function on $[0, \infty)$ with $H(0) = 0$, and suppose $h = H'$ satisfies (5.1). Suppose also that conditions (5.3a–d) are satisfied. Let $\psi$ satisfy (2.1), let $u \in C^{0,1}(\Omega)$ and suppose $u \geq \psi$ in $\Omega$. If $u$ is a solution of (5.2) with $K$ the set of all $v \in W^{1,H}$ with $v - u \in W_0^{1,H}$ and $v \geq \psi$, then there are constants $\sigma_0(\Lambda, \delta, g_0)$ and $C(n, \Lambda, \delta, g_0, \sup |Du|, \Psi_1, \Lambda_1, \alpha, \text{diam} \Omega)$ such that
\[
(5.9) \quad |Du(x_1) - Du(x_2)| \leq C \left[ \zeta(|x_1 - x_2|) + \left(1 + \frac{\sup |Du|}{d(x_1)}\right) |x_1 - x_2|^{\theta} \right]
\]
for all $x_1$ and $x_2$ in $\Omega$ with $|x_1 - x_2| \leq \frac{1}{4} d(x_1)$, where $\sigma = \min\{\sigma_0, \alpha/(2n + 2\theta)\}$ and $\theta$ is the constant from Lemma 5.1.

The corresponding boundary regularity result is similar, and the proof is similar.

**Theorem 5.4.** Let $\partial \Omega \in C^{1,\alpha}$ for some $\alpha \in (0, 1)$, let $A$ and $B$ be, respectively, a vector-valued function and a scalar-valued function on $\Omega \times \mathbb{R} \times \mathbb{R}^n$, let $a_0$ be a scalar valued function on $\partial \Omega \times \mathbb{R}$, and let $H$ be a convex, $C^2$ function on $[0, \infty)$ with $H(0) = 0$, and suppose $h = H'$ satisfies (5.1). Suppose also that conditions (5.3a–d) are satisfied and that
\[
(5.10) \quad |a_0(x, z) - a_0(y, w)| \leq \Lambda_2(|x - y| + |z - w|)^{\alpha}
\]
for all \((x, z)\) and \((y, w)\) in \(\partial \Omega \times \mathbb{R}\). Let \(\psi\) satisfy (2.1) and
\[
A(x, \psi, Y) \cdot \gamma + a_0(x, \psi) \geq 0
\]
on \(\partial \Omega\). If \(u\) is a solution of
\[
\int_{\Omega} [A(x, u, Du) \cdot D(u - v) - B(x, u, Du)(u - v)] \, dx \leq \int_{\partial \Omega} a_0(x, u)(u - v) \, d\sigma
\]
with \(K\) the set of all \(v \in W^{1, \infty}\) with \(v \geq \psi\), then there are constants \(\sigma_0(\Lambda, \delta, g_0)\) and \(C(n, \Lambda, \delta, g_0, \sup |Du|, \Psi_1, \Lambda_1, \alpha, \Omega)\) such that
\[
|Du(x_1) - Du(x_2)| \leq C[\zeta(|x_1 - x_2|) + |x_1 - x_2|^\sigma]
\]
for all \(x_1\) and \(x_2\) in \(\Omega\), where \(\sigma = \min\{\sigma_0, \alpha/(2n + 2\theta)\}\).

**Proof.** To prove the analog of (5.7), we let \(U\) solve the variational inequality
\[
\int_{\Omega} A(x_0, u(x_0), DU) \cdot D(U - v) \, dx \leq \int_{\partial \Omega} [a_0(x_0, u(x_0)) + C_0 R] \, d\sigma,
\]
where \(K\) is the set of all \(v \in W^{1, \infty}\) with \(v \geq \psi\) in \(\Omega\) and \(v = u\) on \(\Omega \setminus \Omega[R]\); the constant \(C_0\), which is independent of \(x_0\) and \(R\), is chosen so that
\[
A(x_0, \psi(x_0), Y(x)) \cdot \gamma(x) + a_0(x_0, \psi(x_0)) + C_0 R \geq 0
\]
for all \(x \in \partial \Omega[R]\). The appropriate Hölder gradient estimate was proved in [15, Section 4] for the special case that \(A\) depends only on \(p\), \(a_0\) is constant, and \(B\) is identically zero, and the general estimate follows from the perturbation argument in [17, Section 5].

We leave the straightforward modifications of these results for double obstacle problems to the reader. We do observe that the previous results for double obstacle problems (specifically [7, 18, 26]) all assume that the obstacle has Hölder continuous first derivatives. Thus, we have improved these results by considering general moduli of continuity and also suitable one-sided conditions.


A suitable existence theory for our obstacle problem is based on known \textit{a priori} estimates and the penalization method of Lions (see [14, Section IV.5]). We assume first that \(\partial \Omega \in C^3\) (although this assumption can be relaxed by the remarks at the end of [21, Section 3]), and we assume that \(\psi\) satisfies (2.1) with \(\zeta(t) = z_0 t\). In addition, we assume that (3.9) holds. For \(\rho\) a \(C^2(\Omega)\) function such that \(D\rho = \gamma\) on \(\partial \Omega\) (which always exists), we suppose that there are nonnegative constants \(M_0\) and \(M_1\) such that
\[
\begin{align*}
&zF(x, z, -M_1 D\rho, -M_1 D^2 \rho) < 0 \text{ in } \Omega, \\
&zG(x, z, -M_1 \gamma) < 0 \text{ on } \partial \Omega
\end{align*}
\]

for \( z \geq M_0 \). Next, we assume that there are increasing functions \( \mu \) and \( \mu_0 \) such that

(6.2a) \[
\lambda(x, z, p, r) |\xi|^2 \leq F^{ij}(x, z, p, r) \xi_i \xi_j \leq \Lambda(x, z, p, r) |\xi|^2,
\]

(6.2b) \[
\Lambda(x, z, p, r) \leq \mu_0(|z|) \lambda(x, z, p, r)
\]

(6.2c) \[
|F(x, z, p, 0)| \leq \mu_0(|z|) \lambda(x, z, p, r)[1 + |p|^2]
\]

for all \((x, z, p, r) \in \Gamma\) and

(6.2d) \[
|G(x, z, p')| \leq \mu_0(|z|) G_p(x, z, p) \cdot \gamma[1 + |p'|]
\]

for all \((x, z, p) \in \Gamma'\), where \( p' = p - (\gamma \cdot p) \gamma \). We also assume that there is an increasing function \( \mu_1 \) such that

(6.3) \[
(1 + |p|)|F_p| + |F_z| + |F_x| \leq \mu_1(|z|) \lambda[1 + |p|^2 + |r|]
\]

on \( \Gamma \) and

(6.4) \[
(1 + |p|)|G_p| + |G_z| + |F_x| \leq \mu_1(|z|) G_p \cdot \gamma[1 + |p|]
\]

on \( \Gamma' \). Finally we assume that \( F \) is concave (or concave) with respect to \( r \) and that \( \lambda \) is uniformly bounded above and uniformly positive on bounded sets of \( \Gamma \), and we assume that \( G_p(x, z, p) \cdot \gamma \) is uniformly bounded and uniformly positive on bounded subsets of \( \Gamma' \). Note that \([21, \text{Lemma 7.1}]\) implies the upper bound \( u \leq M_1 \sup \rho \) while the obstacle condition imply that \( u \geq \min \psi \). Hence, we may assume that conditions (6.2)–(6.4) hold with \( \mu, \mu_0, \) and \( \mu_1 \) independent of \( z \) by redefining \( F \) and \( G \) for large \(|z|\) as needed. In particular, we may assume that \( F \) and \( G \) are independent of \( z \) for \( z \leq \psi(x) \).

Now for \( \varepsilon \in (0, 1) \), we define \( \beta_\varepsilon \) by

\[
\beta_\varepsilon(t) = (\min\{t, 0\})^2 / \varepsilon.
\]

It then follows from \([21, \text{Lemma 7.1, Theorems 3.3, 4.1, and 7.8}]\) along with \([29, \text{Theorem 3.3}]\) (see also \([19, \text{Theorems 14.22 and 14.23}]\)) that the problem

\[
F(x, u_\varepsilon, Du_\varepsilon, D^2 u_\varepsilon) + \beta_\varepsilon(u_\varepsilon - \psi) = 0 \quad \text{in } \Omega,
\]

\[
G(x, u_\varepsilon, Du_\varepsilon) + \varepsilon = 0 \quad \text{on } \partial \Omega
\]

has a \(C^{2, \theta}(\Omega)\) solution for any \( \varepsilon \in (0, 1) \) and some \( \theta \in (0, 1) \) upon recalling our previous observations that we may take \( \mu, \mu_0, \) and \( \mu_1 \) independent of \( z \). As previously remarked, \([21, \text{Lemma 7.1}]\) implies that \( (u_\varepsilon) \) is uniformly bounded, independent of \( \varepsilon \).

Now we estimate \( \beta_\varepsilon(u_\varepsilon - \psi) \). If the minimum of \( u_\varepsilon - \psi \) is nonnegative, then \( \beta_\varepsilon = 0 \). In addition, at a boundary minimum,

\[
-\varepsilon = G(x, u_\varepsilon, Du_\varepsilon) \geq G(x, u_\varepsilon, Y),
\]
so if the minimum of \( u_\varepsilon - \psi \) is negative, it must occur at some \( x_0 \in \Omega \). In this case, \( Du_\varepsilon(x_0) = Y(x_0) \) and \( D^2u_\varepsilon \geq -2z_0 I \), where \( I \) denotes the \( n \times n \) identity matrix, so
\[
\beta_\varepsilon(u_\varepsilon - \psi)(x_0) = -F(x_0, u_\varepsilon, Du_\varepsilon, D^2u_\varepsilon) \leq F(x_0, \psi(x_0), Y(x_0), -2z_0 I).
\]
It follows that \( \beta_\varepsilon(u_\varepsilon - \psi_\varepsilon) \leq c_1 \) for some nonnegative constant \( c_1 \) independent of \( \varepsilon \). We can then use [21, Theorem 3.3] to infer a global gradient bound for \( u_\varepsilon \), which is uniform with respect to \( \varepsilon \). We can then apply [3, Theorem 2] (see [19, Lemma 13.21, and Theorems 14.14 and 14.20] for a discussion of the extension to the oblique derivative boundary condition) to infer that \( |Du_\varepsilon|_\alpha \leq c_2 \) for constants \( \alpha \in (0,1) \) and \( c_2 \) independent of \( \varepsilon \). Finally, [3, Theorem 1] shows that \( (D^2u_\varepsilon) \) is bounded in \( L^p_{\text{loc}}(\Omega) \) for any \( p < \infty \).

From these estimates and the argument on pages 44 and 45 of [1], we infer that there is a sequence \( (\varepsilon(j)) \) such that \( (u_\varepsilon(j)) \) converges to a function \( u \in W^{2,n}_{\text{loc}}(\Omega) \cap C^{1,\alpha} \) and that \( u \) solves (0.1). Theorem 2.3 then implies that \( u \in C^{1,1}_{\text{loc}}(\Omega) \).

Note that a more thorough existence theory can be derived via approximation of the obstacle; however, the convergence of the approximating solutions to a function in \( W^{2,n}_{\text{loc}}(\Omega) \) requires at least that the obstacle be a supremum of \( W^{2,n}_{\text{loc}}(\Omega) \) functions. On the other hand, the extension to two-obstacle problems, which we leave to the reader, is very simple.

References


Received January 24, 2000.

Department of Mathematics
Iowa State University
Ames, Iowa 50011
E-mail address: lieb@iastate.edu
ENTROPY IN TYPE I ALGEBRAS

SERGEY NESHVEYEV AND ERLING STÖRMER

It is shown that if \((M, \phi, \alpha)\) is a \(W^*\)-dynamical system with \(M\) a type I von Neumann algebra then the entropy of \(\alpha\) w.r.t. \(\phi\) equals the entropy of the restriction of \(\alpha\) to the center of \(M\). If furthermore \((N, \psi, \beta)\) is a \(W^*\)-dynamical system with \(N\) injective then \(h_{\phi \otimes \psi}(\alpha \otimes \beta) = h_\phi(\alpha) + h_\psi(\beta)\).

1. Introduction.

In the theory of non-commutative entropy the attention has almost exclusively been concentrated on non-type I algebras. We shall in the present paper remedy this situation by proving the basic facts on entropy of automorphisms of type I C\(^*\)- and von Neumann-algebras. The results are as nice as one can hope. The CNT-entropy of an automorphism of a von Neumann algebra of type I with respect to an invariant normal state is the classical entropy of the restriction of the automorphism to the center of the algebra. If one factor of a tensor product of two von Neumann algebras is of type I and the other injective, then the entropy of a tensor product automorphism with respect to an invariant product state is the sum of the entropies. The results have obvious corollaries to type I C\(^*\)-algebras. The main idea behind our proofs is the use of conditional expectations of finite index, as employed in [GN].

We shall use the notation \(h_\phi(\alpha)\) for the CNT-entropy of a C\(^*\)-dynamical system as defined by Connes, Narnhofer and Thirring in [CNT], and \(h'_\phi(\alpha)\) for the ST-entropy defined by Sauvageot and Thouvenot in [ST].

2. Main results.

We first prove a general result for the Sauvageot-Thouvenot entropy for the restriction of an automorphism to a globally invariant C\(^*\)-subalgebra of finite index.

**Proposition 1.** Let \((A, \phi, \alpha)\) be a unital C\(^*\)-dynamical system. Let \(B \subset A\) be an \(\alpha\)-invariant C\(^*\)-subalgebra (with \(1 \in B\)). Suppose there exists a conditional expectation \(E : A \to B\) such that \(E \circ \alpha = \alpha \circ E\), \(\phi \circ E = \phi\) and \(E(x) \geq cx\) for all \(x \in A^+\) for some \(c > 0\). Then \(h'_\phi(\alpha) = h'_\phi(\alpha|_B)\).
Proof. Let \((C, \mu, \beta)\) be a \(C^*\)-dynamical system with \(C\) abelian. Using \(E\) we can lift any stationary coupling on \(B \otimes C\) to a stationary coupling on \(A \otimes C\). This, together with the property of monotonicity of relative entropy, shows that \(h'_\phi(\alpha) \geq h'_\phi(\alpha|_B)\).

Conversely, suppose \(\lambda\) is a stationary coupling of \((A, \phi, \alpha)\) with \((C, \mu, \beta)\), and \(P\) is a finite-dimensional subalgebra of \(C\) with atoms \(p_1, \ldots, p_n\). Let

\[
\phi_i(a) = \frac{\lambda(a \otimes p_i)}{\mu(p_i)} \quad \text{for } a \in A.
\]

Then in the notations of \([ST]\)

\[
h'_\phi(\alpha) = \sup \left\{ H_{\mu}(P|P^c) - H_{\mu}(P) + \sum_{i=1}^n \mu(p_i) S(\phi, \phi_i) \right\}.
\]

Since \(\phi_i \leq \frac{1}{\mu(p_i)} \phi\), \(\phi_i\) is normal in the GNS-representation of \(\phi\). Since \(E\) is \(\phi\)-invariant, it extends to a normal conditional expectation of the closure of \(A\) in the GNS-representation onto the closure of \(B\). Thus we can apply \([OP\), Theorem 5.15\] to \(\phi\) and \(\phi_i\), and (as in the proof of Lemma 1.5 in \([GN]\)) get

\[
\sum_{i=1}^n \mu(p_i) S(\phi, \phi_i) = \sum_{i=1}^n \mu(p_i) (S(\phi|_B, \phi_i) + S(\phi_i \circ E, \phi_i)) \leq \sum_{i=1}^n \mu(p_i) S(\phi|_B, \phi_i) - \log c.
\]

It follows that \(h'_\phi(\alpha) \leq h'_\phi(\alpha|_B) - \log c\). Then for each \(m \in \mathbb{N}\)

\[
h'_\phi(\alpha) = h'_\phi(\alpha|_B) - \frac{1}{m} \log c = \frac{1}{m} h'_\phi(\alpha\alpha^m|_B) - \frac{1}{m} \log c.
\]

Thus \(h'_\phi(\alpha) \leq h'_\phi(\alpha|_B)\). \(\square\)

By \([ST\), Proposition 4.1\] the Sauvageot-Thouvenot entropy coincides with the CNT-entropy for nuclear \(C^*\)-algebras. In fact, what is really necessary for the coincidence of the entropies, is the existence of a net of unital completely positive mappings \(\gamma_i\) of finite-dimensional \(C^*\)-algebras into \(A\) such that \(S(\phi, \psi) = \lim_i S(\phi \circ \gamma_i, \psi \circ \gamma_i)\) for any positive linear functional \(\psi\) on \(A\), \(\psi \leq \phi\). We therefore have:

**Corollary 2.** If in the above proposition \(A\) and \(B\) are injective von Neumann algebras and \(\phi\) is normal then \(h_\phi(\alpha) = h_\phi(\alpha|_B)\).

To prove our main result we need also two simple lemmas. The first lemma is more or less well-known.

**Lemma 3.** Let \((M, \phi, \alpha)\) be a \(W^*\)-dynamical system. Then

(i) if \(p\) is an \(\alpha\)-invariant projection in \(M\) such that \(\text{supp} \phi \leq p\), then \(h_\phi(\alpha) = h_\phi(\alpha|_{M_p})\);
(ii) if \( \{p_i\}_{i \in I} \) is a set of mutually orthogonal \( \alpha \)-invariant central projections in \( M \), \( \sum_i p_i = 1 \), then
\[
h_\phi(\alpha) = \sum_i \phi(p_i)h_\phi(\alpha_i),
\]
where \( \phi_i = \frac{1}{\phi(p_i)}\phi \) is the normalized restriction of \( \phi \) to \( Mp_i \), and \( \alpha_i = \alpha|_{Mp_i} \).

Proof. (i) easily follows from the definitions; (ii) follows from [CNT, VII.5(iii)], (i) and [SV, Lemma 3.3] applied to the subalgebras \( M(p_i + \cdots + p_n) + \mathbb{C}(1 - p_i - \cdots - p_n) \).

The proof of the following lemma is left to the reader.

Lemma 4. Let \( T \) be an automorphism of a probability space \( (X, \mu) \), \( f \in L^\infty(X, \mu) \) a \( T \)-invariant function such that \( f \geq 0 \) and \( \int_X f \, d\mu = 1 \). Let \( \mu_f \) be the measure on \( X \) such that \( d\mu_f/d\mu = f \). Then \( h_{\mu_f}(T) \leq \|f\|_\infty h_\mu(T) \).

Theorem 5. Let \( (M, \phi, \alpha) \) be a \( W^* \)-dynamical system with \( M \) a von Neumann algebra of type I. Let \( Z \) denote the center of \( M \). Then \( h_\phi(\alpha) = h_\phi(\alpha|_Z) \).

Proof. By Lemma 3(i) we may suppose that \( \phi \) is faithful. Then \( M \) is a direct sum of homogeneous algebras of type I, \( n \in \mathbb{N} \cup \{\infty\} \). By Lemma 3(ii) we may assume that \( M \) is homogeneous of type I, \( n \). We first assume that \( n \in \mathbb{N} \). Then \( Z = L^\infty(X, \mu) \), where \( (X, \mu) \) is a probability space and \( \phi|_Z = \mu \). Thus
\[
M \cong Z \otimes \text{Mat}_n(\mathbb{C}) = L^\infty(X, \text{Mat}_n(\mathbb{C})), \quad \phi = \int_X \phi_x d\mu(x),
\]
where \( \phi_x = \text{Tr}(\cdot Q_x) \) is a state on \( \text{Mat}_n(\mathbb{C}) \), \( \text{Tr} \) the canonical trace on \( \text{Mat}_n(\mathbb{C}) \). We first assume \( Q_x \geq c > 0 \) for all \( x \).

If \( s \in M^+ \), \( s \) is a function in \( L^\infty(X, \text{Mat}_n(\mathbb{C})) \). Define the \( \phi \)-preserving conditional expectation \( E: M \to Z \) by \( E(s)(x) = \phi_x(s(x)) \). Then
\[
E(s)(x) = \text{Tr}(s(x)Q_x) \geq c\text{Tr}(s(x)) \geq cs(x),
\]
so \( E(s) \geq cs \), and it follows from Corollary 2 that \( h_\phi(\alpha) = h_\phi(\alpha|_Z) \).

If there is no \( c > 0 \) such that \( Q_x \geq c \) for all \( x \), let \( X_c = \{x \in X \mid Q_x \geq c\} \), \( (c > 0) \),
\[
N_c = L^\infty(X_c, \text{Mat}_n(\mathbb{C})) \quad \text{and} \quad M_c = N_c + \mathbb{C}\chi_{X\setminus X_c},
\]
where \( \chi_{X\setminus X_c} \) is the characteristic function of \( X\setminus X_c \). Since \( \phi \) is \( \alpha \)-invariant so is \( M_c \), so by the above argument and Lemma 3, letting \( \phi_c = \frac{1}{\mu(X_c)}\phi|_{N_c} \) and \( \mu_c = \frac{1}{\mu(X_c)}\mu|_{X_c} \), we obtain
\[
h_\phi(\alpha|_{M_c}) = \mu(X_c)h_{\phi_c}(\alpha|_{N_c}) = \mu(X_c)h_{\mu_c}(T|_{X_c}) \leq h_{\mu}(T),
\]
where \( T \) is the automorphism of \((X, \mu)\) induced by \( \alpha \). Letting \( c \to 0 \) and using [SV, Lemma 3.3] we obtain the Theorem when \( M \) is finite.

If \( M \) is homogeneous of type \( \text{I}_\infty \), we have \( M \cong L^\infty(X, \mu) \otimes B(H) \), where \( H \) is a separable Hilbert space. Let \( \text{Tr} \) denote the canonical trace on \( B(H) \).

Write again
\[
\phi = \int_X \phi_x \, d\mu(x), \quad \phi_x = \text{Tr}(Q_x),
\]
and let \( E_x(U) \) denote the spectral projection of \( Q_x \) corresponding to a Borel set \( U \). Let \( P_c \in M = L^\infty(X, B(H)) \) be the projection defined by \( P_c(x) = E_x([c, +\infty)) \), where \( c > 0 \). Then \( P_c \) is an \( \alpha \)-invariant finite projection. Let
\[
M_c = P_c M P_c + \mathbb{C}(1 - P_c).
\]
Then \( M_c \) is a finite type \( \text{I} \) von Neumann algebra. Its center is isomorphic to \( L^\infty(\hat{X}_c, \mu_c) \oplus \mathbb{C} \), and the restriction of \( \phi \) to it is \( \phi(P_c) \mu_c \oplus \phi(1 - P_c) \), where \( \hat{X}_c = \{ x \in X \mid P_c(x) \neq 0 \} \) and
\[
\int_{\hat{X}_c} f(x) \, d\mu_c(x) = \frac{1}{\phi(P_c)} \int_{\hat{X}_c} f(x) \phi_x(P_c(x)) \, d\mu(x).
\]
So we can apply the first part of the proof to \( M_c \). Since \( d\mu_c/d\mu \leq \frac{1}{\phi(P_c)} \), applying Lemma 4 we get
\[
h_{\phi}(\alpha|_{M_c}) = \phi(P_c) h_{\mu_c}(T|_{\hat{X}_c}) \leq h_{\mu}(T).
\]
Now letting \( c \to 0 \) we conclude that \( h_{\phi}(\alpha) = h_{\mu}(T) \). \( \square \)

It should be remarked that in a special case the above theorem was proved in [GS, Proposition 2.4].

If \( A \) is a \( \text{C}^* \)-algebra and \( \phi \) a state on \( A \), the central measure \( \mu_\phi \) of \( \phi \) is the measure on the spectrum \( \hat{A} \) of \( A \) defined by \( \mu_\phi(F) = \phi(\chi_F) \), where \( \phi \) is regarded as a normal state on \( A'' \), see [P, 4.7.5]. Thus by Theorem 5 and [P, 4.7.6] we have the following:

**Corollary 6.** Let \((A, \phi, \alpha)\) be a \( \text{C}^* \)-dynamical system with \( A \) a separable unital type \( \text{I} \) \( \text{C}^* \)-algebra. Then \( h_{\phi}(\alpha) = h_{\mu_\phi}(\hat{\alpha}) \), where \( \hat{\alpha} \) is the automorphism of the measure space \((\hat{A}, \mu_\phi)\) induced by \( \alpha \).

Since inner automorphisms act trivially on the center we have:

**Corollary 7.** If \((M, \phi, \alpha)\) is a \( \text{W}^* \)-dynamical system with \( M \) of type \( \text{I} \) and \( \alpha \) an inner automorphism then \( h_{\phi}(\alpha) = 0 \).

Note that in the finite case the above corollary also follows from a result of N. Brown [Br, Lemma 2.2].

The next result was shown in [S] when \( \phi \) is a trace.
Corollary 8. Let \( R \) denote the hyperfinite II\(_1\)-factor. Let \( A \) be a Cartan subalgebra of \( R \) and \( u \) a unitary operator in \( A \). If \( \phi \) is a normal state such that \( u \) belongs to the centralizer of \( \phi \) then \( h_\phi(\text{Ad} u) = 0 \).

Proof. As in [S], it follows from [CFW] that there exists an increasing sequence of full matrix algebras \( N_1 \subset N_2 \subset \ldots \) with union weakly dense in \( R \) such that \( A \cong A_n \otimes B_n \), where \( A_n = N_n \cap A \) and \( B_n = (N'_n \cap R) \cap A \) for all \( n \in \mathbb{N} \). Let \( M_n = N_n \otimes B_n \). Then \( M_n \) is of type I and contains \( u \). Hence \( h_\phi(\text{Ad} u|_{M_n}) = 0 \). Since \((\cup_n M_n)^- = R \), \( h_\phi(\text{Ad} u) = 0 \) by [SV, Lemma 3.3].

If \((A, \phi, \alpha)\) and \((B, \psi, \beta)\) are C*-dynamical systems we always have

\[
h_{\phi \otimes \psi}(\alpha \otimes \beta) \geq h_\phi(\alpha) + h_\psi(\beta),
\]

see [SV, Lemma 3.4]. Equality does not always hold, see [NST] or [Sa]. However, we have:

Theorem 9. Let \((A, \phi, \alpha)\) and \((B, \psi, \beta)\) be \( W^* \)-dynamical systems. Suppose that \( A \) is of type I, and \( B \) is injective. Then

\[
h_{\phi \otimes \psi}(\alpha \otimes \beta) = h_\phi(\alpha) + h_\psi(\beta).
\]

Proof. We shall rather prove that \( h_{\phi \otimes \psi}(\alpha \otimes \beta) = h_\phi(\alpha|_{Z(A)}) + h_\psi(\beta) \). For this it suffices to consider the case when \( A \) is abelian; the general case will follow by the same arguments as in the proof of Theorem 5. (Note that the mapping \( x \mapsto \text{Tr}(x) - x \) on \( \text{Mat}_n(\mathbb{C}) \) is not completely positive, but the mapping \( x \mapsto \text{Tr}(x) - \frac{1}{n} x \) is by the Pimsner-Popa inequality. Thus replacing \( M \) with \( M \otimes B \) and \( Z \) with \( Z \otimes B \) in the proof of Theorem 5 we have to replace the inequality \( E(s) \geq cs \) in the proof with \( E(s) \geq \frac{\pi}{2h} s \).

So suppose that \( A \) is abelian. It is clear that it suffices to prove that if \( A_1, \ldots, A_n \) are finite-dimensional subalgebras of \( A \), and \( B_1, \ldots, B_n \) are finite-dimensional subalgebras of \( B \), then

\[
H_{\phi \otimes \psi}(A_1 \otimes B_1, \ldots, A_n \otimes B_n) = H_\phi(A_1, \ldots, A_n) + H_\psi(B_1, \ldots, B_n).
\]

We always have the inequality "\( \geq \)", [SV, Lemma 3.4]. To prove the opposite inequality consider a decomposition

\[
\phi \otimes \psi = \sum_{i_1, \ldots, i_n} \omega_{i_1 \ldots i_n}.
\]

Let \( H_{\{\phi \otimes \psi = \sum \omega_{i_1 \ldots i_n}\}}(A_1 \otimes B_1, \ldots, A_n \otimes B_n) \) be the entropy of the corresponding abelian model, so

\[
H_{\{\phi \otimes \psi = \sum \omega_{i_1 \ldots i_n}\}}(A_1 \otimes B_1, \ldots, A_n \otimes B_n)
\]

\[
= \sum_{i_1, \ldots, i_n} \eta \omega_{i_1 \ldots i_n}(1) + \sum_{k=1}^n \sum_i S \left( \phi \otimes \psi|_{A_k \otimes B_k} \right) \sum_{i_k = i} \omega_{i_1 \ldots i_n}|_{A_k \otimes B_k}.
\]
Set $C = \bigvee_{k=1}^{n} A_k$. Let $p_1, \ldots, p_r$ be those atoms $p$ of $C$ for which $\phi(p) > 0$. Define positive linear functionals $\psi_{m,i_1 \ldots i_n}$ on $B$,

$$
\psi_{m,i_1 \ldots i_n}(b) = \frac{\omega_{i_1 \ldots i_n}(p_m \otimes b)}{\phi(p_m)}.
$$

Let also $\phi_m$ be the linear functional on $C$ defined by the equality $\phi_m(a) = \phi(ap_m)$. Then

$$
\omega_{i_1 \ldots i_n} = \sum_{m=1}^{r} \phi_m \otimes \psi_{m,i_1 \ldots i_n} \text{ on } C \otimes B,
$$

and

$$
\psi = \sum_{i_1 \ldots i_n} \psi_{m,i_1 \ldots i_n} \text{ for } m = 1, \ldots, r.
$$

Since the supports of the positive functionals $\phi_m$ are mutually orthogonal minimal projections in $C$, we have

$$
\sum_{k=1}^{n} \sum_{i} S \left( \phi \otimes \psi|_{A_k \otimes B_k}, \sum_{i_k=i} \omega_{i_1 \ldots i_n}|_{A_k \otimes B_k} \right) 
\leq \sum_{k=1}^{n} \sum_{i} S \left( \phi \otimes \psi|_{C \otimes B_k}, \sum_{i_k=i} \omega_{i_1 \ldots i_n}|_{C \otimes B_k} \right) 
= \sum_{k=1}^{n} \sum_{i} S \left( \phi \otimes \psi|_{C \otimes B_k}, \sum_{m=1}^{r} \phi_m \otimes \left( \sum_{i_k=i} \psi_{m,i_1 \ldots i_n} \right)|_{C \otimes B_k} \right) 
= \sum_{k=1}^{n} \sum_{i} \sum_{m=1}^{r} \phi(p_m) S \left( \psi|_{B_k}, \sum_{i_k=i} \psi_{m,i_1 \ldots i_n}|_{B_k} \right).
$$

If $a_i \geq 0$ then $\eta \left( \sum_{i} a_i \right) \leq \sum_{i} \eta(a_i)$. Hence we have

$$
\sum_{i_1 \ldots i_n} \eta\omega_{i_1 \ldots i_n}(1) 
\leq \sum_{m=1}^{r} \sum_{i_1 \ldots i_n} \eta(\phi_m \otimes \psi_{m,i_1 \ldots i_n})(1) 
= \sum_{m=1}^{r} \eta\phi(p_m) \sum_{i_1 \ldots i_n} \psi_{m,i_1 \ldots i_n}(1) + \sum_{m=1}^{r} \phi(p_m) \sum_{i_1 \ldots i_n} \eta\psi_{m,i_1 \ldots i_n}(1) 
= \sum_{m=1}^{r} \eta\phi(p_m) + \sum_{m=1}^{r} \phi(p_m) \sum_{i_1 \ldots i_n} \eta\psi_{m,i_1 \ldots i_n}(1).
$$
Thus
\[
H_{\{\phi \otimes \psi = \sum \omega_1 \ldots \omega_n\}}(A_1 \otimes B_1, \ldots , A_n \otimes B_n)
\leq \sum_{m=1}^{r} \eta \phi(p_m) \sum_{m=1}^{r} \phi(p_m)H_{\{\psi = \sum \psi_{\omega_1 \ldots \omega_n}\}}(B_1, \ldots , B_n).
\]
Since \(\sum_m \eta \phi(p_m) = H_\phi(C) = H_\phi(A_1, \ldots , A_n)\), we conclude that
\[
H_{\phi \otimes \psi}(A_1 \otimes B_1, \ldots , A_n \otimes B_n) \leq H_\phi(A_1, \ldots , A_n) + H_\psi(B_1, \ldots , B_n),
\]
completing the proof of the Theorem. \(\square\)

References


Received January 31, 2000. The first author was partially supported by NATO grant SA (PST.CLG.976206)5273. The second author was partially supported by the Norwegian Research Council.

B. Verkin Institute for Low Temperature Physics and Engineering
National Academy of Sciences of Ukraine,
47, Lenin Ave.
310164, Kharkov
Ukraine

Department of Mathematics
University of Oslo
P.O. Box 1053
Blindern, 0316 Oslo
Norway

E-mail address: erlings@math.uio.no
HOLOMORPHY OF IGUSA’S AND TOPOLOGICAL ZETA FUNCTIONS FOR HOMOGENEOUS POLYNOMIALS

B. Rodrigues and W. Veys

Let $F$ be a number field and $f \in F[x_1, \ldots, x_n] \setminus F$. To any completion $K$ of $F$ and any character $\kappa$ of the group of units of the valuation ring of $K$ one associates Igusa’s local zeta function $Z^K(\kappa, f, s)$. The holomorphy conjecture states that for all except a finite number of completions $K$ of $F$ we have that if the order of $\kappa$ does not divide the order of any eigenvalue of the local monodromy of $f$ at any complex point of $f^{-1}\{0\}$, then $Z^K(\kappa, f, s)$ is holomorphic on $\mathbb{C}$. The second author already showed that this conjecture is true for curves, i.e., for $n = 2$. Here we look at the case of an homogeneous polynomial $f$, so we can consider $\{f = 0\} \subseteq \mathbb{P}^{n-1}$. Under the condition that $\chi(\mathbb{P}^{n-1}_C \setminus \{f = 0\}) \neq 0$ we prove the holomorphy conjecture. Together with some results in the case when $\chi(\mathbb{P}^{n-1}_C \setminus \{f = 0\}) = 0$, we can conclude that the holomorphy conjecture is true for an arbitrary homogeneous polynomial in three variables.

We also prove the so-called monodromy conjecture for a homogeneous polynomial $f \in F[x_1, x_2, x_3]$ with $\chi(\mathbb{P}^2_C \setminus \{f = 0\}) \neq 0$.

0. Introduction.

0.1. Let $K$ be a finite extension of the field $\mathbb{Q}_p$ of $p$-adic numbers, $R_K$ the valuation ring of $K$, $P_K$ the maximal ideal of $R_K$, $\pi$ a fixed uniformizing parameter for $R_K$, and $\overline{K} = R_K/P_K$ the residue field of $K$ with cardinality $q$. For $z \in K$, $\mathrm{ord}_\pi z \in \mathbb{Z} \cup \{+\infty\}$ denotes the valuation of $z$, $|z| = q^{-\mathrm{ord}_\pi z}$ and $\mathrm{ac}(z) = z\pi^{-\mathrm{ord}_\pi z}$ is the angular component of $z$.

Let $f(x) \in K[x], x = (x_1, \ldots, x_n)$, be a nonconstant polynomial and $\kappa$ a character of $R_K^\times$, i.e., a homomorphism $\kappa: R_K^\times \rightarrow \mathbb{C}^\times$ with finite image, where $R_K^\times$ denotes the group of units of $R_K$. (We formally put $\kappa(0) = 0$.) To these data one associates Igusa’s local zeta function $Z^K(\kappa, f, s)$, which is the meromorphic continuation to $\mathbb{C}$ of

$$s \mapsto \int_{R_K^\times} \kappa(ac f(x)) |f(x)|^s |dx|,$$
implies that when the order of \( \kappa \) does not divide the order of any eigenvalue of the local monodromy on \( f^{-1}(0) \), then \( Z^K(\kappa, f, s) \) is holomorphic on \( \mathbb{C} \).

0.2. We can write \( Z^K(\kappa, f, s) \) and \( Z_0^K(\kappa, f, s) \) in terms of an embedded resolution \( h^K : Y^K \to \mathbb{A}^n(K) \) of \( f^{-1}(0) \) in \( \mathbb{A}^n(K) \), see Theorem 1.5.1. Let \( E_j^K, j \in T^K, \) be the (reduced) irreducible components of \( (h^K)^{-1}(f^{-1}(0)) \), and let \( N_j \) be the multiplicity of \( E_j^K \) in the divisor of \( f \circ h^K \) on \( Y^K \). Then Theorem 1.5.1 implies that when the order of \( \kappa \) divides \( N_j \) at all, the zeta functions \( Z^K(\kappa, f, s) \) and \( Z_0^K(\kappa, f, s) \) will be holomorphic on \( \mathbb{C} \). Now the \( N_j \) are not intrinsically associated to \( f^{-1}(0) \); but the order (as root of unity) of any eigenvalue of the local monodromy on \( f^{-1}(0) \) divides some \( N_j \), and those eigenvalues are intrinsic invariants of \( f^{-1}(0) \) (see (1.6)). This observation inspired Denef [2, Conjecture 4.4.2] to propose the following.

0.3. Holomorphy Conjecture. Let \( f \in F[x_1, \ldots, x_n] \setminus F \) for some number field \( F \). Then for almost all completions \( K \) of \( F \) (i.e., for all except a finite number) we have the following for any character \( \kappa \) of \( R_K \). If the order of \( \kappa \) does not divide the order of any eigenvalue of the (complex) local monodromy of \( f \) at any complex point of \( f^{-1}(0) \), then \( Z^K(\kappa, f, s) \) is holomorphic on \( \mathbb{C} \).

0.4. Remark. Denef also formulated this conjecture for \( Z_0^K(\kappa, f, s) \) and for a generalization of those two involving a Schwartz-Bruhat function, i.e., a locally constant function with compact support.

0.5. The second author showed in [10] that this conjecture is true for curves, i.e., for \( f \in F[x_1, x_2] \). In this paper we consider the case of a homogeneous polynomial \( f \), but in an arbitrary number of variables; thus \( f \in F[x_1, \ldots, x_n] \). Remark that for such \( f \) we can consider \( \{f = 0\} \subseteq \mathbb{P}^{n-1} \). Under the condition that \( \chi(\mathbb{P}^{n-1} \setminus \{f = 0\}) \neq 0 \) we will prove the holomorphy conjecture for \( f \). (Here \( \chi(\cdot) \) denotes the topological Euler-Poincaré characteristic.) If this condition is not fulfilled we will formulate a sort of ‘projective holomorphy conjecture’ implying Conjecture 0.3 as we will prove in Theorem 3.5. Important is that with this projective version of the holomorphy conjecture we actually drop the dimension by one. This will enable us to prove the holomorphy conjecture for an arbitrary homogeneous polynomial in three variables by reducing the problem to the situation of curves. We will also prove all these results for the so-called topological zeta function (see (2.3) for the definition of this function).

In the last section we will prove the monodromy conjecture for a homogeneous polynomial \( f \) in three variables (under the condition that \( \chi(\mathbb{P}^2 \setminus \{f = 0\}) \neq 0 \)). This conjecture roughly states that if \( s_0 \) is a pole of \( Z^K(\kappa, f, s) \), then \( \exp(2\pi i \text{Re}(s_0)) \) is an eigenvalue of the local monodromy of \( f \) at some complex point of \( f^{-1}(0) \).
1. Explicit formulas.

1.1. In this section we will construct some embedded resolutions playing the key role in the proof of our results. We will also state some general formulas for Igusa’s local zeta function and for eigenvalues of the local monodromy of \( f \) in terms of those embedded resolutions.

1.2. Let \( f \in F[x_1, \ldots, x_n]\backslash F \) be a homogeneous polynomial over some field \( F \) of characteristic zero, \( \mathcal{P} = \text{Proj} \ F[x_1, \ldots, x_n] \) and \( \mathcal{D} = \text{Proj} \ (F[x_1, \ldots, x_n]/(f)) \). An embedded resolution of \( \mathcal{D} \) in \( \mathcal{P} \) consists of a nonsingular variety \( \mathcal{Y} \) and a proper birational morphism \( \varphi : \mathcal{Y} \to \mathcal{P} \) such that the restriction \( \varphi : \mathcal{Y}\backslash \varphi^{-1}(\mathcal{D}) \to \mathcal{P}\backslash \mathcal{D} \) is an isomorphism and \( \varphi^{-1}(\mathcal{D}) \) has normal crossings in \( \mathcal{Y} \). By Hironaka [6] we can choose such an embedded resolution \((\mathcal{Y}, \varphi)\) of \( \mathcal{D} \) in \( \mathcal{P} \) over \( F \) by means of blowing-ups. Let \( \mathcal{E}_i, i \in \mathcal{T} = \mathcal{T}_e \cup \mathcal{T}_s \), denote the (reduced) \( F \)-irreducible components of \( \varphi^{-1}(\mathcal{D}) \), where \( i \in \mathcal{T}_s \) if and only if \( \mathcal{E}_i \) is a (reduced) irreducible component of the strict transform of \( \mathcal{D} \).

Then we can write \( f = \prod_{i \in \mathcal{T}_s} f_i^{N_i} \), where each \( f_i \) is an irreducible homogeneous polynomial over \( F \) and \( f_i \) corresponds with \( \mathcal{E}_i \) \( (i \in \mathcal{T}_s) \) in the obvious way. Let \( P \) be the divisor of \( f \) on \( \mathcal{P} \), i.e., \( P = \text{div} \ f = \sum_{i \in \mathcal{T}_s} N_i P_i \), with \( P_i = \text{Proj} \ (F[x_1, \ldots, x_n]/(f_i)) \). Then for \( i \in \mathcal{T} \) we define \( N_i \) to be the multiplicity of \( \mathcal{E}_i \) in the divisor \( \varphi^*(P) \) on \( \mathcal{Y} \).

For a fixed point \( b \) of any \( \mathcal{E}_i \), we can choose local coordinates \((u_1, \ldots, u_{n-1})\) around \( b \) and local coordinates \((v_1, \ldots, v_{n-1})\) around \( \varphi(b) \). Then \( \nu_i - 1 \) is defined to be the multiplicity of \( \mathcal{E}_i \) in the local divisor defined by \( \det \left( \frac{\partial(v_1, \ldots, v_{n-1})}{\partial(u_1, \ldots, u_{n-1})} \right) \). Remark that this is independent of the choice of \( b \) on \( \mathcal{E}_i \) and the choice of local coordinates around \( b \) and \( \varphi(b) \).

The ordered pairs of positive integers \((N_i, \nu_i), i \in \mathcal{T}_s\) are called the numerical data of the resolution \((\mathcal{Y}, \varphi)\). For \( i \in \mathcal{T} \) and \( I \subseteq \mathcal{T} \) we denote \( \mathcal{E}^0 := \mathcal{E}_i \setminus \bigcup_{j \notin i} \mathcal{E}_j \), \( \mathcal{E}^I := \bigcap_{i \in I} \mathcal{E}_i \) and \( \mathcal{E}^I := \mathcal{E}_i \setminus \bigcup_{j \notin I} \mathcal{E}_j \). In particular when \( I = \emptyset \), we have that \( \mathcal{E}_\emptyset = \mathcal{Y} \). Remark that \( \mathcal{Y} \) is the disjoint union of the \( \mathcal{E}^I \).

1.3. For any field extension \( L \) of \( F \) we can take the base extension of the resolution \((\mathcal{Y}, \varphi)\). The result will be an embedded resolution \((\mathcal{Y}^L, \varphi^L)\) of \( \text{Proj} \ (L[x_1, \ldots, x_n]/(f)) \) in \( \text{Proj} \ L[x_1, \ldots, x_n] \) over \( L \), with \( \mathcal{Y}^L = \mathcal{Y}^F \times_F L \).

If there is any danger of confusion we will include the field \( L \) in the notation of (1.2) and thus write \( \mathcal{P}^L, \mathcal{D}^L, \mathcal{E}^L, \mathcal{T}^L, \mathcal{T}^L, \mathcal{E}^L, \mathcal{E}_j^L, \mathcal{E}_j^L \).

For any field extension \( L_1 \subseteq L_2 \) we have that \( \mathcal{E}^{L_1 \cap L_1} L_2 \cong \bigcup_{j \in \mathcal{T}_i} \mathcal{E}_j^{L_2} \), where \( T_i \) is some subset of \( \mathcal{T}^{L_2} \) and all \( \mathcal{E}_j^{L_2}, j \in T_i \), have the same numerical data as \( \mathcal{E}_i^{L_1} \).
1.4. Now we can start our construction of a suitable embedded resolution of \( f^{-1}\{0\} \) in \( \mathbb{A}^n(F) \).

1.4.1. First consider the blowing-up \( \pi : V \to \mathbb{A}^n(F) \) of \( \mathbb{A}^n(F) \) with center the origin. Denote the strict transform of \( f^{-1}\{0\} \) in \( V \) with \( \overline{f^{-1}\{0\}} \) and the inverse image of the origin (by \( \pi \)) with \( E_0 \). Remark that \( E_0 \cong \mathcal{P} \). Using the crucial ingredient that \( f \) is homogeneous the following facts are not difficult to verify:

(i) the intersection \( E_0 \cap \overline{f^{-1}\{0\}} \) is isomorphic to \( \mathcal{D}_{\text{red}} \);

(ii) there is an open covering \( \{V_i\} \) of \( E_0 \) (respectively \( \{U_i\} \) of \( E_0 \cap \overline{f^{-1}\{0\}} \)) such that \( V \) (respectively \( \overline{f^{-1}\{0\}} \)) is obtained by gluing products of the form \( \mathbb{A}^1 \times V_i \) (respectively \( \mathbb{A}^1 \times U_i \)).

1.4.2. Then we can find an embedded resolution \( (Y, h) \) of \( f^{-1}\{0\} \) in \( \mathbb{A}^n(F) \) over \( F \) by combining the point-centered blowing-up \( \pi \) with the resolution of (1.2), taking into account the product with \( \mathbb{A}^1 \) pointed out in (ii) of (1.4.1). Let \( E_i, i \in T = T_c \cup T_s \), be the (reduced) \( F \)-irreducible components of \( h^{-1}(f^{-1}\{0\}) \), where \( E_i \) is a (reduced) irreducible component of the exceptional divisor for \( i \in T_c \) and of the strict transform of \( f^{-1}\{0\} \) in \( Y \) for \( i \in T_s \). For each \( i \in T \) let \( N_i \) and \( \nu_i - 1 \) be the multiplicities of \( E_i \) in the divisor of respectively \( f \circ h \) and \( h^*(dx_1 \wedge \cdots \wedge dx_n) \) on \( Y \). The \( (N_i, \nu_i) \), for \( i \in T \), are called the numerical data of the resolution \( (Y, h) \). For \( i \in T \) and \( I \subseteq T \) we denote \( E_i^c := E_i \setminus \bigcup_{j \neq i} E_j \), \( E_I := \bigcap_{i \in I} E_i \) and \( E_i^s := E_i \setminus \bigcup_{j \in T \setminus I} E_j \).

1.4.3. If \( E_0 \) also denotes the strict transform of \( E_0 \) in \( Y \) (remark that \( 0 \in T_c \) in this situation), then the following remarks are easy consequences of the homogeneity of \( f \) and the choice of our embedded resolution \( (Y, h) \):

(i) The intersection \( E_0 \cap E_i \) for \( i \in T \setminus \{0\} \) is canonically isomorphic to one of the components \( E_j \) with \( j \in T \). This will give us a bijection between \( T \setminus \{0\} \) and \( T \), so from now on we will assume \( T = T \setminus \{0\} \).

(ii) Under this identification of \( T \) with \( T \setminus \{0\} \) also \( T_c \) and \( T_s \) coincide with respectively \( T_c \setminus \{0\} \) and \( T_s \), and corresponding \( N_i \) and \( \nu_i \) will be the same.

(iii) The second fact of (1.4.1) will also hold for \( Y \) (respectively \( E_i \)) instead of \( V \) (respectively \( \overline{f^{-1}\{0\}} \)).

(iv) The numerical data of \( E_0 \) are \( N_0 = \deg f \) and \( \nu_0 = n \); and \( h^{-1}\{0\} = E_0 \).

1.4.4. In the same way as in (1.3) we can extend everything in (1.4) to any field extension \( L \) of \( F \).

1.4.5. Remark. From now on we associate to any homogeneous polynomial the embedded resolutions and other notations of (1.2)-(1.4).
1.5. Now we are ready to state some general results in terms of the embedded resolutions constructed above. For the next two theorems we fix a homogeneous polynomial \( f \in F[x_1, \ldots, x_n] \setminus F \), where \( F \) is some number field.

**Theorem 1.5.1** ([2, Section 3]). For almost all completions \( K \) of \( F \) (i.e., for all except a finite number) we have the following for a character \( \kappa \) of \( \mathbb{R}^* \) of order \( d \).

(i) If \( \kappa \) is not trivial on \( 1 + P_K \), then \( Z^K(\kappa, f, s) \) and \( Z^0_K(\kappa, f, s) \) are constant on \( \mathbb{C} \).

(ii) If \( \kappa \) is trivial on \( 1 + P_K \), then

\[
Z^K(\kappa, f, s) = q^{-n} \sum_{I \subseteq T_K, \forall i \in I: d | N_i} C^K_{I, \kappa} \prod_{i \in I} q^{\nu_i + sN_i - 1},
\]

with \( C^K_{I, \kappa} = \sum_k (-1)^k \text{Tr}[\text{Frob}, H^k(\mathbb{Q}_K, \mathcal{L}_\kappa)] \).

Here \((\cdot)_K\) denotes reduction modulo \( P_K \), \( \mathcal{L}_\kappa \) is a certain \( \ell \)-adic sheaf on \((Y^K)_K\) associated to \( \kappa \), \( \text{Tr} \) denotes the trace, and \( \text{Frob} \) is the geometric Frobenius of \( K \).

For \( Z^0_K(\kappa, f, s) \) we have an analogous formula replacing \( C^K_{I, \kappa} \) by a similar constant \( C^K_{I, \kappa, 0} \).

(The explicit expression of \( C^K_{I, \kappa} \) is just given for completeness; we will not need it in this paper.)

**Theorem 1.5.2** ([5, Proposition 2 and Theorem 7]). For almost all completions \( K \) of \( F \) we have the following for a character \( \kappa \) of \( \mathbb{R}^* \) of order \( d \).

(i) If \( d \) does not divide \( \deg f \), then \( Z^K(\kappa, f, s) = Z^0_K(\kappa, f, s) = 0 \).

(ii) If \( d \) divides \( \deg f \) and \( \kappa \) is trivial on \( 1 + P_K \), then

\[
Z^K(\kappa, f, s) = \frac{(1 - q^{-1})q^{-(n-1)}}{1 - q^{-(\deg f)s-n}} \sum_{I \subseteq T_K, \forall i \in I: d | N_i} C^K_{I, \kappa} \prod_{i \in I} q^{\nu_i + sN_i - 1},
\]

where \( C^K_{I, \kappa} \) is a certain constant.

**1.5.3. Remark.** By adapting the proof of Theorem 7 in [5] we easily find that under the same conditions as in (1.5.2)(ii)

\[
Z^0_K(\kappa, f, s) = q^{-(\deg f)s-n} \frac{(1 - q^{-1})q^{-(n-1)}}{1 - q^{-(\deg f)s-n}} \sum_{I \subseteq T_K, \forall i \in I: d | N_i} C^K_{I, \kappa} \prod_{i \in I} q^{\nu_i + sN_i - 1}.
\]
1.5.4. Remark. Although we stated the two theorems above only for a homogeneous polynomial, Theorem 1.5.1 also holds for an arbitrary polynomial. But Theorem 1.5.2 is specific for the homogeneous case.

1.6. We now remind the definition of local monodromy [9]. Fix \( g \in \mathbb{C}[x_1, \ldots, x_n] \setminus \mathbb{C} \) and \( b \in \mathbb{C}^n \) with \( g(b) = 0 \). Let \( B \subseteq \mathbb{C}^n \) be a small enough ball with center \( b \); the restriction \( g|_B \) is a locally trivial \( \mathcal{C}^\infty \) fibration over a small enough pointed disc \( D \subseteq \mathbb{C} \setminus \{0\} \) with center 0. Hence the diffeomorphism type of the Milnor fiber \( M_{(g,b)} := g^{-1}\{t\} \cap B \) of \( g \) around \( b \) does not depend on \( t \in D \), and the counterclockwise generator of the fundamental group of \( D \) induces an automorphism of \( H^q(M_{(g,b)}, \mathbb{C}) \) which is called the local monodromy of \( g \) at \( b \). By an eigenvalue of the local monodromy of \( g \) at \( b \) we mean an eigenvalue of the monodromy action on (at least) one of the \( H^q(M_{(g,b)}, \mathbb{C}) \) for \( q = 0, \ldots, n-1 \).

1.6.1. Remark. In the same way as in (1.4.2) we associate to an arbitrary embedded resolution of \( g^{-1}\{0\} \) in \( \mathbb{A}^n(\mathbb{C}) \) the notations of that section. Now fix such an embedded resolution.

Theorem 1.6.2 ([1, Theorem 3]). For \( b \in g^{-1}\{0\} \) let \( P_q(t) \) denote the characteristic polynomial of the monodromy action on \( H^q(M_{(g,b)}, \mathbb{C}) \) for \( q = 0, \ldots, n-1 \). Then

\[
\prod_{q=0}^{n-1} (P_q(t))^{(-1)^{q+1}} = \prod_{i \in T} (1 - t^{N_i})^{-\chi(E^q_i \cap h^{-1}\{b\})}.
\]

1.6.3. In particular if \( b \) is the origin and if \( g \) is homogeneous, then (1.4.3)(iv) implies that \( \prod_{q=0}^{n-1} (P_q(t))^{(-1)^{q+1}} = (1 - t^{\deg g})^{-\chi(E^q_0)} \). This assertion is also classically known and follows for example from [9, Section 9].

2. Holomorphy conjecture for homogeneous polynomials.

2.1. In this section we will use the embedded resolutions of Section 1 to provide in arbitrary dimension an easy proof of the holomorphy conjecture for homogeneous polynomials under the additional characteristic-assumption mentioned before in the introduction. In the next section we will treat the case in which this assumption is not fulfilled.

Theorem 2.2. Let \( F \) be a number field and \( f \in F[x_1, \ldots, x_n] \setminus F \) a homogeneous polynomial such that \( \chi(\mathbb{P}^{n-1}_\mathbb{C} \setminus \{f = 0\}) \neq 0 \). For almost all completions \( K \) of \( F \) we have the following for any character \( \kappa \) of \( \mathbb{R}^+_K \). If the order of \( \kappa \) does not divide the order of any eigenvalue of the (complex) local monodromy of \( f \) at any point of \( f^{-1}\{0\} \), then \( Z^K(\kappa, f, s) \) and \( Z_0^K(\kappa, f, s) \) are holomorphic on \( \mathbb{C} \). In fact they are identically zero.
Proof. We use the construction and notation of (1.2)-(1.4). It follows from this construction (especially from (1.4.1)(i)) that $\chi(E^o_0) = \chi(\mathbb{P}^{n-1}_\mathbb{C} \setminus \{f = 0\}) \neq 0$. So (1.6.3) implies that $e^{2\pi i \deg f}$ is an eigenvalue of the local monodromy of $f$ at the origin. From the conditions in the holomorphy conjecture we find that the order of the character $\kappa$ does not divide $\deg f$, which gives us the result by Theorem 1.5.2(i). \qed

2.3. Now we introduce the so-called topological zeta function $Z^{(r)}_{\text{top}}(g, s)$, which is associated to $g \in \mathbb{C}[x_1, \ldots, x_n]$ and $r \in \mathbb{N} \setminus \{0\}$ by Denef and Loeser [4, Section 3]. With the notations from remark (1.6.1) we have that

$$Z^{(r)}_{\text{top}}(g, s) = \sum_{I \subseteq T} \chi(E^o_I) \prod_{i \in I} \frac{1}{\nu_i + N_i s},$$

with $s \in \mathbb{C}$. Replacing $\chi(E^o_I)$ by $\chi(E^o_I \cap h^{-1}\{0\})$ we analogously define $Z^{(r)}_{\text{top}, 0}(g, s)$. When $g$ is homogeneous it will easily follow from the proof of Theorem 2.4 that $Z^{(r)}_{\text{top}}(g, s) = Z^{(r)}_{\text{top}, 0}(g, s)$. So from now on we just have to deal with one of them. We can also formulate a holomorphy conjecture for this topological zeta function.

**Conjecture 2.3.1.** If $r \in \mathbb{N} \setminus \{0\}$ does not divide the order of any eigenvalue of the local monodromy of $g$ at any point of $g^{-1}\{0\}$, then $Z^{(r)}_{\text{top}}(g, s)$ is holomorphic on $\mathbb{C}$.

**Theorem 2.4.** Let $f \in \mathbb{C}[x_1, \ldots, x_n] \setminus \mathbb{C}$ be a homogeneous polynomial, with $\chi(\mathbb{P}^{n-1}_\mathbb{C} \setminus \{f = 0\}) \neq 0$. Then the holomorphy conjecture is true for $Z^{(r)}_{\text{top}}(f, s)$.

Proof. Let $r \in \mathbb{N} \setminus \{0\}$ such that $r$ does not divide the order of any eigenvalue of the local monodromy of $f$ at any point of $f^{-1}\{0\}$. As in the previous proof we find that $r$ does not divide $\deg f$, so with the notations of Section 1 the topological zeta function reduces to

$$Z^{(r)}_{\text{top}}(f, s) = \sum_{0 \notin I \subseteq T \forall i \in I : r \mid N_i} \chi(E^o_I) \frac{1}{\prod_{i \in I} \nu_i + N_i s}.$$

Now consider a subset $I$ of $T$ with $0 \notin I$ and $\forall i \in I : r \mid N_i$. From (1.4.3)(iii) we know that (locally) $E^o_I = H_I \times (\mathbb{A}^1 \setminus \{\text{point}\})$, where $H_I$ is some subvariety of $E_0$. Because $\chi(\mathbb{A}^1 \setminus \{\text{point}\}) = 0$, we can easily conclude that $\chi(E^o_I) = 0$, implying $Z^{(r)}_{\text{top}}(f, s)$ to be zero. \qed
3. The case of Euler-Poincaré characteristic zero.

3.1. When \( \chi(\mathbb{P}^{n-1}_\mathbb{C} \setminus \{f = 0\}) = 0 \) for the homogeneous polynomial \( f \) in question, we are still able to prove the holomorphy conjecture by assuming a sort of ‘projective holomorphy conjecture in \( \mathbb{P}^{n-1} \)’. For \( n = 3 \) this will actually give us a tool to prove the holomorphy conjecture by using the fact that the holomorphy conjecture is true for curves. (See [10].)

3.2. Suppose that \( K \) is the completion of a number field \( F \) with respect to some maximal ideal of its ring of integers and that \( \kappa \) is a character of \( R_K^\times \) of order \( d \) such that \( d \) divides \( \deg f \) and \( \kappa \) is trivial on \( 1 + P_K \). We define the projective local zeta function associated to \( f \) and \( \kappa \) to be

\[
Z^K_{\text{proj}}(\kappa, f, s) = \frac{1 - q^{-(\deg f)s-n}}{1 - q^{-1}} Z^K(\kappa, f, s).
\]

By Theorem 1.5.2(ii) we have, using the notations of Section 1, that

\[
Z^K_{\text{proj}}(\kappa, f, s) = q^{-(n-1)} \sum_{I \subset T^K} \prod_{i \in I} \frac{q - 1}{q^{\nu_i + sN_i} - 1},
\]

for almost all fields \( K \) and all characters \( \kappa \) as above. By comparing this expression with Theorem 1.5.1(ii) the underlying inspiration for this definition should be clear.

3.3. Under the holomorphy conjecture for \( Z^K_{\text{proj}}(\kappa, f, s) \) we understand: For almost all fields \( K \) and for all characters \( \kappa \) as in (3.2) we have the following. If the order of \( \kappa \) does not divide the order of any eigenvalue of the (complex) local monodromy of any \( f_j \) at any complex point of \( f_j^{-1}\{0\} \), then \( Z^K_{\text{proj}}(\kappa, f, s) \) is holomorphic on \( \mathbb{C} \).

Here \( f_j \) denotes the polynomial you get by putting \( x_j = 1 \) in the homogeneous polynomial \( f \). Before stating the next theorem we formulate a lemma that we will need in its proof.

**Lemma 3.4** ([3, proof of Proposition 3.4]). Let \( g \in \mathbb{C}[x_1, \ldots, x_n] \setminus \mathbb{C} \). If \( \lambda \) is an eigenvalue of the local monodromy of \( g \) at some point \( b \) on \( g^{-1}\{0\} \), then there is a point \( c \) on \( g^{-1}\{0\} \) such that \( \lambda \) is a zero or a pole of the alternating product of the characteristic polynomials of the monodromy action on \( H^i(M_{(g,c)}, \mathbb{C}) \) for \( i = 0, \ldots, n-1 \).

**Theorem 3.5.** Let \( F \) be a number field and \( f \in F[x_1, \ldots, x_n] \setminus F \) a homogeneous polynomial. The holomorphy conjecture for \( Z^K_{\text{proj}}(\kappa, f, s) \) implies the holomorphy conjecture for \( Z^K(\kappa, f, s) \) (and for \( Z^0_K(\kappa, f, s) \)).
Remark. Of course this theorem holds when \( \chi(\mathbb{P}^{n-1}_C \setminus \{ f = 0 \}) \neq 0 \), but we know from the proof of Theorem 2.2 that this condition implies that the order of \( \kappa \) does not divide \( \deg f \). So in the light of Theorem 1.5.2(i), we only have to deal with a nontrivial case if \( \chi(\mathbb{P}^{n-1}_C \setminus \{ f = 0 \}) = 0 \).

Proof of Theorem 3.5. By Theorem 1.5.1(i) and Theorem 1.5.2(i) we may assume that \( \kappa \) is trivial on \( 1 + P_K \) and that the order \( d \) of \( \kappa \) divides \( \deg f \). So we let \( K \) and \( \kappa \) be as in (3.2).

Suppose that the order of \( \kappa \) divides the order of some eigenvalue \( \lambda \) of the local monodromy of some \( f_j \) at some point \( b \) of \( f_j^{-1}\{0\} \). Then by Lemma 3.4 we may assume that \( \lambda \) is a zero or a pole of the alternating product of the characteristic polynomials of the monodromy action on \( H^i(M(f_j,b), \mathbb{C}) \) for \( i = 0, \ldots, n-2 \). Now, using Theorem 1.6.2 and the construction in (1.4) (especially the results of (1.4.3)), it is not hard to find a point \( c \) (different from the origin) of \( f_j^{-1}\{0\} \) such that the product above equals the alternating product of the characteristic polynomials of the monodromy action on \( H^i(M(f,c), \mathbb{C}) \) for \( i = 0, \ldots, n-1 \). So \( \lambda \) will also be an eigenvalue of the local monodromy of \( f \) at \( c \).

We can conclude that the condition in the holomorphy conjecture (for \( Z^K(\kappa, f, s) \) and \( Z^K_0(\kappa, f, s) \)) implies the condition in the holomorphy conjecture for \( Z^K_{\text{proj}}(\kappa, f, s) \), and hence by assumption that \( Z^K_{\text{proj}}(\kappa, f, s) \) is holomorphic on \( \mathbb{C} \) (for almost all completions \( K \) of \( F \)). Because \( Z^K_{\text{proj}}(\kappa, f, s) \) is a rational function in \( q^{-s} \) (of non-positive degree), this yields that \( Z^K_{\text{proj}}(\kappa, f, s) \) is constant as function of \( s \), and more concretely

\[
Z^K_{\text{proj}}(\kappa, f, s) = \frac{1}{q^{n-1}C^K_{\emptyset, \kappa}}.
\]

From (1.5.3) and (3.2) we find that

\[
Z^K_0(\kappa, f, s) = q^{-(\deg f)s-n} Z^K(\kappa, f, s) = \frac{(1-q^{-1})q^{-(\deg f)s-n}}{1-q^{-(\deg f)s-n}} \frac{1}{q^{n-1}} C^K_{\emptyset, \kappa},
\]

which is of degree zero (as rational function in \( q^{-s} \)) if \( C^K_{\emptyset, \kappa} \) is different from zero. But since Theorem 1.2 in [3] says that for almost all completions \( K \) of \( F \) the degree of \( Z^K_0(\kappa, f, s) \) has to be strictly negative, we can conclude that \( C^K_{\emptyset, \kappa} = 0 \) and so that \( Z^K(\kappa, f, s) = 0 \) (and of course also that \( Z^K_0(\kappa, f, s) = 0 \)). \( \square \)
3.6. Topological zeta function. Let \( f \in \mathbb{C}[x_1, \ldots, x_n] \setminus \mathbb{C} \) be a homogeneous polynomial and \( r \in \mathbb{N} \setminus \{0\} \) such that \( r \) divides \( \deg f \). Looking at the definition of \( Z_{\text{top}}^{(r)}(f, s) \), but now thinking projectively, we are stimulated to define
\[
Z_{\text{top,proj}}^{(r)}(f, s) := \sum_{I \subseteq T} \chi(I^S) \prod_{i \in I} \frac{1}{\nu_i + sN_i}.
\]
Then by the proof of Theorem 2.4 and by the facts of (1.4.3) we see that \( Z_{\text{top}}^{(r)}(f, s) = 1^{n+(\deg f)s}Z_{\text{top,proj}}^{(r)}(f, s) \), so obviously \( Z_{\text{top,proj}}^{(r)}(f, s) \) is independent of the chosen embedded resolution of \( D \) in \( \mathcal{P} \) as in (1.2).

3.6.1. The holomorphy conjecture for \( Z_{\text{top,proj}}^{(r)}(f, s) \) can be formulated as follows. If \( r \in \mathbb{N} \setminus \{0\} \) does not divide the order of any eigenvalue of the local monodromy of any \( f_j \) at any point of \( f_j^{-1}\{0\} \) and if \( r|\deg f \), then \( Z_{\text{top,proj}}^{(r)}(f, s) \) is holomorphic on \( \mathbb{C} \).

3.6.2. Under the condition \( \chi(\mathbb{P}_F^{n-1} \setminus \{f = 0\}) = 0 \) the analogue of Theorem 3.5 will hold for the topological zeta function. Indeed, if \( r \not| \deg f \), then \( Z_{\text{top}}^{(r)}(f, s) \) is holomorphic on \( \mathbb{C} \) by the proof of Theorem 2.4. If \( r|\deg f \), then the same idea as in the proof of Theorem 3.5 will give us (under the assumption of the holomorphy conjecture for \( Z_{\text{top,proj}}^{(r)}(f, s) \)) that \( Z_{\text{top}}^{(r)}(f, s) = \frac{1}{n+(\deg f)s}\chi(\mathbb{P}_F^{n-1} \setminus \{f = 0\}) = 0 \).

Theorem 3.7. For a homogeneous polynomial \( f \) in three variables the holomorphy conjectures for Igusa’s local zeta functions \( Z^K(\kappa, f, s) \) and \( Z_0^K(\kappa, f, s) \) and for the topological zeta function \( Z_{\text{top}}^{(r)}(f, s) \) are true.

Proof. From the previous results it is clear that we just have to prove the holomorphy conjecture for \( Z_{\text{proj}}^{(r)}(\kappa, f, s) \) (and \( Z_{\text{top,proj}}^{(r)}(f, s) \)).

Take for the embedded resolution \((\chi^F, \varphi^F)\) of (1.2) the (scheme-theoretical) canonical embedded resolution of \( D^F \) in \( \mathcal{P}^F \). By restricting we get the canonical embedded resolution of \( f_j^{-1}\{0\} \) in \( \mathcal{A}^2(F) \) for any \( j \in \{1, 2, 3\} \), so we can use on every affine chart the same arguments as in the proof of the holomorphy conjecture for curves [10]. By noting that the number \( N_i \) of an irreducible component of the strict transform does not change by passing to an affine chart, the proof of Theorem 3.7 is done.

4. The monodromy conjecture.

4.1. Under the usual characteristic-assumption we will show in this section how to prove the so-called monodromy conjecture of Igusa for an homogeneous polynomial in three variables using mainly the same ideas as in the proof of Theorem 3.7.
Theorem 4.2. Let $F$ be a number field and $f \in F[x_1, x_2, x_3] \setminus F$ a homogeneous polynomial such that $\chi(\mathbb{P}_C^2 \setminus \{f = 0\}) \neq 0$. For almost all completions $K$ of $F$ we have the following for any character $\kappa$ of $R^K$. If $s_0$ is a pole of $Z^K(\kappa, f, s)$, then $\exp(2\pi \sqrt{-1} \Re(s_0))$ is an eigenvalue of the local monodromy of $f$ at some complex point of $f^{-1}\{0\}$.

Proof. As in the proof of Theorem 3.7 we take for the embedded resolution $(\mathcal{O}^{E}, \varphi^{E})$ of (1.2) the (scheme-theoretical) canonical embedded resolution of $\mathcal{O}^{E}$ in $\mathcal{P}^{E}$. By Theorem 1.5.1(i) and Theorem 1.5.2(ii) we may assume that $Z^K(\kappa, f, s) = \frac{1}{1-q^{-\deg f}} Z^K_{\text{proj}}(\kappa, f, s)$. Because $\exp(2\pi i(\frac{-3}{\deg f}))$ is an eigenvalue of the local monodromy of $f$ at the origin (see the proof of Theorem 2.2), we may also assume that $s_0$ is a pole of $Z^K_{\text{proj}}(\kappa, f, s)$. As in the proof of Theorem 5.2.1 in [2], we find that $\Re(s_0) = \frac{-\nu_j}{\pi}$ for some $j \in T^C$, with $|E^C_j \setminus E^0_{\text{red}}| \geq 3$ or $j \in T^C_s$. First suppose that $j \in T^C$. Then for some $k \in \{1, 2, 3\}$ there is an irreducible component of $f^{-1}_k\{0\}$ in $\mathbb{A}^2(C)$, whose strict transform (by the restriction of $\varphi^C$) has numerical data $(N_j, \nu_j) = (N_j, 1)$. It is well-known that in this case $\exp(2\pi i(\frac{-\nu_j}{\deg f}))$ is an eigenvalue of the local monodromy of $f_k$ at some nonsingular point of $f^{-1}_k\{0\}$. As in the proof of Theorem 3.5 this implies that $\exp(2\pi i \Re(s_0))$ is an eigenvalue of the local monodromy of $f$ at some complex point of $f^{-1}\{0\}$. Next suppose that $j \notin T^C$ and $|E^C_j \setminus E^0_{\text{red}}| \geq 3$. Then $E^C_j$ will also be an exceptional curve of the restriction of $\varphi^C$ to some affine chart, which is in fact the canonical embedded resolution of $f^{-1}_k\{0\}$ in $\mathbb{A}^2(C)$ for some $k \in \{1, 2, 3\}$. Now it is known, see for example [8], that then $\exp(2\pi i \Re(s_0))$ is an eigenvalue of the local monodromy of $f_k$ at some point of $f^{-1}_k\{0\}$. Then again we can conclude that $\exp(2\pi i \Re(s_0))$ is also an eigenvalue of the local monodromy of $f$ at some complex point of $f^{-1}\{0\}$. This completes the proof of Theorem 4.2.

4.2.1. Remark. From Remark 1.5.3 we know that

$$Z^K_0(\kappa, f, s) = q^{-(\deg f)s-n} Z^K(\kappa, f, s),$$

so the monodromy conjecture is clearly also true for $Z^K_0(\kappa, f, s)$ when $\chi(\mathbb{P}_C^2 \setminus \{f = 0\}) \neq 0$.

4.2.2. Remark. In a similar way we can prove the monodromy conjecture for the topological zeta function (again for a homogeneous polynomial $f$ in three variables and under the condition that $\chi(\mathbb{P}_C^2 \setminus \{f = 0\}) \neq 0$). This conjecture states the following for $r \in \mathbb{N} \setminus \{0\}$. If $s_0$ is a pole of $Z^{(r)}_{\text{top}}(f, s)$, then $\exp(2\pi \sqrt{-1}s_0)$ is an eigenvalue of the local monodromy of $f$ at some complex point of $f^{-1}\{0\}$.
References


Received February 1, 2000 and revised November 8, 2000. The first author is a Research Assistant of the Belgian Fund for Scientific Research - Flanders.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF LEUVEN
CELESTIJENLAAN 200B
B-3001 LEUVEN
BELGIUM
E-mail address: bart.rodrigues@wis.kuleuven.ac.be

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF LEUVEN
CELESTIJENLAAN 200B
B-3001 LEUVEN
BELGIUM
E-mail address: wim.veys@wis.kuleuven.ac.be
HAUSDORFF DIMENSION OF LIMIT SETS FOR PARABOLIC IFS WITH OVERLAPS

K. Simon, B. Solomyak, and M. Urbański

We study parabolic iterated function systems with overlaps on the real line. We show that if a $d$-parameter family of such systems satisfies a transversality condition, then for almost every parameter value the Hausdorff dimension of the limit set is the minimum of 1 and the least zero of the pressure function. Moreover, the local dimension of the exceptional set of parameters is estimated. If the least zero is greater than 1, then the limit set (typically) has positive Lebesgue measure. These results are applied to some specific families including those arising from a class of continued fractions.

1. Introduction.

Let $\Phi = \{\phi_1, \ldots, \phi_k\}$ be a collection of self-maps on a closed interval $X \subset \mathbb{R}$. We call $\Phi$ an iterated function system (IFS). Under standard contractivity hypotheses, there is a unique non-empty compact set $J_{\Phi}$ such that $J_{\Phi} = \bigcup_{j=1}^{k} \phi_j(J_{\Phi})$, called the limit set, or attractor, of the IFS.

If the sets $\phi_j(X)$ are mutually disjoint, then $J_{\Phi}$ is a Cantor set. If, in addition, $\phi_j$ are monotone, the limit set is known as a “cookie-cutter”; then it is more common to view $J_{\Phi}$ as the repeller of an expanding map $f : \bigcup_{j=1}^{k} \phi_j(X) \to X$ defined by $f(x) = \phi_j^{-1}(x)$ for $x \in \phi_j(X)$. Suppose that all the maps $\phi_j$ are in $C^{1+\theta}(X)$ for some $\theta \in (0, 1]$ and are hyperbolic, that is, $0 < |\phi_j'(x)| < 1$ on $X$. Then the Hausdorff dimension $\dim_H(J_{\Phi})$ is given by Bowen’s formula [Bo2, R]:

$$\dim_H(J_{\Phi}) = s(\Phi) \quad \text{where} \quad P_{\Phi}(s(\Phi)) = 0.$$ 

Here $P_{\Phi}(t)$ is the pressure function, which can be defined by

$$P_{\Phi}(t) = \lim_{n \to \infty} \frac{1}{n} \log \sum_{\omega \in I^n} \|\phi_{\omega}'\|^t$$

where $I = \{1, \ldots, k\}$, $\phi_{\omega} = \phi_{\omega_1} \circ \cdots \circ \phi_{\omega_n}$, and $\|\cdot\|$ is the supremum norm on $X$.

The dimension formula was extended in [U] to the parabolic case, where some of the maps have a neutral fixed point. A parabolic IFS is not uniformly contracting, which makes the situation more subtle. It was proved in [U]...
that the Hausdorff dimension of the limit set is the least zero of the pressure function. In contrast with the hyperbolic case (where the zero is unique), in the parabolic case the pressure function is identically zero for all \( t \) larger than the least zero.

We study the case when the “pieces” of the limit set are allowed to overlap. One-dimensional IFS with overlaps arise naturally in the study of higher-dimensional dynamical systems \([\text{PrU, BU, Si1, Si2, SSo}]\) (in this sentence the “dimension” refers to the phase space rather than to the limit set). Moreover, IFS with overlaps occur in some problems on random matrix products, random continued fractions, and in prediction theory, see \([\text{Pi, Ly, LL}]\).

There are many open problems related to IFS with overlaps, which are notoriously difficult even when the maps are linear, see \([\text{PSo3}]\). The dimension of the limit set may be strictly less than the least zero of the pressure function, for instance, if \( \phi_\omega \equiv \phi_\tau \) for two distinct words \( \omega \) and \( \tau \). Since it is often hard to analyze an individual system, one can try to investigate a “typical” (in the sense of Lebesgue measure) IFS in a parameterized family. This method was first used by Falconer \([\text{F1}]\) who considered families of linear contractions with a linear dependence on parameter. Further work in this direction was done in \([\text{PoS, So1, PSo2, So2, SSo}]\). An important role in these papers was played by a certain “transversality condition”, which controls the way the IFS depends on parameters.

In this paper this approach is extended to a class of non-linear IFS. Our main result (Theorem 6.1) states that if a parameterized family \( \{\Phi_t\} \) of parabolic IFS satisfies the transversality condition, then for Lebesgue-a.e. parameter \( t \) the Hausdorff dimension of the limit set is given by

\[
\dim_H(J_{\Phi_t}) = \min\{1, s(t)\} \quad \text{where} \quad s(t) := \min\{s : P_{\Phi_t}(s) = 0\}.
\]

Moreover, the limit set \( J_{\Phi_t} \) has positive Lebesgue measure for a.e. \( t \) such that \( s(t) > 1 \). If a slightly stronger version of transversality is imposed, then the local dimension of the exceptional set in the first statement can be estimated above by \( s(t) + (d - 1) \), where \( d \) is the number of independent parameters involved.

We illustrate our results by the following example (see Corollaries 7.4 and 7.5). Let \( \phi(x) = \frac{x}{1+x} \). This function is parabolic on \([0,1]\). Let \( A = \{a_1, \ldots, a_k\} \), with \( k \geq 3 \), and consider the family of IFS on \([0,1]\)

\[
\Phi_A = \{\phi(x + a_j)\}_{j=1}^k
\]

for \( A \in U := \{A \in \mathbb{R}^k : a_k = 0, \ a_j > 0, \ j = 1, \ldots, k-1\} \).

Denote by \( J_A \) the limit set of the IFS \( \Phi_A \) and let \( s(A) = \min\{t > 0 : P_{\Phi_A}(t) = 0\} \). We will prove that \( s : U \to \mathbb{R} \) is a continuous function, so \( U_{<1} = \{A \in U : s(A) < 1\} \) and \( U_{>1} = \{A \in U : s(A) > 1\} \) are open sets.
Proposition 1.1.  

(i) For Lebesgue-a.e. \( A \in U \),
\[
\dim_H(J_A) = \min\{s(A), 1\}.
\]

(ii) For any subset \( G \subset U_{<1} \) we have
\[
\dim_H\{A \in G : \dim_H(J_A) < s(A)\} \leq \sup_G s(A) + (k - 2).
\]

(iii) For Lebesgue-a.e. \( A \in U_{>1} \) the set \( J_A \) has positive Lebesgue measure.

(iv) Similar results hold for the one-parameter family \( \Phi_A \) where \( A = \{a, 2, 0\} \) and \( a \in (0, 2) \).

Remarks.  

1. We assumed that \( k \geq 3 \) since for \( k = 2 \) either the limit set is an interval, or the IFS has no overlaps, so the result is true for all parameters by \([U]\).

2. Proposition 1.1(iii) concerning the positive measure of the limit set reflects a phenomenon which cannot occur in the non-overlapping case. It is an open problem whether such limit sets can be “fat” Cantor sets or they necessarily contain intervals.

3. The limit set \( J_A \) can be described as the set of continued fractions of the form
\[
y = [1, Y_1, 1, Y_2, 1, Y_3, \ldots] = \frac{1}{1 + \frac{1}{Y_1 + \frac{1}{1 + \ldots}}}
\]
where \( Y_i \in A \). The dimension of sets arising by some restriction in their expansions (continued fractions, \( \lambda \)-expansions, etc.) was studied by many authors. IFS with overlaps arise when the expansion for some numbers is non-unique. The family of linear IFS \( \{\lambda x, \lambda x + 1, \lambda x + 3\} \), with \( \lambda \in (\frac{1}{4}, \frac{2}{5}) \), investigated in \([KSS, PoS, So1]\), was an important “testing ground” in the study of IFS with overlaps. The family \( \{\phi(x), \phi(x + \alpha), \phi(x + 2)\} \), with \( \alpha \in (0, 2) \), that we consider in Proposition 1.1(iv) is a non-linear parabolic analog.

4. A related problem is to analyze invariant (stationary) measures on the limit set of an IFS. The fundamental question is whether this measure is singular or absolutely continuous. This is interesting already for \( k = 2 \), when the limit set is an interval. R. Lyons \([Ly]\) investigated a family of such measures for the IFS \( \Phi_A \) with \( A = \{\alpha, 0\} \). He showed singularity for a certain interval of parameters and asked if the measure is absolutely continuous for small \( \alpha \). In \([SSU2]\), using some of the techniques developed in this paper, we establish that the invariant measure is indeed absolutely continuous for a.e. \( \alpha \) in some interval.

Here is a brief outline of the contents of the paper.

Section 2 contains preliminaries concerning infinite hyperbolic IFS, including properties of the pressure function.
Section 3 deals with families of infinite hyperbolic IFS depending on parameters. Its main result, which is of independent interest, is Theorem 3.1, which computes the Hausdorff dimension and Lebesgue measure of the limit set of a.e. infinite hyperbolic IFS with overlaps from a family satisfying a transversality condition.

The exceptional set of parameters, associated with Theorem 3.1, is analyzed in Section 4 where we estimate its local dimension from above.

In Section 5 we consider a single parabolic IFS. Following the approach of [MU2, MU3], we reduce the parabolic IFS to an infinite hyperbolic IFS. The limit sets of the parabolic and infinite hyperbolic systems differ in a countable set, so they have the same dimension. We prove that the unique zero of the pressure for the infinite hyperbolic IFS coincides with the least zero of the pressure for the parabolic IFS, even though the pressure functions for these systems differ (see Proposition 5.10(ii)).

In Section 6 we study families of parabolic IFS. Our main result, Theorem 6.1, computes the Hausdorff dimension and Lebesgue measure of the limit set of a.e. parabolic IFS with overlaps in a family satisfying a transversality condition. Moreover, applying the results of Section 4 we estimate the local dimension of the exceptional set, in the spirit of Proposition 1.1(ii).

Section 7 is devoted to examples. We consider two general classes of examples. The most difficult part in applying Theorem 6.1 is checking the transversality condition. In Propositions 7.1 and 7.2 we obtain effective sufficient conditions for transversality. We conclude with specific examples arising from continued fractions.

Notation. We write $B_δ(t_0)$ for the open ball of radius $δ$ centered at $t_0$ and $L_d$ for the Lebesgue measure in $\mathbb{R}^d$. If $μ$ is a measure we often write $μA$ without parentheses. The symbol $≤$ means that the inequality holds up to an absolute multiplicative constant, and $≃$ means that both $≤$ and $≥$ are true. $\text{Int}(X)$ denotes the interior of a set $X$.

2. Preliminaries.

Let $X \subset \mathbb{R}$ be a closed interval. We consider a collection $Ψ = \{ψ_i\}_{i ∈ I}$ of continuous self-maps of $X$, where the set $I$ may be finite or countable. We set $I^* := \bigcup_{n≥1} I^n$ and denote by $I^∞$ the set of all infinite sequences of elements of $I$. If $ω ∈ I^*$, then by $|ω|$ we denote the length of $ω$. If $ω ∈ I^* ∪ I^∞$ and $|ω| > n$ then $ω|_n = ω_1ω_2⋯ω_n$ is the word consisting of the first $n$ letters of $ω$ and $σ^nω = ω_{n+1}ω_{n+2}⋯ω_{|ω|}$. The shift map $σ : I^∞ → I^∞$ sends an element $\{ω_k\}_{k=1}^∞$ to the element $\{ω_{k+1}\}_{k=1}^∞$. If $ω ∈ I^n$, then by $ψ_ω : X → X$ we denote the composition $ψ_ω = ψ_2 ◦ ψ_2 ◦ ⋯ ◦ ψ_ω$. Notice that given $ω ∈ I^∞$, the sequence of compact sets $\{ψ_ω|_n(X)\}_{n=1}^∞$ is descending and therefore $\bigcap_{n≥1} ψ_ω|_n(X) ≠ \emptyset$. If for every $ω ∈ I^∞$ this intersection is a
singleton, the collection $\Psi$ is said to be a **topological IFS**. We can then define the map $\pi_\Psi : I^\infty \to X$ by setting

$$\{ \pi_\Psi(\omega) \} = \bigcap_{n \geq 1} \psi_{\omega|_n}(X).$$

We call this map the **natural projection** induced by the topological IFS $\Psi$, and its range, the set $J_\Psi = \pi_\Psi(I^\infty)$, is called the **limit set** of $\Psi$. Thus,

$$\lim_{n \to \infty} \text{diam} \left( \psi_{\omega|_n}(X) \right) = 0$$

and therefore,

$$\pi_\Psi(\omega) = \lim_{n \to \infty} \psi_{\omega|_n}(x)$$

for every $x \in X$. We also have the following useful identity:

$$\pi_\Psi(\omega) = \psi_{\omega|_n}(\pi_\Psi(\sigma^n \omega)) \quad \text{for any } n \geq 1.$$  \hfill (2.1)

The limit set satisfies $J_\Psi = \bigcup_{i \in I} \psi_i(J_\Psi)$ but it need not be compact when $I$ is infinite [MU1]. We call $\Psi$ a **smooth IFS** if the following condition is satisfied.

**Smoothness:** There exists $\theta \in (0, 1]$ such that

$$\psi_i \in C^{1+\theta}(X \to \text{Int}(X)), \quad \text{and} \quad \psi_i'(x) \neq 0 \quad \text{for all } x \in X \text{ and } i \in I.$$  \hfill (2.2)

Given $t \geq 0$ we define the pressure function $P_\Psi(t)$ by the formula

$$P_\Psi(t) = \lim_{n \to \infty} \frac{1}{n} \log Z_n(\Psi, t)$$

where $Z_n(\Psi, t) = \sum_{|\omega| = n} \| \psi_\omega' \|^t$ and $\| \cdot \|$ denotes the supremum norm on $X$. Observe that the limit in (2.3) really exists since the sequence $n \mapsto \log Z_n(\Psi, t)$ is subadditive. Comparing this with the usual definition of topological pressure (see e.g., [Bo1]) we see that $P_\Psi(t)$ equals the pressure of the shift map $\sigma$ on $I^\infty$ with the potential

$$\omega \mapsto t \log |\psi_{\omega_1}'(\pi(\sigma \omega))|.$$  \hfill (2.3)

We call a smooth IFS **hyperbolic** if the following conditions are satisfied:

**Hyperbolicity:** For some $\gamma \in (0, 1),$

$$\| \psi_i' \| \leq \gamma < 1, \quad \text{for all } i \in I$$

(2.4)

(any map satisfying this property will be called hyperbolic), and

**Bounded Distortion Property:** There exists $K > 1$ such that for all $n \in \mathbb{N}$ and $\omega \in I^n$,

$$K^{-1} \leq \frac{|\psi_{\omega}(x)|}{|\psi_{\omega}(y)|} \leq K \quad \text{for all } x, y \in X.$$  \hfill (2.5)

Such $\Psi$ is a one-dimensional example of **conformal IFS**, introduced (for infinite $I$) and explored in [MU1], where also the open set condition was
assumed. It is well-known that (2.5) follows from (2.2) and (2.4) when $I$ is finite, see e.g., [B1] (see [MU1] for more general sufficient conditions).

From now on throughout the Sections 2-4, unless otherwise stated, we assume that the smooth IFS $\Psi$ is hyperbolic. Let
\begin{equation}
\Theta(\Psi) = \inf \{ t \geq 0 : P_\Psi(t) < \infty \}.
\end{equation}
The behavior of the pressure function is described in the following lemma.

**Lemma 2.1.** The function $t \mapsto P_\Psi(t)$, for $t \in (\Theta(\Psi), \infty)$, is finite, strictly decreasing and continuous, and $\lim_{t \to \infty} P_\Psi(t) = -\infty$.

**Proof.** The first statement is immediate from the definition of the number $\Theta = \Theta(\Psi)$. Now, given $t > \Theta$ and $s > 0$, we have by (2.4) for all $n \geq 1$:
\begin{equation}
Z_n(\Psi, t + s) = \sum_{|\omega| = n} \| (\psi_\omega)' \|^t s \leq \sum_{|\omega| = n} \| (\psi_\omega)' \|^t \gamma^{ns},
\end{equation}
and therefore, $P_\Psi(t + s) \leq s \log \gamma + P_\Psi(t) < P_\Psi(t)$. Thus, the function $t \mapsto P_\Psi(t)$ is strictly decreasing on $t \in (\Theta, \infty)$. Now, an application of Hölder’s inequality shows that each function $t \mapsto Z_n(\Psi, t)$ is log convex. Therefore the function $t \mapsto P_\Psi(t)$, $t \in (\Theta, \infty)$, is convex and, consequently, continuous. By the definition of $\Theta$ we have $P_\Psi(\Theta + 1) < \infty$. Hence, for every $t > 0$ and every $n \geq 1$,
\begin{equation}
Z_n(\Psi, \Theta + 1 + t) = \sum_{|\omega| = n} \| \psi_\omega' \|^{\Theta + 1 + t} \leq \sum_{|\omega| = n} \| \psi_\omega' \|^{1+\Theta} \cdot \| \psi_\omega' \|^t \\
\leq \gamma^{tn} \sum_{|\omega| = n} \| \psi_\omega' \|^{1+\Theta}.
\end{equation}
Therefore, $P_\Psi(\Theta + 1 + t) \leq t \log \gamma + P_\Psi(1 + \Theta)$ and hence $\lim_{s \to +\infty} P_\Psi(s) = -\infty$. The proof is complete. \qed

**Definition 2.2.** Following [MU1] we call a hyperbolic system $\Psi$ regular if
\begin{equation}
\exists s(\Psi) \geq \Theta(\Psi) : P_\Psi(s(\Psi)) = 0.
\end{equation}
We denote by $\Xi_X(K, \gamma, \theta)$ the class of regular hyperbolic IFS on $X$.

In view of Lemma 2.1, if the number $s(\Psi)$ exists, then it is unique. Also, if $\#I < \infty$ then $\Psi$ is regular since then $\Theta(\Psi) = 0$ and $P_\Psi(0) = \log(\#I) > 0$. The following lemma shows that $s(\Psi)$ is always an upper bound for the Hausdorff dimension of the limit set. The argument is well-known but we include it for completeness. We write $H^\alpha(A) = \lim_{s \to 0} H^\alpha_s(A)$ for the $\alpha$-dimensional Hausdorff measure of a set $A$.

**Lemma 2.3.** If $\Psi \in \Xi_X(K, \gamma, \theta)$, then $\dim_H(J_\Psi) \leq s(\Psi)$.
Proof. Fix $\epsilon, \delta > 0$ and take $n_1$ so large that $\gamma^{n_1} \leq \delta$. By the definition of $s(\Psi)$ and Lemma 2.1 there exists $n_2 \geq n_1$ and $\eta > 0$ such that for all $n \geq n_2$,
\[
\frac{1}{n} \log \left( \sum_{|\omega|=n} \|\psi'_\omega\|^s(\Psi)+\epsilon \right) \leq -\eta.
\]
Hence, for all $n \geq n_2$,
\[
H^s(\Psi)+\epsilon \delta (J_\Psi) \leq \sum_{|\omega|=n} \diam (\psi_\omega (X))^{s(\Psi)+\epsilon} 
\leq \diam (X)^{s(\Psi)+\epsilon} \sum_{|\omega|=n} \|\psi'_\omega\|^s(\Psi)+\epsilon 
\leq \diam (X)^{s(\Psi)+\epsilon} e^{-\eta n},
\]
and, consequently, $H^s(\Psi)+\epsilon (J_\Psi) = 0$. Thus, $H^s(\Psi)+\epsilon (J_\Psi) = 0$, and letting $\epsilon \downarrow 0$ we conclude that $\dim_H (J_\Psi) \leq s(\Psi)$. \qed

Given an IFS $\Psi = \{\psi_i : i \in I\}$ and $F \subset I$ let $\Psi_F = \{\psi_i : i \in F\}$. Denote $\text{Fin}(I) = \{F \subset I : \# F < \infty\}$. We are going to show that $s(\Psi)$ is the supremum of $s(\Psi_F)$ over $F \in \text{Fin}(I)$. Along the way we obtain an estimate on the speed of convergence in (2.3), which will be useful later.

Lemma 2.4. Let $\Psi \in \Xi_X (K, \gamma, \theta)$. Then

(i) for every $t > 0$ and every $n \geq 1$,
\[
(2.8) \quad P_\Psi (t) \leq \frac{1}{n} \log Z_n (\Psi, t) \leq P_\Psi (t) + \frac{t \log K}{n}.
\]

(ii) $s(\Psi) = \sup \{ s(\Psi_F) : F \in \text{Fin}(I) \}$.

Proof. (i) The left-hand side inequality is immediate from the subadditivity of the sequence $n \mapsto \log Z_n (\Psi, t)$. In order to prove the right-hand side inequality, fix $n \geq 1$ and consider an arbitrary integer $q \geq 1$. Then
\[
\frac{1}{qn} \log Z_{qn} (\Psi, t) = \frac{1}{qn} \log \sum_{\omega \in I^{qn}} \|\psi'_{\omega}\|^t 
\geq \frac{1}{qn} \log \left( K^{-qt} \sum_{\omega \in (I^n)^q} \prod_{i=1}^{q} \|\psi'_{\tau_i}\|^t \right) 
\geq -t \log K + \frac{1}{qn} \log \left( \sum_{\tau \in I^n} \|\psi'_{\tau}\|^t \right)^q 
= -t \log K + \frac{1}{n} \log Z_n (\Psi, t),
\]
where in the second displayed line we used (2.5) and \( \omega = \tau_1 \ldots \tau_q \), with \( \tau_i \in I^n \). Letting \( q \to \infty \) we obtain the right-hand side inequality of (2.8).

(ii) Since for every \( t \geq 0 \) and every \( F \subset I \) we have \( P_{\Psi_F}(t) \leq P_{\Psi}(t) \), the inequality \( s(\Psi) \geq \sup\{s(\Psi_F) : F \in \text{Fin}(I)\} \) is obvious. The opposite inequality will be deduced from (i). Fix an arbitrary \( \Theta(\Psi) < t < s(\Psi) \).

Then \( 0 < P_{\Psi}(t) < \infty \) and there exists \( n \in \mathbb{N} \) so large that \( P_{\Psi}(t) > \frac{2t \log K}{n} \). Fix such an \( n \). Clearly, \( Z_n(\Psi, t) = \sup\{Z_n(\Psi_F) : F \in \text{Fin}(I)\} \), hence we can find \( F \in \text{Fin}(I) \) satisfying

\[
\frac{\log Z_n(\Psi_F, t)}{n} \geq \frac{\log Z_n(\Psi, t)}{n} - \frac{t \log K}{n} \geq P_{\Psi}(t) - \frac{t \log K}{n} > \frac{t \log K}{n}.
\]

But now, applying (2.8) to \( \Psi_F \in \Xi_X(K, \gamma, \theta) \) we obtain

\[
P_{\Psi_F}(t) \geq \frac{\log Z_n(\Psi_F, t)}{n} - \frac{t \log K}{n} > 0
\]

which implies that \( t \leq s(\Psi_F) \). Thus, \( s(\Psi) \leq \sup\{s(\Psi_G) : G \in \text{Fin}(I)\} \), and the proof is complete. \( \square \)

3. Families of hyperbolic IFS.

Let \( X \subset \mathbb{R} \) be a compact interval and \( U \subset \mathbb{R}^d \) an open set. Here we consider families of hyperbolic IFS \( \Psi^t \in \Xi_X(K, \gamma, \theta) \) depending on a parameter \( t \in U \).

By \( J_t \) we denote the limit set of \( \Psi^t \) and by \( \pi_t = \pi_{\Psi^t} : I^\infty \to J_t \) we denote the natural projection introduced in Section 2. We need two conditions concerned with the dependence of the IFS on \( t \).

Distortion Continuity: For any \( \eta > 0 \) there exists \( \delta > 0 \) such that

\[
t_1, t_2 \in U, \|t_1 - t_2\| \leq \delta \implies \forall \omega \in I^*, \ e^{-|\omega| \eta} \leq \frac{\| (t_1 \omega) \|^t}{\| (t_2 \omega) \|^t} \leq e^{|\omega| \eta}.
\]

Transversality Condition: For any \( \omega \) and \( \tau \) in \( I^\infty \) with \( \omega_1 \neq \tau_1 \), there exists a constant \( C_1 = C_1(\omega_1, \tau_1) \) such that

\[
\mathcal{L}_d\{t \in U : |\pi_t(\omega) - \pi_t(\tau)| \leq r\} \leq C_1 r \quad \text{for all } r > 0.
\]

We emphasize that \( C_1 \) depends only on \( \omega_1 \) and \( \tau_1 \). Thus, \( C_1 \) can be assumed independent of \( \omega \) and \( \tau \) if \( I \) is finite.

Now we can state the main result of this section. We write \( s(t) = s(\Psi^t) \).

Theorem 3.1. Suppose that \( \{\Psi^t\}_{t \in U} \) is a family of IFS in \( \Xi_X(K, \gamma, \theta) \) satisfying (3.1) and (3.2). Then the function \( t \mapsto s(t) \) is continuous on \( U \) and

(i) \( \dim_H(J_t) = \min\{s(t), 1\} \) for Lebesgue-a.e. \( t \in U \);

(ii) \( \mathcal{L}_1(J_t) > 0 \) for Lebesgue-a.e. \( t \in U \) such that \( s(t) > 1 \).
The rest of the section is devoted to the proof of this theorem. We begin with two lemmas which are easy consequences of (3.1) and (3.2).

**Lemma 3.2.** Given $\epsilon, a > 0$ define $\eta = \frac{-\epsilon \log \gamma}{4a + \epsilon}$ and take $\delta = \delta(\eta)$ coming from (3.1) ascribed to $\eta$. Then for all $\omega \in I^*$,

$$\|t_0 - t\| < \delta \implies \|\psi^t_\omega\|^{a + \frac{\epsilon}{4}} \leq \|\psi^t_\omega\|^a.$$

**Proof.** By (3.1) we have

$$\|\psi^t_\omega\|^{a + \frac{\epsilon}{4}} \leq e^{\|\omega\| \eta (a + \frac{\epsilon}{4})} \cdot \|\psi^t_\omega\|^{a + \frac{\epsilon}{4}} \leq e^{\|\omega\| \eta (a + \frac{\epsilon}{4})} \cdot \|\psi^t_\omega\|^{a + \frac{\epsilon}{4}} \log \gamma \cdot \|\psi^t_\omega\|^{a + \frac{\epsilon}{4}} = e^{\|\omega\| \eta (a + \frac{\epsilon}{4}) + \frac{\epsilon}{4} \log \gamma} \cdot \|\psi^t_\omega\|^{a + \frac{\epsilon}{4}}.$$

The proof is complete. \(\square\)

**Lemma 3.3.** Suppose that the family $\Psi^t$ satisfies (3.2). Then for every $0 < \alpha < 1$ and for all $\omega, \tau \in I^\infty$ with $\omega_1 \neq \tau_1$, there exists $C_2 = C_2(\alpha, \omega_1, \tau_1) > 0$ such that

$$\int_U \frac{dt}{|\pi_t(\omega) - \pi_t(\tau)|^\alpha} \leq C_2.$$

**Proof.** In view of (3.2), we can estimate as follows:

$$\int_U \frac{dt}{|\pi_t(\omega) - \pi_t(\tau)|^\alpha} = \int_0^\infty \mathcal{L}_d \left\{ t \in U : \frac{1}{|\pi_t(\omega) - \pi_t(\tau)|^\alpha} \geq x \right\} dx$$

$$= \int_0^\infty \mathcal{L}_d \left\{ t \in U : |\pi_t(\omega) - \pi_t(\tau)| \leq r \right\} r^{-\alpha - 1} dr$$

$$= \int_0^{\omega_1} \mathcal{L}_d \left\{ t \in U : |\pi_t(\omega) - \pi_t(\tau)| \leq r \right\} r^{-\alpha - 1} dr +$$

$$\int_0^{\alpha - 1} \mathcal{L}_d \left\{ t \in U : |\pi_t(\omega) - \pi_t(\tau)| \leq r \right\} r^{-\alpha - 1} dr$$

$$\leq C_1(\omega_1, \tau_1)(1 - \alpha)^{-1} |X|^{1 - \alpha} + \mathcal{L}_d(U) \alpha^{-1} |X|^{-\alpha},$$

and the lemma is proved. \(\square\)

The next lemma is proved following the scheme of [SSo, Lemma 4.1(ii)]; it implies the continuity statement in Theorem 3.1.

**Lemma 3.4.** If the family $\Psi^t \in \Xi_X(K, \gamma, 0)$, with $t \in U$, satisfies (3.1), then the function $t \mapsto s(t)$ is continuous on $U$. 

Proof. Fix an arbitrary $\Phi \in \Xi_X(K, \gamma, \theta)$. Then for every $t > \Theta(\Phi)$ and $u \geq 0$ we have

$$P_{\Phi}(t + u) - P_{\Phi}(t) = \lim_{n \to \infty} \frac{1}{n} \left( \log \left( \sum_{|\omega| = n} \| \phi'_\omega \|^t u \right) - \log \left( \sum_{|\omega| = n} \| \phi'_\omega \|^t \right) \right) \leq \lim_{n \to \infty} \frac{1}{n} \left( \log \left( \sum_{|\omega| = n} \| \phi'_\omega \|^t u \right) - \log \left( \sum_{|\omega| = n} \| \phi'_\omega \|^t \right) \right) = u \log \gamma.$$ 

Therefore, for all $t > \Theta(\Phi)$ and $u > \Theta(\Phi) - t$,

$$(3.3) \quad |P_{\Phi}(t + u) - P_{\Phi}(t)| \geq |u| \cdot |\log \gamma|.$$ 

Recall that $P_{\Psi^t_i}(s(t_i)) = 0$ by the definition of $s(t_i) = s(\Psi^t_i)$. Fix $\epsilon > 0$, consider $\delta > 0$ produced by (3.1) with $\eta = \epsilon$, and suppose that $\|t_2 - t_1\| < \delta$.

Then by (2.3) and (3.1),

$$|P_{\Psi^t_1}(s(t_2)) - P_{\Psi^t_2}(s(t_2))| = \lim_{n \to \infty} \frac{1}{n} \left( \log \left( \sum_{|\omega| = n} \| (\psi^t_1)' \|^s(t_2) \right) - \log \left( \sum_{|\omega| = n} \| (\psi^t_2)' \|^s(t_2) \right) \right) \leq \epsilon.$$ 

Therefore, $s(t_2) > \Theta(\Psi^t_1)$ and in view of (3.3) we have

$$|s(t_2) - s(t_1)| \leq \frac{1}{|\log \gamma|} |P_{\Psi^t_1}(s(t_2)) - P_{\Psi^t_1}(s(t_1))| \leq \frac{1}{|\log \gamma|} |P_{\Psi^t_1}(s(t_2))| \leq \frac{s(t_2) \epsilon}{|\log \gamma|},$$

and the desired statement follows. 

Following [SSo], we now prove the main ingredient needed for the proof of Theorem 3.1.

Lemma 3.5. Suppose that the family $\{\Psi^t\}_{t \in U}$ satisfies (3.1) and (3.2). Then

(i) for any $t_0 \in U$ and any $\epsilon > 0$ there exists $\delta > 0$ such that

$$\dim_H(J_t) \geq \min\{s(t_0), 1\} - \epsilon \quad \text{for } \mathcal{L}_d\text{-a.e. } t \in B_\delta(t_0).$$

(ii) Suppose that $s(t_0) > 1 + \epsilon$ for some $0 < \epsilon < 1$. Then there exists $\delta > 0$ such that

$$\mathcal{L}_1(J_t) > 0 \quad \text{for } \mathcal{L}_d\text{-a.e. } t \in B_\delta(t_0).$$
Proof. Let \( s = \min \{ s(t_0), 1 \} \). By Lemma 2.4(ii), there exists a finite subset \( F \) of \( I \) such that \( s(\Psi_F^t) > s(t_0) - \frac{c}{2} \geq s - \frac{c}{2} \). To simplify notation we set

\[
\Psi = \{ \psi_i \}_{i \in F} = \Psi_F^t.
\]

Consider the function \( f : F^\infty \to \mathbb{R} \) defined by \( f(\omega) = \log |\psi_i'| (\pi_\omega (\sigma^n \omega)) \). It follows from (2.2) that \( f \) is Hölder continuous, and (2.1) implies

\[
\sum_{i=0}^{n-1} f(\sigma^i \omega) = \log |\psi_{\omega^n} (\pi_\omega (\sigma^n \omega))| \quad \text{for all } \omega \in F^\infty.
\]

Since \( P_{\Psi}(s(\Psi)) = 0 \), the theory of Gibbs states (see [Bo1], cf. [MU1] for a more general setting) produces a Borel probability shift-invariant measure \( \mu \) on \( F^\infty \) such that for some constant \( C_3 \geq 1 \), all \( \omega \in F^\infty \), and all \( n \geq 1 \),

\[
\mu(\omega^n) \in (C_3^{-1}, C_3) |\psi_{\omega^n} (\pi_\omega (\sigma^n \omega))| s(\Psi).
\]

Here \( [\omega^n] \) is the cylinder set of all sequences starting with \( \omega_1 \ldots \omega_n \). The measure \( \mu \) is called the Gibbs state for the potential \( \omega \mapsto s(\Psi)f(\omega) \). Bounded distortion (2.5) implies that there exists a constant \( C_4 \geq 1 \) such that for all \( \omega \in F^\infty \) and all \( n \geq 1 \),

\[
(3.4) \quad \mu([\omega^n]) \in (C_4^{-1}, C_4) ||\psi'_{\omega^n}|| s(\Psi).
\]

Denote the product measure \( \mu \times \mu \) by \( \mu_2 \). First we prove part (i) of the lemma. By the potential-theoretic characterization of the Hausdorff dimension (see [F2, p. 79]) it is enough to show that

\[
(3.5) \quad R(t) = \int_{F^\infty < F^\infty} \frac{d\mu_2(\omega, \tau)}{\pi_t(\omega) - \pi_t(\tau)}^{s-\epsilon} < \infty
\]

for a.e. \( t \in B_\delta(t_0) \), where \( \pi_t = \pi_{\Psi t} \). Indeed, (3.5) means that the \( (s - \epsilon) \)-energy of the “push-down” measure \( \mu \circ \pi_t^{-1} \), supported on the limit set \( J_{\Psi t} \subset J_t \), is finite.

Following the scheme of Kaufman [K] we prove that

\[
\int_{B_\delta(t_0)} R(t)dt < \infty
\]

where \( \delta = \delta(\eta) \) comes from (3.1) and \( \eta = \frac{-\epsilon \log \gamma}{4(s(\Psi) - \frac{1}{2}) \tau^\epsilon} \). For \( \rho \in F^n \) denote

\[
A_\rho = \{ (\omega, \tau) \in F^\infty \times F^\infty : \omega \wedge \tau = \rho \}
\]

where \( \omega \wedge \tau \) is the largest common initial segment of \( \omega \) and \( \tau \). For \( (\omega, \tau) \in A_\rho \) we have by (2.1), (2.5), and the Mean Value Theorem:

\[
(3.6) \quad |\pi_t(\omega) - \pi_t(\tau)| = |(\psi_{\rho^n}')(\epsilon)| \cdot |\pi_t(\sigma^n \omega) - \pi_t(\sigma^n \tau)|
\]

\[
\geq K^{-1} ||\psi_{n}'|| \cdot |\pi_t(\sigma^n \omega) - \pi_t(\sigma^n \tau)|.
\]
By Lemma 3.2,
\begin{equation}
\|\psi_t^\varepsilon\|^s - \varepsilon \geq \|\psi_t^\varepsilon\|^s(\Psi) - \frac{\varepsilon}{t} \geq \|\psi^\varepsilon\|^s(\Psi) - \frac{\varepsilon}{t} \quad \text{for } t \in B_\delta(t_0)
\end{equation}

where \(\psi^\varepsilon = \psi(t_0)\). Now we can estimate using (3.6) and (3.7):
\[
\int_{B_\delta(t_0)} R(t)dt = \sum_{n \geq 0} \sum_{\rho \in F^n} \int_{A_{\rho}} \left( \int_{B_\delta(t_0)} \frac{dt}{|\pi_t^\varepsilon(\omega) - \pi_t^\varepsilon(\tau)|^s - \varepsilon} \right) d\mu_2(\omega, \tau)
\]
\[
\leq \sum_{n \geq 0} \sum_{\rho \in F^n} \int_{A_{\rho}} \left( \|\psi^\varepsilon\|^s - \frac{\varepsilon}{sT} \int_{B_\delta(t_0)} \frac{dt}{|\pi_t^\varepsilon(\sigma^n\omega) - \pi_t^\varepsilon(\sigma^n\tau)|^s - \varepsilon} \right) d\mu_2(\omega, \tau)
\]
\[
\leq \sum_{n \geq 0} \sum_{\rho \in F^n} \int_{A_{\rho}} \|\psi^\varepsilon\|^s(\Psi) - \frac{\varepsilon}{sT} d\mu_2(\omega, \tau).
\]

In the last inequality we applied Lemma 3.3 which is possible because \((\sigma^n\omega)_1 \neq (\sigma^n\tau)_1\). The multiplicative constant depends on \((\sigma^n\omega)_1\) and \((\sigma^n\tau)_1\), but since the set \(F\) is finite this does not cause a problem. Next, applying (3.4) and (2.4) and denoting by \([\rho]\) the cylinder set of \(\rho \in F^n\) we obtain:
\[
\int_{B_\delta(t_0)} R(t)dt \leq \sum_{n \geq 0} \sum_{\rho \in F^n} \int_{A_{\rho}} \frac{\|\psi^\varepsilon\|^s}{\|\rho\|} d\mu_2(\omega, \tau)
\]
\[
\leq \sum_{n \geq 0} \gamma \frac{n}{4T} \sum_{\rho \in F^n} \frac{\mu_2(A_{\rho})}{\|\rho\|}
\]
\[
\leq \sum_{n \geq 0} \gamma \frac{n}{4T} \sum_{\rho \in F^n} \frac{\|\rho\|^2}{\|\rho\|} = \sum_{n \geq 0} \gamma \frac{n}{4T} < \infty.
\]

This concludes the proof of part (i).

(ii) Let \(\eta = \frac{-\varepsilon \log \gamma}{4 + \varepsilon}\) and determine \(\delta = \delta(\eta)\) from (3.1). In view of Lemma 2.4, there exists a finite set \(F \subset I\) such that \(s(\Psi_{t_0}) \geq 1 + \frac{4}{T}\). We use the same set-up as in the proof of part (i) and let \(\Psi = \Psi_{t_0}\). Recall that \(\mu\) is the Gibbs state for the potential \(\omega \mapsto s(\Psi) \log |\psi_{\omega_1}^\varepsilon(\pi(\sigma \omega))|\) satisfying (3.4). For every \(t \in B_\delta(t_0)\) consider
\[
\nu_t = \mu \circ \pi_t^{-1},
\]
the push-down measure on the limit set \(J_{\Psi_t} \subset J_t\). It is enough to show that \(\nu_t\) is absolutely continuous with respect to the Lebesgue measure \(\mathcal{L}_1\) for a.e. \(t \in B_\delta(t_0)\). We prove that
\[
I = \int_{B_\delta(t_0)} \int_{\mathbb{R}} D(\nu_t, x) d\nu_t dt < \infty
\]
where
\[ D(\nu_t, x) = \liminf_{r \downarrow 0} \frac{\nu_t[x - r, x + r]}{2r} \]
is the lower density of the measure \( \nu_t \) at the point \( x \). This will be sufficient since then for a.e. \( t \in B_\delta(t_0) \) we will have \( D(\nu_t, x) < \infty \) for \( \nu_t \)-a.e. \( x \) and [M, Lemma 2.12] will imply that \( \nu_t \) is absolutely continuous. The argument below follows the scheme of [PS01]. First we apply Fatou’s Lemma to get
\[
\mathcal{I} \leq \liminf_{r \downarrow 0} \int_{B_\delta(t_0)} \int_{\mathbb{R}} \frac{\nu_t[x - r, x + r]}{2r} \, d\nu_t \, dt.
\]
Next we use the definition of \( \nu_t \) to change the variable, write \( \nu_t[x - r, x + r] \) as an integral of the indicator function, and change the variable once again to obtain
\[
\int_{\mathbb{R}} \nu_t[x - r, x + r] \, d\nu_t = \int_{F^\infty \times F^\infty} 1_{\{\omega \in F^\infty : |\pi_t(\omega) - \pi_t(\tau)| \leq r\}} \, d\mu_2(\omega, \tau).
\]
Substituting this into (3.8) and exchanging the order of integration yields
\[
\mathcal{I} \leq \liminf_{r \downarrow 0} (2r)^{-1} \int_{F^\infty \times F^\infty} \mathcal{L}_d \{ t \in B_\delta(t_0) : |\pi_t(\omega) - \pi_t(\tau)| \leq r \} \, d\mu_2(\omega, \tau)
\]
By (3.6), Lemma 3.2, and (3.2), we have for all \( (\omega, \tau) \in A_\rho \):
\[
\mathcal{L}_d \{ t \in B_\delta(t_0) : |\pi_t(\omega) - \pi_t(\tau)| \leq r \} \leq \left\{ t \in B_\delta(t_0) : |\pi_t(\sigma^n \omega) - \pi_t(\sigma^n \tau)| \leq \frac{K r}{\|\psi^r\|} \right\} \leq \left\{ t \in B_\delta(t_0) : |\pi_t(\sigma^n \omega) - \pi_t(\sigma^n \tau)| \leq \frac{K r}{\|\psi^r\|^{1+\xi}} \right\} \leq r\|\psi^r\|^{-1-\xi}.
\]
Here we used again that the constant in (3.2) can be made independent of \( \omega \) and \( \tau \), due to the fact that \( F \) is finite. Now we can estimate the integral \( \mathcal{I} \) as follows:
\[
\mathcal{I} \leq \sum_{n \geq 0} \sum_{\rho \in F^n} \int_{A_\rho} \|\psi^r\|^{-1-\xi} \, d\mu_2(\omega, \tau).
\]
By (3.4) and (2.4),
\[
\|\psi^r\|^{-1-\xi} \propto (\mu[\rho])^{-1+\xi/4} \leq (\mu[\rho])^{-1+\xi/4} \leq \gamma^{\frac{n}{n+1}} (\mu[\rho])^{-1},
\]
since $s(\Psi) > 1 + \frac{\epsilon}{2}$ and $(1 + \frac{\epsilon}{2})/(1 + \frac{\epsilon}{2}) < 1 - \frac{\epsilon}{12}$ for $\epsilon < 1$. Thus,

$$I \leq \sum_{n \geq 0} \gamma_n \sum_{\rho \in F^n} \frac{\mu(A_{\rho})}{\mu(\rho)} \leq \sum_{n \geq 0} \gamma_n \sum_{\rho \in F^n} \mu(\rho) = \sum_{n \geq 0} \gamma_n^{1/2} \leq \infty.$$ 

The proof is complete. □

**Proof of Theorem 3.1.** By Lemma 2.3, the function $s(t) = \min\{s(\Psi^t), 1\}$ is an upper bound for the Hausdorff dimension of the limit set $J_t$. So we just have to show that

$$\dim_H(J_t) \geq s(t)$$

for a.e. $t \in U$. Suppose that this is not the case. Then we can find $\epsilon > 0$ and $t_0$, a density point of those $t$ for which

$$\dim_H(J_t) < s(t) - \epsilon.$$ 

Then there exists $\delta_0 > 0$ such that for each $\delta < \delta_0$,

$$(3.9) \quad \mathcal{L}_d \{t \in B_\delta(t_0) : \dim_H(J_t) < \min\{s(t), 1\} - \epsilon\} > 0.$$ 

However, by the continuity of the function $s(t)$ (see Lemma 3.4), if $\delta$ is small enough then $s(t) < s(t_0) + \frac{\epsilon}{2}$ for all $t \in B_\delta(t_0)$. Thus, for all $\delta$ sufficiently small we obtain from (3.9) that

$$\mathcal{L}_d \{t \in B_\delta(t_0) : \dim_H(J_t) < \min\{s(t_0), 1\} - \frac{\epsilon}{2}\} > 0.$$ 

This contradicts Lemma 3.5(i) and completes the proof of the first part of Theorem 3.1. The second part follows immediately from Lemma 3.5(ii). □

**4. Exceptional parameters.**

In this section, following the scheme of Kaufman [K], we obtain an estimate from above for the local Hausdorff dimension of the set of exceptional parameters in Theorem 3.1(i). As in Section 3, we assume that $\{\Psi^t\}_{t \in U}$ is a family of IFS in $\Xi_X(K, \gamma, \theta)$ satisfying (3.1), but we will need the following stronger transversality condition which will be checked for all the examples that we consider. Denote by $N_r(F)$ the minimal number of balls of radius $r$ needed to cover the set $F \subset \mathbb{R}^d$.

**Strong Transversality Condition:** For all $\omega$ and $\tau$ in $F^\infty$ with $\omega_1 \neq \tau_1$, there exists a constant $C_1 = C_1(\omega_1, \tau_1)$ such that for all $r > 0$,

$$(4.1) \quad N_r \left( \{t \in U : |\pi_t(\omega) - \pi_t(\tau)| \leq r\} \right) \leq C_1 r^{1-d}.$$ 

Of course, the strong transversality condition implies the transversality condition (3.2). In the same way as Lemma 3.3 we can prove the following.
Lemma 4.1. Suppose that the family $Ψ^t$ satisfies the strong transversality condition (4.1). Let $m$ be a Borel probability measure in $\mathbb{R}^d$ such that $m(B_r(x)) \leq Cr^u$ for some $C, u > 0$ and all $x \in \mathbb{R}^d, r > 0$. Then for every $\alpha < u - d + 1$ and for all $\omega, \tau \in \mathcal{A}^\infty$ with $\omega_1 \neq \tau_1$, there exists $C_2 = C_2(\alpha, \omega_1, \tau_1) > 0$ such that

$$\int_U \frac{dm(t)}{|\pi_t(\omega) - \pi_t(\tau)|^\alpha} \leq C_2.$$ 

In the sequel any measure with the properties required in Lemma 4.1 will be called a Frostman measure with exponent $u$. Next we prove the analog of Lemma 3.5(i).

Lemma 4.2. Suppose that the family $\{Ψ^t\}_{t \in U}$ satisfies (3.1) and (4.1). Then for any $t_0 \in U$ and any $\epsilon > 0$ there exists $\delta = \delta(t_0, \epsilon) > 0$ such that if $m$ is a Frostman measure on $B_\delta(t_0)$ with exponent $u$, then

$$\dim_H(J_t) \geq \min\{s(t_0), u - d + 1\} - \epsilon$$

for $m$-a.e. $t \in B_\delta(t_0)$.

Proof. We let $s = \min\{s(t_0), u - d + 1\}$ and then repeat the proof of Lemma 3.5(i) almost word by word. The only change is that now we prove that $\int_{B_\delta(t_0)} R(t)dm(t) < \infty$ using Lemma 4.1 in the place where Lemma 3.3 was used. □

Now we can prove the main result of this section.

Theorem 4.3. Suppose that the $d$-parameter family of IFS $\{Ψ^t\}_{t \in U}$ satisfies (3.1) and (4.1). If $G$ is an arbitrary subset of $U$, then for every $\xi > 0$ we have

$$\dim_H(\{t \in G : \dim_H(J_t) < \min\{\xi, s(t)\}\}) \leq \min\{\xi, \sup_{G} s(t)\} + d - 1.$$ 

Proof. Denote $\kappa := \min\{\xi, \sup_{G} s(t)\} + d - 1$. By the countable stability of the Hausdorff dimension, it is enough to prove that for all $n \in \mathbb{N}$,

$$\dim_H(\left\{ t \in G : \dim_H(J_t) < \min\{\xi, s(t)\} - \frac{1}{n} \right\}) \leq \kappa.$$

Fix $n$ and observe that it suffices to show that for all $t_0$ in $G$ there exists $\delta = \delta(t_0)$ such that

$$\dim_H(\left\{ t \in B_\delta(t_0) : \dim_H(J_t) < \min\{\xi, s(t)\} - \frac{1}{n} \right\}) \leq \kappa$$

(just use that any cover of $G$ contains a countable subcover and again the countable stability of the Hausdorff dimension). To establish our claim, suppose that it is false. Then there exists $t_0$ such that for all $\delta > 0$

$$\dim_H(\left\{ t \in B_\delta(t_0) : \dim_H(J_t) < \min\{\xi, s(t)\} - \frac{1}{n} \right\}) > \kappa.$$
Choose $\delta > 0$ so small that the statement of Lemma 4.2 holds with $\epsilon = \frac{1}{2n}$ and $|s(t) - s(t_0)| < \frac{1}{2n}$ for all $t \in B_\delta(t_0)$ (by the continuity of $s(t)$). Then

$$\left\{ t \in B_\delta(t_0) : \dim_H(J_t) < \min\{\xi, s(t)\} - \frac{1}{n} \right\}$$

$$\subset \left\{ t \in B_\delta(t_0) : \dim_H(J_t) < \min\{\xi, s(t_0)\} - \frac{1}{2n} \right\} =: E,$$

hence $\dim_H(E) > \kappa$. By Frostman’s Lemma (see [M, Th. 8.8]), there is a Frostman measure $m$ on the set $E$ with exponent $u = \kappa$. By Lemma 4.2, for $m$-a.e. $t$ we have

$$\dim_H(J_t) \geq \min\{s(t_0), \kappa - d + 1\} - \frac{1}{2n} = \min\left\{ s(t_0), \min_{G} \left\{ \xi, \sup_{G} s(t) \right\} \right\} - \frac{1}{2n}.$$

This is a contradiction since for all $t \in E$ we have $\dim_H(J_t) < \min\{\xi, s(t_0)\} - \frac{1}{2n}$ and

$$\min\{\xi, s(t_0)\} \leq \min\left\{ s(t_0), \min_{G} \left\{ \xi, \sup_{G} s(t) \right\} \right\}.$$

□

Since the function $t \mapsto s(t)$, $t \in U$, is continuous, as an immediate consequence of Theorem 4.3 we get the following estimate for the local dimension of the exceptional set.

**Corollary 4.4.** For every $t_0 \in U$ we have

$$\lim_{r \to 0} \dim_H(\{t \in B_r(t_0) : \dim_H(J_t) < \min\{\xi, s(t)\}\}) \leq \min\{\xi, s(t_0)\} + d - 1.$$

## 5. Parabolic IFS.

Let $X \subset \mathbb{R}$ be a compact interval. We say that a $C^{1+\theta}$ map $\phi : X \to X$ is **parabolic** if the following requirements are fulfilled:

- there is only one point $v \in X$ such that $\phi(v) = v$;
- $|\phi'(v)| = 1$ and $0 < |\phi'(x)| < 1$ for all $x \in X \setminus \{v\}$.
- There exists $L_1 \geq 1$ and $\beta = \beta(\phi) < \theta/(1 - \theta)$ ($= \infty$ if $\theta = 1$) such that

$$L_1^{-1} \leq \liminf_{x \to v} \frac{|\phi'(x)| - 1}{|x - v|^\beta} \leq \limsup_{x \to v} \frac{|\phi'(x)| - 1}{|x - v|^\beta} \leq L_1.$$

At the beginning of this section we state some useful properties of a single parabolic map. They are very similar to [U, Lemmas 2.1-2.3]. First, integrating the partial sums of the series $\sum_{n=1}^{\infty} |(\phi^n)'(x)|$ we get the following.
Lemma 5.1. For all $x \in \phi(X) \setminus \{v\}$ we have

$$\frac{|x - v|}{|\phi^{-1}(x) - x|} \leq \sum_{n=1}^{\infty} |(\phi^n)'(x)| \leq \frac{\left|\phi(x) - v\right|}{\left|x - \phi(x)\right|}.$$

Sending a sufficiently small neighborhood of $v$ to infinity via the mapping $x \mapsto 1/(x - v),$ one can easily prove the following two local results.

Lemma 5.2. For every neighborhood $V$ of $v$ there exists $L_2(V) \geq 1$ such that for all $x \in X \setminus V$ and all $n \geq 1$,

$$\frac{1}{L_2(V)} \leq |\phi^n(x) - v| \cdot n^{1/\beta} \leq L_2(V).$$

Lemma 5.3. For every neighborhood $V$ of $v$ there exists $L_3(V) \geq 1$ such that for all $x \in X \setminus V$ and all $n \geq 1$

$$\frac{1}{L_3(V)} \leq |(\phi^n)'(x)| \cdot n^{\frac{\beta + 1}{\beta}} \leq L_3(V).$$

Since $\beta < \theta/(1 - \theta),$ the following is immediate from Lemma 5.3.

Corollary 5.4. For every neighborhood $V$ of $v$ there exists $L_4(V) < \infty$ such that

$$\sum_{n=1}^{\infty} \| (\phi^n)' \|_{X \setminus V}^\theta < L_4(V)$$

where $\| \cdot \|_{X \setminus V}$ denotes the sup-norm on $X \setminus V.$

Turning now our attention to iterated function systems we recall that a $C^{1+\theta}$ map $\phi$ is hyperbolic if $0 < |\phi'(x)| < 1$ for all $x \in X.$

Definition 5.5. Let $\Phi = \{\phi_1, \ldots, \phi_k\}$ be a collection of $C^{1+\theta}$ functions on a closed interval $X \subset \mathbb{R}$ such that $\phi_k$ is parabolic with the fixed point $v$ and the other functions are hyperbolic. We write $\Phi \in \Gamma_X(\theta)$ if, in addition,

$$(5.1) \quad \phi_i(X) \subset \text{Int}(X) \setminus \{v\} \quad \text{for all } i \leq k - 1.$$

Remark. We consider IFS with just one parabolic element. The case of more than one parabolic function can also be handled, but at the cost of additional technical complications.

Let $\mathcal{A} = \{1, \ldots, k\}, \mathcal{A}^* := \bigcup_{n \geq 1} \mathcal{A}^n,$ and suppose that

$$\max\{\|\phi_i'\| : i \leq k - 1\} \leq \gamma < 1.$$  

Lemma 5.6. An IFS $\Phi \in \Gamma_X(\theta)$ is a topological IFS.
Proof. All we need to show is that the intersections \( \bigcap_{n \geq 1} \phi_{\omega|_n}(X) \) are singletons for all \( \omega \in \mathcal{A}^\infty \). By (5.2),
\[
\lim_{n \to \infty} \text{diam} (\phi_{\omega|_n}(X)) \to 0
\]
if the sequence \( \omega \) has infinitely many symbols other than \( k \). The remaining possibility is \( \omega = wk^\infty \) for some \( w \in \mathcal{A}^* \) but then (5.3) is still true since \( \bigcap_{n \geq 1} \phi_{kn}(X) = \{v\} \).
\[\square\]

Now, following [MU2, MU3], given a parabolic IFS \( \Phi \in \Gamma_X(\theta) \), consider an associated infinite IFS
\[
\Phi_* = \{\phi^n_k \phi_i : n \geq 0, \ i \leq k - 1\}.
\]
We also write \( \Phi_* = \{\phi^*_b\} \) where \( \phi^*_b = \phi^n_k \phi_i \) and
\[
I = \{b = (n, i) : n \geq 0, \ i \leq k - 1\}.
\]
The following properties of \( \Phi_* \) are immediate from the definitions.

**Corollary 5.7.** Let \( \Phi \in \Gamma_X(\theta) \). Then \( \Phi_* = \{\phi^*_b\} \) satisfies
\( i \) \( \phi^*_b \in C^{1+\theta}(X \to \text{Int}(X)) \) for all \( b \in I \);
\( ii \) \( 0 < \|\phi^*_b\| \leq \gamma < 1 \) for all \( b \in I \);
\( iii \) \( J_\Phi = J_{\Phi_*} \cup \{\phi_w(v)\}_{w \in \mathcal{A}^*} \) so \( \dim_H(J_\Phi) = \dim_H(J_{\Phi_*}) \).

Thus there is an infinite hyperbolic IFS \( \Phi_* \) associated with a finite parabolic IFS \( \Phi \). This idea essentially goes back to Schweiger’s “jump transformation” [Sc]. Our next goal is to show that \( \Phi \in \Gamma_X(\theta) \) implies \( \Phi_* \in \Xi_X(K, \gamma, \theta) \) for some \( K \), see Definition 2.2. To achieve this, two more properties have to be verified: The bounded distortion property (2.5) and regularity (2.7). Bounded distortion properties for parabolic IFS (without overlaps) were investigated in [U]. Here a different version is needed but the approach is similar.

In the next section we study families of parabolic IFS, and it will be very important to know exactly what the various constants depend on. Therefore, we introduce the following notation. Let \( \Phi \in \Gamma_X(\theta) \). We write
\[
\Phi \in \Gamma_X(\theta, V, \gamma, u, M)
\]
if \( V \) is a connected open neighborhood of the parabolic point \( v \) such that
\[
\bigcup_{i=1}^{k-1} \phi_i(X) = \emptyset,
\]
\[
\max\{|\phi'_i(x)|, \ i \leq k - 1\} \leq \gamma \in (0, 1),
\]
\[
\min\{|\phi'_i(x)|, \ x \in X, \ i \leq k\} \geq u \in (0, 1),
\]
and
\[ \| \Phi' \|_\theta := \max_{i \leq k} \sup \{ |\phi'_i(x) - \phi'_i(y)| \cdot |x - y|^{-\theta} \} \leq M. \]

By Definition 5.5, every \( \Phi \in \Gamma_X(\theta) \) belongs to some \( \Gamma_X(\theta, V, \gamma, u, M) \).

The next lemma will also be useful when we consider families of parabolic IFS. Recall that \( \| \cdot \|_{X \setminus V} \) denotes the supremum norm on \( X \setminus V \).

**Lemma 5.8.** There exist constants \( L_5 = L_5(X, \theta, V, \gamma, u, M) > 1 \) and \( L_6 = L_6(\theta, V, \gamma) > 0 \) such that for every \( \Phi \in \Gamma_X(\theta, V, \gamma, u, M) \), all \( \omega \in A^* \), and all \( x, y \in X \setminus V \),
\[ L_5^{-|x-y|^\theta} \leq \frac{|\phi'_\omega(y)|}{|\phi'_\omega(x)|} \leq L_5^{|x-y|^\theta} \]

and
\[ \sum_{j=0}^{|\omega|} \| \phi'_{\sigma^j \omega} \|_{X \setminus V}^\theta \leq L_6. \]

**Proof.** We start with (5.10). Every \( \tau \in A^* \) can be written as
\[ \tau = k^{r_1} j_1 k^{r_2} j_2 \ldots k^{r_l} j_l k^{r_{l+1}} \]
where \( l \geq 0 \) and \( r_p \geq 0, j_p \leq k - 1 \) for \( p \leq l \). When \( l = 0 \), Equation (5.11) becomes \( \tau = k^{r_1} \). One readily estimates using (5.5) and (5.6):
\[ \| \phi'_\tau \|_{X \setminus V} \leq \gamma^l \prod_{p=1}^{l+1} \| \phi'^{r_p}_{k^p} \|_{X \setminus V}^\theta. \]
Applying this inequality to \( \tau = \sigma^j \omega \) and summing over \( j \) we obtain
\[ \sum_{j=0}^{|\omega|} \| \phi'_{\sigma^j \omega} \|_{X \setminus V}^\theta \]
\[ \leq \gamma^{l \theta} \cdot \left( \sum_{j=1}^{r_1} \| (\phi^j_k)' \|_{X \setminus V}^\theta + 1 \right) \cdot \prod_{p=2}^{l+1} \| (\phi'^{r_p}_{k^p})' \|_{X \setminus V}^\theta \]
\[ + \gamma^{(l-1) \theta} \cdot \left( \sum_{j=1}^{r_2} \| (\phi^j_k)' \|_{X \setminus V}^\theta + 1 \right) \cdot \prod_{p=2}^{l+1} \| (\phi'^{r_p}_{k^p})' \|_{X \setminus V}^\theta \]
\[ + \cdots + \gamma^\theta \cdot \left( \sum_{j=1}^{r_{l+1}} \| (\phi^j_k)' \|_{X \setminus V}^\theta + 1 \right) \cdot \| (\phi'^{r_{l+1}}_{k^p})' \|_{X \setminus V}^\theta + \sum_{j=1}^{r_{l+1}} \| (\phi^j_k)' \|_{X \setminus V}^\theta. \]
Applying Corollary 5.4 for \( \phi = \phi_k \) and using that \( \|(\phi'_k)^r\|_{X \setminus V} \leq 1 \) for all \( r \), we obtain
\[
\sum_{j=0}^{\lfloor \omega \rfloor - 1} \|\phi'_{\sigma_j\omega}\|_{X \setminus V}^\theta \leq (L_4(V) + 1) \sum_{i=0}^{\infty} \gamma^i \theta =: L_6 = L_6(\theta, V, \gamma) < \infty,
\]
and (5.10) is proved.

Now we turn to (5.9); clearly, it suffices to prove the right-hand side inequality. Suppose first that \( x \) and \( y \) belong to the same connected component of \( X \setminus V \). Then using the Mean Value Theorem we conclude that for every \( 0 \leq j \leq \lfloor \omega \rfloor \) there exists \( c_j \in X \setminus V \) such that \( |\phi_{\sigma_j\omega}(y) - \phi_{\sigma_j\omega}(x)| = |\phi'_{\sigma_j\omega}(c_j)| \cdot |y - x| \). We have for all \( 1 \leq i \leq \lfloor \omega \rfloor \) by (5.7):
\[
\left| \log |\phi'_\omega(y)| - \log |\phi'_\omega(x)| \right|
= \sum_{j=1}^{\lfloor \omega \rfloor} \log |\phi'_{\omega_j}(\phi_{\sigma_j\omega}(y))| - \sum_{j=1}^{\lfloor \omega \rfloor} \log |\phi'_{\omega_j}(\phi_{\sigma_j\omega}(x))|
\leq \sum_{j=1}^{\lfloor \omega \rfloor} \left| \log |\phi'_{\omega_j}(\phi_{\sigma_j\omega}(y))| - \log |\phi'_{\omega_j}(\phi_{\sigma_j\omega}(x))| \right|
\leq \sum_{j=1}^{\lfloor \omega \rfloor} \frac{\|\Phi'\|_\theta}{u} |\phi'_{\sigma_j\omega}(c_j)| \theta |y - x|\theta
= \sum_{j=1}^{\lfloor \omega \rfloor} \frac{\|\Phi'\|_\theta}{u} L_6 |y - x|\theta \leq \frac{M}{u} L_6 |y - x|\theta.
\]

In the last displayed line we used (5.10) and (5.8). This completes the proof of (5.9) when \( x \) and \( y \) are in the same connected component of \( X \setminus V \). Now suppose that they are in different components. Since then \( |y - x| \geq \text{diam} \, (V) \), it suffices to show the existence of a constant \( L_7 = L_7(X, \theta, V, \gamma, u, M) \geq 1 \) such that \( |\phi'_\omega(y)| \leq L_7 \cdot |\phi'_\omega(x)| \) for all \( \omega \in \mathcal{A}^* \). To this end, suppose \( \omega = \tau k^n \) where \( n \geq 0 \) and \( \tau_k \neq k \). Let \( |\tau| = \tau \). Observe that the points \( \phi_{\tau_k^n}(x) \) and
\( \phi_{n,k^n}(y) \) belong to \( \phi_n(X) \) and hence are in the same connected component of \( X \setminus V \). Thus, in view of Lemma 5.3 and (5.12) we get
\[
\frac{|\phi'_n(y)|}{|\phi'_n(x)|} = \frac{|\phi'_{n,\tau_{l-1}}(\phi_{n,k^n}(y))| \cdot |\phi'_n(\phi_{k^n}(y))| \cdot |\phi'_n(y)|}{|\phi'_{n,\tau_{l-1}}(\phi_{n,k^n}(x))| \cdot |\phi'_n(\phi_{k^n}(x))| \cdot |\phi'_n(x)|}
\]
\[
\leq \exp\left( \frac{M}{u} L_0 |y - x|^\theta \right) \frac{\gamma}{u} L_3(V)^2,
\]
and it remains to note that \( |y - x| \leq \text{diam}(X) \). The only possibility left is \( \omega = k^n \) but then \( |\phi'_n(y)| \leq L_3(V)^2 |\phi'_n(x)| \) by Lemma 5.3. The proof is complete.

**Corollary 5.9.** There exists a constant \( K_1 = K_1(X, \theta, V, \gamma, u, M) > 1 \) such that for every \( \Phi \in \Gamma_X(\theta, V, \gamma, u, M) \), the associated infinite IFS \( \Phi_\ast \) satisfies (2.5) with \( K = K_1 \). More precisely, for all \( \tau \in I^+ \) and all \( x, y \in X \),
\[
K_1^{-|y-x|^\theta} \leq \frac{|(\phi_{\tau^n}^\ast)'(y)|}{|(\phi_{\tau^n}^\ast)'(x)|} \leq K_1^{|y-x|^\theta}.
\]

**Proof.** It is enough to prove only the right-hand side of this formula. By the definition of the system \( \Phi_\ast \) we can write \( \phi_{\tau^n}^\ast = \phi_{\omega^n} \phi_i \), where \( \omega \in \mathcal{A}^* \) and \( i \in \{1, 2, \ldots, k - 1\} \). Then, applying (5.9) and (5.7) we can estimate as follows:
\[
\frac{|(\phi_{\tau^n}^\ast)'(y)|}{|(\phi_{\tau^n}^\ast)'(x)|} = \frac{|(\phi_{\omega^n} \phi_i)'(y)|}{|(\phi_{\omega^n} \phi_i)'(x)|} = \frac{|\phi_{\omega^n}'(\phi_i(y))|}{|\phi_{\omega^n}'(\phi_i(x))|} \cdot \frac{|\phi'_i(y)|}{|\phi'_i(x)|}
\]
\[
\leq L_5 |\phi_{\omega^n}'(\phi_i(x))|^\theta \exp\left( \frac{|\phi'_i(y)|}{|\phi'_i(x)|} - 1 \right)
\]
\[
\leq L_5^{|y-x|^\theta} \exp\left( \frac{\|\Phi\|^\theta}{u} |y - x|^\theta \right)
\]
\[
\leq L_5^{|y-x|^\theta} \exp\left( \frac{M}{u} |y - x|^\theta \right),
\]
and the proof is finished. \( \Box \)

The pressure function \( P_\Phi(t) \) for the parabolic IFS \( \Phi \) is defined by
\[
P_\Phi(t) = \lim_{n \to \infty} \frac{1}{n} \sum_{|\omega|=n} \|\phi_{\tau^n}\|^t,
\]
in accordance with (2.3). Observe that \( P_\Phi(0) = \log k \). In contrast with the hyperbolic case, \( P_\Phi(t) \geq 0 \) for all \( t > 0 \) since \( \|\phi_{k^n}\| = 1 \) for all \( n \). It is easy to see that \( P_\Phi(t) \) is non-increasing and continuous. Thus, \( P_\Phi(s) = 0 \) implies \( P_\Phi(t) = 0 \) for all \( t > s \). Denote
\[
s(\Phi) = \min\{t > 0 : P_\Phi(t) = 0\}.
\]
If the pressure function has no zeros, we let $s(\Phi) = \infty$, however, the next proposition implies that $s(\Phi) < \infty$ for any $\Phi \in \Gamma_X(\theta)$.

**Proposition 5.10.** Let $\Phi \in \Gamma_X(\theta)$ and let $\Phi_* \in \Gamma_X(\theta)$ be the associated infinite hyperbolic IFS. Then

(i) $\Theta(\Phi_*) = \frac{\beta}{\beta + 1}$ and $\Phi_*$ is regular, i.e., there exists a unique $s(\Phi_*) > \Theta(\Phi_*)$ such that $P_{\Phi_*}(s(\Phi_*)) = 0$;

(ii) $s(\Phi_*) = s(\Phi)$.

Before the proof, we point out the following.

**Corollary 5.11.** If $\Phi \in \Gamma_X(\theta)$ then

(i) $\Phi_* \in \Xi_X(K_1, \gamma, \theta)$ where $K_1$ is from Corollary 5.9;

(ii) $\dim_H(J_{\Phi}) \leq s(\Phi)$.

**Proof.** (i) follows from Proposition 5.10(i) and Definition 2.2.

(ii) follows from Proposition 5.10(ii), Lemma 2.3, and Corollary 5.7. $\Box$

**Proof of Proposition 5.10.** (i) First we compute $\Theta(\Phi_*)$, see (2.6) for the definition. It follows from (5.1) and Lemma 5.3 that there exists a constant $K_2 > 1$ such that for all $x \in X$ and $b = (l, i) \in I$,

\begin{equation}
(\phi^*_b)'(x) = |(\phi^*_b)'(x)| \in (K_2^{-1}, K_2) (l + 1)^{-\frac{\beta + 1}{\beta}}.
\end{equation}

Therefore,

\begin{equation}
|(\phi^*_b)'(x)| \geq K_2^{-2} ||(\phi^*_b)'||_t \quad \text{for all } x \in X.
\end{equation}

Let $t > \frac{\beta}{\beta + 1}$. We have by (5.14), writing $\phi^*_r = \phi^*_{r_1} \cdots \phi^*_{r_n}$:

\begin{align*}
Z_n(\Phi_*, t) &= \sum_{|\tau| = n} ||(\phi^*_\tau)'||_t \sum_{|\tau'| = n} ||(\phi^*_\tau)'||_t \cdots ||(\phi^*_\tau)'||_t \\
&= (K_2^{-2tn}, 1) \left( \sum_{b \in I} ||(\phi^*_b)'||_t \right)^n \\
&= (K_2^{-2tn}, 1) \left( \sum_{i \leq k - 1} \sum_{l \geq 0} ||(\phi^*_b)'||_t \right)^n.
\end{align*}

Next we use (5.13) to get

\begin{align*}
Z_n(\Phi_*, t) &= (K_2^{-3tn}, K_2^{3tn}) \left( \sum_{i \leq k - 1} \sum_{l \geq 0} (l + 1)^{-\frac{\beta + 1}{\beta}t} \right)^n \\
&= (K_2^{-3tn}(k - 1)^n, K_2^{3tn}(k - 1)^n) \left( \sum_{l \geq 0} (l + 1)^{-\frac{\beta + 1}{\beta}t} \right)^n.
\end{align*}
Since \( P_{\Phi_*}(t) = \lim_{n \to \infty} \frac{1}{n} \log Z_n(\Phi_*, t) \) we obtain

\[
\log(k - 1) - 3t \log K_2 \leq P_{\Phi_*}(t) - \sum_{l \geq 0} (l + 1) \frac{\beta_{l+1}}{\beta^l} t \leq \log(k - 1) + t \log K_2.
\]

It follows that \( P_{\Phi_*}(t) < \infty \) for \( t > \frac{\beta}{\beta + 1} \) and \( \lim_{t \to \frac{\beta}{\beta + 1}} P_{\Phi_*}(t) = +\infty \). Thus, \( \Theta(\Phi_*) = \frac{\beta}{\beta + 1} \). Since \( P_{\Phi_*}(t) \) is positive and finite for some \( t \), we conclude from Lemma 2.1 that there exists a unique solution \( s(\Phi_*) \) for the Bowen’s equation \( P_{\Phi_*}(t) = 0 \). Part (i) is proved.

(ii) First we observe that

\[
(5.15) \quad s(\Phi_*) = s' := \inf \left\{ t > 0 : \sum_{n \geq 1} Z_n(\Phi_*, t) < \infty \right\}.
\]

Indeed, since \( P_{\Phi_*}(t) \) is strictly decreasing, \( s(\Phi_*) = \inf \{ t > 0 : P_{\Phi_*}(t) < 0 \} \).

If \( P_{\Phi_*}(t) < -\epsilon < 0 \) then \( Z_n(\Phi_*, t) \leq Ce^{-\epsilon n} \) as \( n \to \infty \) and so \( t \geq s' \). Thus, \( s(\Phi_*) \geq s' \). On the other hand, if \( \sum_{n \geq 1} Z_n(\Phi_*, t) < \infty \) then \( Z_n(\Phi_*, t) \to 0 \) hence \( P_{\Phi_*}(t) \leq 0 \). Therefore, \( s(\Phi_*) \leq s' \) and (5.15) is proved.

Next we demonstrate that for all \( t > 0 \),

\[
(5.16) \quad Z_n(\Phi, t) \leq 1 + \sum_{m \leq n} Z_m(\Phi_*, t).
\]

Indeed, every \( \omega \in A^n \) can be written as \( \omega = k^n \) or

\[
\omega = k^{r_1} i_1 k^{r_2} i_2 \ldots k^{r_l} i_l k^{r_{l+1}}
\]

where \( l \geq 1 \) and \( i_p \leq k - 1 \), \( r_p \geq 0 \), for \( p \leq l + 1 \). Thus, either \( \phi_\omega = \phi_k^t \) or \( \phi_\omega = \phi_\tau \phi_{k^{r_{l+1}}}^t \) for some \( \tau \in I^* \). In the latter case, \( \|\phi_\tau^t\| \leq \|\phi_k^t\| \) and \( |\tau| = l \leq n \). Moreover, every map \( \phi_{k^n}^t \) with \( |\tau| \leq n \) occurs in this procedure at most once. The estimate (5.16) follows by noting that \( \|\phi_{k^n}^t\| = 1 \).

Now we can show that \( s(\Phi_*) \geq s(\Phi) = \min \{ t > 0 : P_{\Phi}(t) = 0 \} \). In fact, if \( \sum_{n \geq 1} Z_n(\Phi_*, r) < \infty \) then \( Z_n(\Phi, r) \) is bounded by (5.16) and hence

\[
P_{\Phi}(r) = \lim_{n \to \infty} \frac{1}{n} \log Z_n(\Phi, r) \leq 0.
\]

Therefore, \( r \geq s(\Phi) \) and the desired inequality follows from (5.15).

It remains to verify that \( s(\Phi_*) \leq s(\Phi) \). Recall Lemma 2.4(i) which says that if \( \Psi \in \Xi_X(K, \gamma, \theta) \) then for every \( t > 0 \) and every \( n \geq 1 \),

\[
(5.17) \quad P_{\Phi}(t) \leq \frac{1}{n} \log Z_n(\Psi, t) \leq P_{\Phi}(t) + \frac{t \log K}{n}.
\]

It is enough to prove that if \( r > s(\Phi) \) then \( r \geq s(\Phi_*) \). We have \( P_{\Phi_*}(r) = 0 \).

Fix an arbitrary \( \delta > 0 \). It suffices to show that \( P_{\Phi_*}(r) \leq \delta \). It is convenient to set \( \Psi := \Phi_* \). By the first part of this proposition, \( \Phi_* \in \Xi_X(K_1, \gamma, \theta) \).
Take \( q \in \mathbb{N} \) so large that \( \frac{r \log K_1}{q} < \frac{\delta}{8} \). Then there exists a finite subset \( F \subset I \) such that
\[
0 < \frac{1}{q} \log Z_q(\Psi, r) - \frac{1}{q} \log Z_q(\Psi_F, r) < \frac{\delta}{4}.
\]
Now applying (5.17) to \( \Psi \) and \( \Psi_F \) (which is obviously in \( \Xi_X(K_1, \gamma, \theta) \) as well) we obtain for all \( n \geq q \):
\[
0 < \frac{1}{n} \log Z_n(\Psi, r) - \frac{1}{n} \log Z_n(\Psi_F, r) < \frac{\delta}{4} + \frac{2r \log K_1}{q} < \frac{\delta}{2}.
\] (5.18) 

Recall that \( \Psi_F \) is a finite IFS so its elements (which are of the form \( \phi^t_i \)) have some finite maximal length, say \( L \), over the alphabet \( A \). Therefore, \( Z_n(\Psi_F, r) \leq \sum_{p \leq n} Z_p(\Phi, r) \). Since \( P_\Phi(r) = 0 \), there exists \( m \in \mathbb{N} \) such that \( Z_p(\Phi, r) \leq e^{\frac{r^2}{2}} \) for \( p \geq m \). Then
\[
Z_n(\Psi_F, r) \leq \sum_{p \leq m} Z_p(\Phi, r) + \sum_{p=m+1}^{nL} e^{\frac{r^2}{2}} \leq C(e^{\frac{r^2}{2}}).
\]
This implies that \( \lim_{n \to \infty} \frac{1}{n} \log Z_n(\Psi_F, r) \leq \frac{\delta}{2} \), and together with (5.18) this yields
\[
P_\Psi(r) = \lim_{n \to \infty} \frac{1}{n} \log Z_n(\Psi, r) \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta,
\]
as desired. The proof is complete. \( \square \)

6. Families of parabolic IFS.

Let \( U \subset \mathbb{R}^d \) be an open set. Here we consider families of parabolic IFS
\[
\Phi^t = \{ \phi^t_1, \ldots, \phi^t_{k-1}, \phi^t_k \}
\]
depending on \( t \in \overline{U} \), the closure of \( U \). We assume that \( \Phi^t \in \Gamma_X(\theta) \) for all \( t \in \overline{U} \) (see Definition 5.5). Although the parabolic function does not depend on the parameter, it is sometimes convenient to write \( \phi^t_k \equiv \phi_k \) for \( t \in U \). Let \( \pi_t : A^\infty \to \mathbb{R} \) be the natural projection associated with \( \Phi^t \) and denote \( J_t = \pi_t(A^\infty) \). Two conditions which control the dependence on \( t \) will be needed.

**Continuity**: The maps
\[
(6.1) \quad t \mapsto \phi^t_i \text{ are continuous from } \overline{U} \text{ to } C^{1+\theta}(X) \text{ for } i \leq k-1.
\]

**Transversality Condition**: There exists a constant \( C'_1 \) such that for all \( \omega \) and \( \tau \) in \( A^\infty \) with \( \omega_1 \neq \tau_1 \),
\[
(6.2) \quad \mathcal{L}_d \{ t \in U : |\pi_t(\omega) - \pi_t(\tau)| \leq r \} \leq C'_1 r \quad \text{for all } r > 0.
\]
This condition is almost identical to the transversality condition (3.2) except that here \( \mathcal{A} \) is finite so \( C'_1 \) is an absolute constant. Let
\[
s(t) = s(\Phi^t) = \min\{u > 0 : P_{\Phi^t}(u) = 0\}.
\]
The following theorem is the main result of the paper.

**Theorem 6.1.** Suppose that \( \{\Phi^t\}_{t \in U} \) is a family of parabolic IFS in \( \Gamma_X(\theta) \) satisfying (6.1) and (6.2). Then the function \( t \mapsto s(t) \) is continuous on \( U \) and
\[
(\text{i}) \quad \dim_H(J_t) = \min\{s(t), 1\} \text{ for Lebesgue-a.e. } t \in U;
\]
\[
(\text{ii}) \quad \mathcal{L}_1(J_t) > 0 \text{ for Lebesgue-a.e. } t \in U \text{ such that } s(t) > 1.
\]

**Proof.** The plan is to apply Theorem 3.1 to the family of associated infinite hyperbolic IFS
\[
\Phi^t_n = \{\phi_k^n \phi^t_i : n \geq 0, \ i \leq k - 1\}.
\]
Since the family \( \{\Phi^t\}_{t \in U} \) satisfies the continuity condition (6.1), we can find \( V, \gamma, u, M \), so that
\[
(6.3) \quad \Phi^t \in \Gamma_X(\theta, V, \gamma, u, M) \quad \text{for all } t \in U,
\]
see (5.4) for the meaning of this notation. These numbers and the neighborhood \( V \) of \( v \) will be fixed for the rest of the proof.

Observe that the constant \( K_1 \) in Corollary 5.9 is independent of \( t \in U \) and by Corollary 5.11(i) we have \( \Phi^t_n \in \Xi_X(K_1, \gamma, \theta) \) for \( t \in U \). Notice also that \( s(t) = s(\Phi^t_n) \) by Proposition 5.10(ii). Thus, to prove the theorem it remains to verify the distortion continuity property (3.1) and transversality condition (3.2) for \( \Phi^t_n \). We begin with the latter since it is easier.

As in Section 5, the alphabet for the IFS \( \Phi^t_n \) will be \( I = \{n, i\} : n \geq 0, i \leq k - 1\}. \) Let \( \zeta \) and \( \xi \) be elements of \( I^\infty \) with \( \zeta_1 \neq \xi_1 \). To distinguish between the IFS \( \Phi^t \) and \( \Phi^t_n \) we denote by \( \pi^* : I^\infty \to \mathbb{R} \) the natural projection corresponding to the IFS \( \Phi^t_n \). Let \( \zeta_1 = (n, i) \) and \( \xi_1 = (m, j) \). Clearly,
\[
\pi^*_t(\zeta) = \pi_t(\omega), \quad \pi^*_t(\xi) = \pi_t(\tau)
\]
where \( \omega \in \mathcal{A}^\infty \) begins with \( k^ni \) and \( \tau \in \mathcal{A}^\infty \) begins with \( k^mj \). Assume without loss of generality that \( n \leq m \). Then
\[
\pi^*_t(\zeta) - \pi^*_t(\xi) = \pi_t(\omega) - \pi_t(\tau) = \phi^*_n(\pi_t(\sigma^n\omega)) - \phi^*_k(\pi_t(\sigma^n\tau))
\]
and \( (\sigma^n\omega)_1 \neq (\sigma^n\tau)_1 \). Thus, by the Mean Value Theorem and (5.7), for some \( c \in X \),
\[
|\pi^*_t(\zeta) - \pi^*_t(\xi)| = |(\phi^*_n)'(c)| \cdot |\pi_t(\sigma^n\omega) - \pi_t(\sigma^n\tau)| \geq u^n|\pi_t(\sigma^n\omega) - \pi_t(\sigma^n\tau)|.
\]
Therefore, by (6.2),
\[
\mathcal{L}_d\{t \in U : |\pi^*_t(\zeta) - \pi^*_t(\xi)| \leq r\}
\]
\[
\leq \mathcal{L}_d\{t \in U : |\pi_t(\sigma^n\omega) - \pi_t(\sigma^n\tau)| \leq ru^{-n}\}
\]
\[
\leq C'r u^{-n}
\]
which implies (3.2) with \( C(\xi_1, \eta_1) := C_1' u^{-n}. \)

The property (3.1) for \( \Phi_*^t \) is immediate from (6.1) and the following lemma where we let
\[
\|\Phi - \Psi\| = \max_{i \leq k-1} \|\phi_i - \psi_i\| \quad \text{and} \quad \|\Phi' - \Psi'\| = \max_{i \leq k-1} \|\phi'_i - \psi'_i\|.
\]

**Lemma 6.2.** There exists a positive constant \( C_2' = C_2'(X, \theta, V, \gamma, u, M) \) such that for any \( \Phi = \{\phi_1, \ldots, \phi_k\} \) and \( \Psi = \{\psi_1, \ldots, \psi_k\} \), two parabolic IFS in \( \Gamma_X(\theta, V, \gamma, u, M) \) with \( \phi_k = \psi_k \), the associated infinite hyperbolic IFS \( \Phi_* \) and \( \Psi_* \) have the following property. For any \( \tau \in I^* \) and any \( x \in X \),
\[
\left| \left( \frac{\phi^*_{\sigma_\tau}(x)}{\psi^*_{\sigma_\tau}(x)} \right) \right| \leq \exp \left( C_2' |\tau| \left( \|\Phi - \Psi\|^2 + \|\Phi' - \Psi'\| \right) \right).
\]

**Proof.** Let \( n = |\tau| \) and observe that for all \( x \in X \) and \( 0 \leq j < n \):
\[
\left| \phi^*_{\sigma_\tau}(x) - \psi^*_{\sigma_\tau}(x) \right| \\
\leq \left| \phi^*_{\sigma_\tau}(\tau_{n-1}) \phi^*_{\tau_{n-1}}(x) \right| - \left| \phi^*_{\sigma_\tau}(\tau_{n-1}) \psi^*_{\tau_{n-1}}(x) \right| \\
+ \left| \phi^*_{\sigma_\tau}(\tau_{n-1}) \psi^*_{\tau_{n-1}}(x) \right| - \left| \phi^*_{\sigma_\tau}(\tau_{n-1}) \psi^*_{\tau_{n-1}}(x) \right| \\
\leq \gamma^{n-j-1} \|\Phi_* - \Psi_*\| + \|\phi^*_{\sigma_\tau}(\tau_{n-1}) - \psi^*_{\sigma_\tau}(\tau_{n-1})\|,
\]
for \( x' = \psi^*_{\tau_n}(x) \). Proceeding inductively we obtain
\[
(6.4) \quad \left| \phi^*_{\sigma_\tau}(x) - \psi^*_{\sigma_\tau}(x) \right| \leq \sum_{i=1}^{n-j} \gamma^{n-i-1} \|\Phi_* - \Psi_*\| < \frac{\|\Phi_* - \Psi_*\|}{1 - \gamma}.
\]

Observe that
\[
(6.5) \quad \|\Phi_* - \Psi_*\| = \max_{i \leq k-1, \ m \geq 0} \|\phi^m_k - \phi^m_k \psi_i\| = \max_{i \leq k-1} \|\phi_i - \psi_i\| = \|\Phi - \Psi\|
\]
since \( \|\phi^m_k\| = 1 \). Now let \( y = \phi^*_{\sigma_\tau}(x) \) and \( z = \psi^*_{\sigma_\tau}(x) \), and suppose that \( \tau_j = (l, i) \). Then \( \phi^*_j = \phi^*_j \phi_i \) and \( \psi^*_j = \phi^*_j \psi_i \), and we can estimate by (5.9):
\[
(6.6) \quad \log \left( \frac{\psi^*_{\sigma_\tau}(x)}{\phi^*_{\sigma_\tau}(x)} \right) = \log \left( \frac{\phi^*_j \psi_i}{\phi^*_j \psi_i} \right) = \log \left( \frac{\phi^*_j \psi_i}{\phi^*_j \psi_i} \right) + \log \left( \frac{\phi^*_j \psi_i}{\phi^*_j \psi_i} \right) \\
\leq \log L_5 \cdot |z - \phi_i(y)|^\theta + u^{-1} \psi_i(y) - \phi_i(y).
\]

Next,
\[
|\psi_i(z) - \phi_i(y)| \leq |\psi_i(z) - \psi_i(y)| + |\psi_i(y) - \phi_i(y)| \\
\leq |z - y| + \|\Phi - \Psi\| \leq \|\Phi - \Psi\| \frac{2 - \gamma}{1 - \gamma}
\]
by (6.4) and (6.5). Furthermore,
\[ |\psi'_i(z) - \phi'_i(y)| \leq |\psi'_i(z) - \psi'_i(y)| + |\psi'_i(y) - \phi'_i(y)| \leq \|\Psi\| |z - y|^\theta + \|\Phi' - \Psi'\| \leq M\|\Phi - \Psi\|^\theta (1 - \gamma)^{-\theta} + \|\Phi' - \Psi'\|. \]

Combining the last two estimates with (6.6) yields
\[ \log \left| \frac{(\psi^*_{\tau_j})'(\psi^*_{\sigma_{\lambda}}(x))}{(\phi^*_{\tau_j})'(\phi^*_{\sigma_{\lambda}}(x))} \right| \leq \|\Phi - \Psi\|^\theta \left( \frac{2 - \gamma}{1 - \gamma} \right)^\theta \log L_5 + u^{-1} \left( M\|\Phi - \Psi\|^\theta (1 - \gamma)^{-\theta} + \|\Phi' - \Psi'\| \right) \leq C'_2 \left( \|\Phi - \Psi\|^\theta + \|\Phi' - \Psi'\| \right) \]
for some constant $C'_2 = C'_2(\theta, V, \gamma, u, M)$. Exchanging the roles of $\Phi$ and $\Psi$ we can see that
\[ \left| \log \left( \frac{(\psi^*_{\tau_j})'(\psi^*_{\sigma_{\lambda}}(x))}{(\phi^*_{\tau_j})'(\phi^*_{\sigma_{\lambda}}(x))} \right) \right| \leq C'_2 \left( \|\Phi - \Psi\|^\theta + \|\Phi' - \Psi'\| \right). \]

Finally,
\[ \left| \log \left( \frac{(\psi^*_{i})'(x)}{(\phi^*_{i})'(x)} \right) \right| = \sum_{j=1}^n \log \left( \frac{(\psi^*_{\tau_j})'(\psi^*_{\sigma_{\lambda}}(x))}{(\phi^*_{\tau_j})'(\phi^*_{\sigma_{\lambda}}(x))} \right) \leq \sum_{j=1}^n C'_2 \left( \|\Phi - \Psi\|^\theta + \|\Phi' - \Psi'\| \right) = nC'_2 \left( \|\Phi - \Psi\|^\theta + \|\Phi' - \Psi'\| \right). \]

The lemma is proved, and this also concludes the proof of Theorem 6.1. \(\square\)

Given $\omega \in \mathcal{A}^*$ let $h(\omega)$ denote the number of hyperbolic letters (i.e., $\neq k$) appearing in $\omega$. We record the following useful corollary for future reference although it is not needed in this paper.

**Corollary 6.3.** There exists a positive constant $C'_2 = C'_2(\theta, V, \gamma, u, M)$ such that for any $\Phi = \{\phi_1, \ldots, \phi_k\}$ and $\Psi = \{\psi_1, \ldots, \psi_k\}$, two parabolic IFS in $\Gamma_X(\theta, V, \gamma, u, M)$ with $\phi_k = \psi_k$, for all $\omega \in \mathcal{A}^*$ and all $x \in X$,
\[ \left| \frac{\phi'_i(x)}{\psi'_i(x)} \right| \leq \exp \left( C'_2 h(\omega) \left( \|\Phi - \Psi\|^\theta + \|\Phi' - \Psi'\| \right) \right). \]
Proof. For any $\omega \in A^*$ we can find $\tau \in I^*$ and $l \geq 0$ such that
\[
\phi_\omega = \phi_\tau^* \phi_k^l \quad \text{and} \quad \psi_\omega = \psi_\tau^* \phi_k^l.
\]
By Lemma 6.2 we have
\[
|\phi'_\omega(x)| = |(\phi_\tau^*)'((\phi_k^l(x))'(x))| = |(\phi_k^l)'(x)| \leq \exp\left(C'_2|\tau|\left(\|\Phi - \Psi\|^{\theta} + \|\Phi' - \Psi'\|\right)\right)
\]
which finishes the proof since $|\tau| = h(\omega)$. \hfill \square

We conclude this section with some finer results concerning exceptional parameters in Theorem 6.1(i). These will turn out to be almost immediate consequences of the results obtained in Section 4. The strong transversality condition is formulated in the context of parabolic systems in the same manner as the strong transversality condition in Section 4. Recall that $N_r(F)$ is the minimal number of balls of radius $r$ needed to cover the set $F \subset \mathbb{R}^d$.

**Strong Transversality Condition:** There exists a constant $C'_1$ such that for all $\omega$ and $\tau$ in $A^\infty$ with $\omega_1 \neq \tau_1$,
\[
(6.7) \quad N_r\left(\{t \in U : |\pi_t(\omega) - \pi_t(\tau)| \leq r\}\right) \leq C'_1 r \quad \text{for all} \quad r > 0.
\]

The same arguments as used in the proof of Theorem 6.1 demonstrate that the strong transversality condition (6.7) for a parabolic IFS implies the strong transversality condition (4.1) for the associated hyperbolic system. Consequently, Theorem 4.3 and Corollary 4.4 respectively imply the following.

**Theorem 6.4.** Suppose that $\{\Phi_t\}_{t \in U^d}$ is a $d$-parameter family of parabolic IFS in $\Gamma_X(\theta)$ satisfying (6.1) and (6.7). If $G$ is an arbitrary subset of $U$, then for every $\xi > 0$ we have
\[
\dim_H\left(\{t \in G : \dim_H(J_t) < \min\{\xi, s(t)\}\}\right) \leq \min\left\{\xi, \sup_{t \in G} s(t)\right\} + d - 1.
\]

**Corollary 6.5.** Suppose that $\{\Phi_t\}_{t \in U^d}$ is a $d$-parameter family of parabolic IFS in $\Gamma_X(\theta)$ satisfying (6.1) and (6.7). Then for every $t_0 \in U$ we have
\[
\lim_{r \to 0} \dim_H\left(\{t \in B(t_0, r) : \dim_H(J_t) < \min\{\xi, s(t)\}\}\right) \leq \min\{\xi, s(t_0)\} + d - 1.
\]
7. Examples.

7.1. General classes of examples. We are going to apply Theorem 6.1 to two more specific types of IFS families. In each case we need to impose some bounds on derivatives to guarantee transversality (even strong transversality), which is the most difficult condition to check. Recall that $\| \cdot \|_Y$ denotes the supremum norm on $Y$.

**Proposition 7.1.** Let $\Phi = \{\phi_1, \ldots, \phi_k\} \in \Gamma_X(\theta)$ and $1 \leq d \leq k - 1$. Assume that
\begin{align}
(7.1) \quad & \phi_i \text{ are increasing for all } i \leq k; \\
(7.2) \quad & \phi_i(X) \cap \phi_j(X) = \emptyset \text{ for all } d < i < j \leq k \\
(7.3) \quad & \|\phi'_i\| + \|\phi'_j\|_{\phi_j^{-1}\phi_i(X)} < 1 \text{ for all } i < j \text{ such that } \phi_i(X) \cap \phi_j(X) \neq \emptyset.
\end{align}

Consider the family
\[ \Phi^t = \{\phi_1(x) + t_1, \ldots, \phi_d(x) + t_d, \phi_{d+1}(x), \ldots, \phi_k(x)\} \]
where $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$. Then there exists $\eta > 0$ such that $\{\Phi^t : t \in B_\eta(0)\}$ satisfies all the hypotheses of Theorem 6.1. Moreover, the strong transversality condition (6.7) is satisfied.

**Remark.** It is not hard to find specific IFS satisfying the above proposition. Notice that non-trivial examples start with $k = 3$. Indeed, if $k = 2$ there is a dichotomy: the limit set is either connected or the projection map $\pi_\Phi : A^\infty \to J_\Phi$ is one-to-one. In the latter case there are no “overlaps”, so the dimension formula holds by [U]. In the former case $J_\Phi$ is an interval, so there is no question about its dimension or measure. To get a non-trivial example with $k = 3$, one can take a parabolic function, say, $\phi_3(x) = \sin x$ on $[0, b]$, with $b < \pi/2$, and add two more functions so that there is a “gap” and an “overlap”. For instance, an increasing function $\phi_2$ may be chosen to satisfy $\phi_2(0) > \sin b$ and $\phi_2(b) = b$, and the increasing function $\phi_1$ to satisfy $\phi_1(0) \in (0, \sin b)$, $\phi_1(b) < \phi_2(0)$. Consider a one-parameter family $\Phi^t = \{\phi_1(x) + t, \phi_2(x), \phi_3(x)\}$ and call $t$ admissible if these properties persist for $\Phi^t$. If $t$ is admissible, then the limit set has the convex hull equal to $[0, b]$. We can define $X = [0, b + \epsilon]$ for a small $\epsilon > 0$, so that $\phi_2(X) \subset \text{Int}(X)$. All the assumptions of Proposition 7.1 are satisfied, provided that the derivative of $\phi_1$ is sufficiently small, so the results of Section 6 apply.

**Example 1.** The family of IFS
\[ \Phi^t = \{0.01(x + 1)^2 + t, 0.1x + (4.5/11)\pi, \sin x\} \text{ on } [0, 5\pi/11 + \epsilon] \]
satisfies all the conditions above for \( t \in [0.3, 0.97] \) if \( \epsilon > \) is sufficiently small.

(We are grateful to P. Hanus for correcting our original faulty example.)

In the next proposition all elements of the IFS are assumed to be of the form \( \phi(x+a_j) \) for a single function \( \phi \). Although this seems much more special than Proposition 7.1, it covers some interesting families, in particular, those which arise from a class of continued fractions.

**Proposition 7.2.** (i) Let \( \Phi = \{ \phi_1, \ldots, \phi_k \} \in \Gamma_X(\theta) \) and \( 1 \leq d \leq k-1 \) be such that (7.2) is satisfied. We further assume that there exists a single increasing function \( \phi \in C^{1+\theta} \) on some interval \( Y \) and \( a_i \in \mathbb{R}, u > 0 \) such that

\[
\phi_i(x) = \phi(x + a_i), \quad i \leq k,
\]

and

\[
\inf_{x \in Y} |\phi'(x)| \geq u > 0.
\]

Let

\[
\Phi_t = \{ \phi_1(x + t_1), \ldots, \phi_d(x + t_d), \phi_{d+1}(x), \ldots, \phi_k(x) \}
\]

where \( t = (t_1, \ldots, t_d) \in \mathbb{R}^d \).

Denote by \( \pi_t \) the projection map corresponding to \( \Phi_t \). If there exists \( \delta > 0 \) such that

\[
\frac{\partial}{\partial t_i} \pi_t(\xi)|_{t=0} < 1 - \delta \quad \text{for all} \quad \xi \in \mathcal{A}^\infty \quad \text{and all} \quad 1 \leq i \leq d,
\]

then there exists \( \eta > 0 \) such that \( \{ \Phi_t : t \in B_\eta(0) \} \) satisfies all the hypotheses of Theorem 6.1. Moreover, the strong transversality condition (6.7) is satisfied.

(ii) Condition (7.6) holds if

\[
\|\phi'_i\| < 1/2 \quad \text{for all} \quad 1 \leq i \leq d.
\]

**Example 2.** Let \( \Phi^a = \{ \ln(x + a), \ln(x + e^2 - 2), \ln(x + 1) \} \). This is a parabolic IFS on \([0,2]\). We have \( \|\phi'_1\| = 1/a \), and so Proposition 7.2 and Theorem 6.1 imply that the dimension formula holds for a.e. \( a \in (2, e^2 - 2) \).

(We need to enlarge the interval slightly and let \( X = [0, 2 + \epsilon] \) so that \( \phi_2(X) \subset \text{Int}(X) \).)

Turning to the proofs of Propositions 7.1 and 7.2, note that checking strong transversality (6.7) is the only issue since all other properties obviously hold for sufficiently small perturbations of \( \Phi \). This will be done with the help of the following elementary lemma.
Lemma 7.3. Let $U \subset \mathbb{R}^d$ be an open, bounded set with smooth boundary. Suppose that $f$ is a $C^1$ real-valued function in a neighborhood of $\overline{U}$ such that for some $i \in \{1, \ldots, d\}$ there exists $\eta > 0$ satisfying

\begin{equation}
(7.8) \quad t \in U, \quad |f(t)| \leq \eta \implies \frac{\partial f(t)}{\partial t_i} \geq \eta.
\end{equation}

Then there exists $C = C(\eta)$ such that

\begin{equation}
(7.9) \quad N_r(\{t \in U : |f(t)| \leq r\}) \leq Cr, \quad \forall r > 0.
\end{equation}

Proof. Recall that $N_r(F)$ denotes the minimal number of balls of radius $r$ needed to cover $F$. Since $U$ is bounded it suffices to prove our lemma for every $r < \eta$. Let $M = f^{-1}(0)$. By (7.8), $\operatorname{grad}(f)(t) \neq 0$ for every $t \in M$. Hence $M$ is a $(d - 1)$-dimensional $C^1$-manifold and $(M \cap \partial U) \cup \partial U$ is contained in a union of finitely many compact connected $C^1$ manifolds with smooth boundaries. Thus there exists a constant $C_1$ such that for every $r < \eta$ the set $(M \cap \overline{U}) \cup \partial U$ can be covered by $C_1r^{1-d}$ balls with radii $r$. In order to complete the proof it is sufficient to show that for each point $t = (t_1, t_2, \ldots, t_d) \in U$ with $|f(t)| \leq r$, the distance between $t$ and $(M \cap \overline{U}) \cup \partial U$ does not exceed $r/\eta$. Indeed, we will then cover the set $f^{-1}([-r, r])$ by at most $C_1r^{1-d}$ balls with radii $(1 + 1/\eta)r$. But each ball of radius $(1 + 1/\eta)r$ can be covered by $C_2(\eta)$ balls of radii $r$, where $C_2(\eta)$ depends only on $\eta$ and the dimension $d$ and we would therefore be done.

Without loss of generality let $i = 1$. Consider the function $g(s) = f(s, t_2, \ldots, t_d)$ defined in a neighborhood of $t_1$. Suppose first that $g(t_1) \in [-r, 0]$. By (7.8), the function $g(s)$ increases for $s > t_1$ until either the point $(s, t_2, \ldots, t_d)$ reaches the boundary of $U$, say at a point $(u, t_2, \ldots, t_d) \in \partial U$ and $g(s) \in [-r, 0]$ for all $s \in [t_1, u]$, or $g(s)$ will take on the value zero earlier. Suppose that the first case materializes. Then $g'(s) \geq \eta$ for every $s \in [t_1, u]$ by (7.8). By the Mean Value Theorem, $|u - t_1| \leq r/\eta$, and therefore, $\operatorname{dist}(t, \partial U) \leq \operatorname{dist}(t, (u, t_2, \ldots, t_d)) = |u - t_1| \leq r/\eta$. So, we are done in this case. If the second case holds, let $w$ be that point for which $g(w) = 0$ and $g(s) \in [-r, 0]$ for every $s \in [t_1, w]$. Then $(w, t_2, \ldots, t_d) \in M$ and $g'(s) \geq \eta$ for every $s \in [t_1, w]$. Again by the Mean Value Theorem, $|w - t_1| \leq r/\eta$. Then $\operatorname{dist}(t, M \cap \overline{U}) \leq \operatorname{dist}(t, (w, t_2, \ldots, t_d)) = |w - t_1| \leq r/\eta$ and we are done in this case as well. If $g(t_1) \in [0, r]$, we proceed similarly letting $s$ go left from $t_1$. The proof is complete. \hfill \Box

Proof of Proposition 7.1. It is convenient to write $\Phi^t = \{\phi^t_i\}_{i \leq k}$. The properties (7.3) and (5.1) persist under a small perturbation, so we can find $\eta \in (0, 1)$ such that

\begin{equation}
(7.10) \quad \inf \{\operatorname{dist}(\phi^t_i(X), v) : i \leq k - 1, \quad \|t\| \leq \eta\} \geq \eta > 0;
\end{equation}
\( \inf \{ \text{dist} (\phi_i^t(X), \phi_j^t(X)) : \phi_i(X) \cap \phi_j(X) = \emptyset, ||t|| \leq \eta \} \geq \eta > 0; \)

\( \text{dist} (\phi_i(X), \phi_j(x)) \leq 3\eta \implies ||\phi_i'|| + |\phi_j'(x)| < 1 - \eta \quad \text{for } i \leq d, i < j. \)

The continuity property (6.1) is obvious so we only need to verify the strong transversality condition (6.7).

Consider \( \omega, \tau \in A^\infty \), with \( \omega_1 \neq \tau_1 \). Let \( i = \omega_1 \) and \( j = \tau_1 \). Without loss of generality assume that \( i < j \). If \( \phi_i(X) \cap \phi_j(X) = \emptyset \) then

\[ |\pi_t^i(\omega) - \pi_t^j(\tau)| = |\phi_i^t(\pi_t^i(\sigma \omega)) - \phi_j^t(\pi_t^j(\sigma \tau))| \geq \eta > 0 \]

for \( t \in B_\eta(0) \) by (7.11) and

\( \{ t \in B_\eta(0) : |\pi_t^i(\omega) - \pi_t^j(\tau)| \leq \eta \} \subset B_\eta(0) \)

(note that the left-hand side is empty for \( r < \eta \)). Thus (6.7) certainly holds in this case. If \( \phi_i(X) \cap \phi_j(X) \neq \emptyset \) then \( i \leq d \) and

\[ |\pi_t^i(\omega) - \pi_t^j(\tau)| = t_i + \phi_i(\pi_t^i(\sigma \omega)) - \phi_j^t(\pi_t^j(\sigma \tau)). \]

We are going to use Lemma 7.3 with \( f(t) = \pi_t^i(\omega) - \pi_t^j(\tau) \) and \( U = B_\eta(0) \), so we need to check (7.8). Since \( |\phi_i^t(\pi_t^i(\sigma \omega)) - \phi_j^t(\pi_t^j(\sigma \tau))| = |t_j| \leq \eta \) and \( |t_i| \leq \eta \), we have the implication

\[ |f(t)| = |\pi_t^i(\omega) - \pi_t^j(\tau)| \leq \eta \implies |\phi_i(\pi_t^i(\sigma \omega)) - \phi_j(\pi_t^j(\sigma \tau))| \leq 3\eta \]

\[ \implies ||\phi_i'|| + |\phi_j'(\pi_t^j(\sigma \tau))| < 1 - \eta \]

by (7.12). Since \( j \neq i \) we have

\[ \frac{\partial}{\partial t_i} (\pi_t^i(\omega) - \pi_t^j(\tau)) = 1 + \frac{\partial}{\partial t_i} \phi_i(\pi_t^i(\sigma \omega)) - \frac{\partial}{\partial t_i} \phi_j(\pi_t^j(\sigma \tau)) \]

\[ = 1 + \phi_i'(\pi_t^i(\sigma \omega)) \cdot \frac{\partial}{\partial t_i} \pi_t^i(\sigma \omega) - \phi_j'(\pi_t^j(\sigma \tau)) \cdot \frac{\partial}{\partial t_i} \pi_t^j(\sigma \tau). \]

Observe that \( \phi_i' > 0 \) and \( \frac{\partial}{\partial t_i} \pi_t^i(\sigma \omega) \geq 0 \) by the assumption (7.1). Suppose that \( (\sigma \tau)_n \) is the first symbol \( i \) in \( \sigma \tau \) (if \( \sigma \tau \) contains no \( i \) we have \( \frac{\partial}{\partial t_i} \pi_t^i(\sigma \tau) = 0 \) and the claim is obvious). Then

\[ \pi_t^i(\sigma \tau) = \phi_{\sigma \tau|_{n-1}}^i \left( t_i + \phi_i \left( \pi_t^i(\sigma^{n+1} \tau) \right) \right), \]

and since \( ||(\phi_{\sigma \tau|_{n-1}}^i)'|| \leq 1 \) we obtain

\[ \frac{\partial}{\partial t_i} \pi_t^i(\sigma \tau) \leq 1 + \phi_i' \cdot \frac{\partial}{\partial t_i} \pi_t^i(\sigma^{n+1} \tau). \]

Proceeding inductively we see that

\[ \frac{\partial}{\partial t_i} \pi_t^i(\sigma \tau) \leq 1 + ||\phi_i'|| + ||\phi_i'||^2 + \ldots = (1 - ||\phi_i'||)^{-1} \]
(recall that \( \phi_i \) is hyperbolic since \( i \leq d < k \)). Together with (7.15) this implies
\[
\frac{\partial}{\partial t_i} (\pi_t(\omega) - \pi_t(\tau)) \geq 1 - |\phi'_j(\pi_t(\sigma\tau))(1 - \|\phi'_i\|)^{-1}
\]
(7.16)
\[
= \frac{1 - \|\phi'_i\| - |\phi'_j(\pi_t(\sigma\tau))|}{1 - \|\phi'_i\|} > \eta
\]
by (7.14). We have verified (7.8), so the strong transversality condition (6.7) holds by Lemma 7.3 and the proof is complete. \( \square \)

**Proof of Proposition 7.2.** is similar to that of Proposition 7.1. We can choose \( \eta > 0 \) so that (7.10) and (7.11) hold, and moreover,
\[
\frac{\partial}{\partial t_i} \pi_t(\xi) < 1 - \eta \quad \text{for all} \quad \xi \in \mathcal{A}^\infty, \ i \leq d, \ t \in B_\eta(0).
\]
Again we only need to check strong transversality. Let \( \omega, \tau \in \mathcal{A}^\infty \) with \( i = \omega_1, j = \tau_1, \) and \( i < j \). If \( \phi_i(X) \cap \phi_j(X) = \emptyset \) we immediately obtain (7.13) by (7.11). If \( \phi_i(X) \cap \phi_j(X) \neq \emptyset \) then \( i \leq d \) and
\[
\pi_t(\omega) - \pi_t(\tau) = \phi_i(t_i + \pi_t(\sigma\omega)) - \phi'_j(\pi_t(\sigma\tau))
\]
\[
= \phi(a_i + t_i + \pi_t(\sigma\omega)) - \phi(a_j + \kappa_j t_j + \pi_t(\sigma\tau))
\]
\[
= \phi'(c)(A(t_j) + t_i + \pi_t(\sigma\omega) - \pi_t(\sigma\tau))
\]
where \( \kappa_j = 1 \) if \( j \leq d \) and 0 otherwise, and \( A(t_j) = a_i - a_j - \kappa_j t_j \) does not depend on \( t_i \). Denoting \( f(t) := A(t_j) + t_i + \pi_t(\sigma\omega) - \pi_t(\sigma\tau) \) we have by (7.5):
\[
\{ t \in B_\eta(0): |\pi_t(\omega) - \pi_t(\tau)| \leq r \} \subset \{ t \in B_\eta(0): |f(t)| \leq u^{-1} r \}.
\]
If we show that \( \frac{\partial f}{\partial t_i} \geq \eta \) then strong transversality will follow from Lemma 7.3. Since all \( \phi_j \) are increasing it is easy to see that \( \frac{\partial}{\partial t_i} \pi_t(\sigma\omega) \geq 0 \), hence
\[
\frac{\partial f}{\partial t_i} \geq 1 - \frac{\partial}{\partial t_i} \pi_t(\sigma\tau),
\]
and the desired statement follows from (7.17).

(ii) It remains to derive (7.6) from (7.7). Let \( 1 \leq i \leq d \). If \( \xi \in \mathcal{A}^\infty \) contains no \( i \), then \( \frac{\partial}{\partial t_i} \pi_t(\xi) = 0 \). Otherwise we write \( \xi = wi\tilde{\xi} \) where \( w \in \mathcal{A}^* \) contains no \( i \) (the word \( w \) may be empty). Then
\[
\pi_t(\xi) = \phi^t_w \phi_i(t_i + \pi_t(\tilde{\xi})),
\]
hence
\[
\frac{\partial}{\partial t_i} \pi_t(\xi) \leq \|\phi^t_w\| \cdot \|\phi'_i\| \left( 1 + \frac{\partial}{\partial t_i} \pi_t(\xi) \right) \leq \|\phi'_i\| \left( 1 + \frac{\partial}{\partial t_i} \pi_t(\tilde{\xi}) \right).
\]
Proceeding inductively we obtain that
\[
\frac{\partial}{\partial t_i} \pi_t(\tilde{\zeta}) \leq \sum_{n=1}^{\infty} \|\phi'_i\|^n = 1 - \frac{1 - 2\|\phi'_i\|}{1 - \|\phi'_i\|} < 1 - \delta
\]
for some $\delta > 0$ by (7.7), and the proof is complete. \hfill \Box

7.2. A class of continued fractions. Here we study in detail IFS arising from the function $\phi(x) = \frac{x}{x+1}$. Let $A = \{a_1, \ldots, a_k\}$ where $a_i > 0$ for $i \leq k-1$ and $a_k = 0$. Let $\Phi_A = \{\phi(x + a_i)\}_{i \leq k}$. Then $\phi_k = \phi$ is parabolic on $[0, 1]$, with the neutral fixed point $v = 0$, for all $A$. The functions $\phi(x + a_i)$, for $i \leq k-1$, are hyperbolic on $[0, 1]$. Clearly, the IFS $\Phi_A$ belongs to $\Gamma_{[0,1]}(1)$ (see Definition 5.5). The IFS $\Phi_A$ with $A = \{\alpha, 0\}$ was recently considered by Lyons [Ly] who studied the properties of the measure which arises from applying the maps randomly and independently with equal probabilities. However, as far as the limit set is concerned, the case $k = 2$ is trivial, since then either the limit set is an interval, or the IFS has no overlaps, so the dimension formula holds for all parameters by [U].

The connection with continued fractions is as follows: The limit set $J_A$ of $\Phi_A$ on $[0, 1]$ coincides with the set of continued fractions of the form
\[
y = [1, Y_1, 1, Y_2, 1, Y_3, 1, \ldots] := 1 + \frac{1}{1 + \frac{1}{Y_1 + \frac{1}{Y_2 + \ldots}}}
\]
where $Y_i \in A$, and also with the set of continued fractions of the form
\[
y = 1 - \frac{1}{(2 + Y_1) - \frac{1}{(2 + Y_2) - \frac{1}{(2 + Y_3) - \ldots}}}
\]
where $Y_i \in A$. Indeed, these representations are immediate by writing
\[
x + \alpha = \frac{1}{1 + \frac{1}{\alpha + x}} \quad \text{and} \quad \frac{x + \alpha}{x + \alpha + 1} = 1 - \frac{1}{(1 + \alpha) + x}.
\]

Example 3. Let $A = \{a, 2, 0\}$. We start with the one-parameter family of IFS $\Phi_A$ corresponding to $a \in (0, 2)$, however, we will see that the non-trivial interval of parameters is smaller.

First observe that the limit set of the IFS $\{\phi(x + a), \phi(x)\}$ is an interval if $a \in (0, 1/2]$. Let $\phi_1(x) = \phi(x + a)$. Then $\phi_1(b) = b$ where $b = \frac{1}{2}(-a + \sqrt{a^2 + 4a})$. This implies that $[0, b] \supset \phi_1([0, b]) \cup \phi([0, b])$, and this becomes equality if and only if $\phi(b) \geq \phi_1(0)$ which is equivalent to $a \leq 1/2$. Since the limit set can only increase when more functions are added, $J_A$ contains an interval if $a \leq 1/2$. 


Since \( \phi_2(x) = \phi(x + 2) \) has the fixed point at \( \sqrt{3} - 1 \), we see that \( Y := [0, \sqrt{3} - 1] \) is the convex hull of \( J_A \). We have, denoting \( \phi_3 = \phi \),

\[
\phi_1(Y) = \left[ \frac{a}{a+1}, \frac{\sqrt{3} - 1 + a}{\sqrt{3} + a} \right],
\]

\[
\phi_2(Y) = [2/3, \sqrt{3} - 1],
\]

\[
\phi_3(Y) = [0, 1 - 1/\sqrt{3}].
\]

Now, \( \phi_2(Y) \cap \phi_3(Y) = \emptyset \). We want to make sure that the IFS has both a “gap” and an “overlap”. Note that \( \phi_1(Y) \cap \phi_2(Y) \neq \emptyset \) if and only if \( a \in (3 - \sqrt{3}, 2) \) and \( \phi_1(Y) \cap \phi_3(Y) \neq \emptyset \) if and only if \( a \in (0, \sqrt{3} - 1) \) (recall that we only consider \( 0 < a < 2 \)). Thus the “interesting” set of parameters is \( U := (1/2, \sqrt{3} - 1) \cap (3 - \sqrt{3}, 2) \).

We want to apply Proposition 7.2 with \( k = 3 \) and \( d = 1 \). Let \( X = [0, \sqrt{3} - 1 + \epsilon] \) where \( \epsilon \) is so small that \( \phi_2(X) \cap \phi_3(X) = \emptyset \). We have

\[
\|\phi'_1\| = \phi'_1(0) = (1 + a)^{-2} < 4/9 < 1/2,
\]

since \( a > \frac{1}{2} \). Thus (7.7) holds and all the assumptions of Proposition 7.2 are satisfied. Let \( s(a) = s(\Phi_A) \) where \( A = \{a, 2, 0\} \); then \( s : U \to \mathbb{R} \) is continuous. Let \( U_{<1} = \{A \in U : s(A) < 1\} \) and \( U_{>1} = \{A \in U : s(A) > 1\} \). We obtain the following:

**Corollary 7.4.**

(i) \( \dim_H(J_A) = \min\{s(a), 1\} \) for Lebesgue-a.e. \( a \in U \).

(ii) For any set \( G \subset U_{<1} \) we have

\[
\dim_H\{a \in G : \dim_H(J_A) < s(a)\} \leq \sup_G s(a).
\]

(iii) \( J_A \) has positive Lebesgue measure for Lebesgue-a.e. \( a \in U_{>1} \).

Note that both \( U_{<1} \) and \( U_{>1} \) are non-empty. Indeed, continuity of \( s(a) \) was established independent of transversality (see Lemma 3.4). Thus \( s(a) \) is continuous on \((0, 2)\). For \( a = \frac{1}{2} \) we know that the limit set of the IFS \( \{\phi_1, \phi_3\} \) is an interval. Since adding an extra function to the IFS makes \( s(a) \) strictly larger, we see that \( s(\frac{1}{2}) > 1 \). By continuity, this implies that \( U_{>1} \neq \emptyset \). On the other hand, for \( a = \sqrt{3} - 1 \) and for \( a = 3 - \sqrt{3} \) the IFS satisfies the open set condition, as the interiors of \( \phi_i(Y) \), \( i = 1, 2, 3 \), are disjoint. Since the limit set is disconnected at these parameter values, it follows, as in [U, Th.6.5], that \( s(\sqrt{3} - 1) < 1 \) and \( s(3 - \sqrt{3}) < 1 \). By continuity, this implies that \( U_{<1} \neq \emptyset \).

**Example 4.** Consider \( \Phi_A = \{\phi(x + a_i)\}_{i \leq k} \) where

\[
A \in U := \{A \in \mathbb{R}^k : a_k = 0, a_j > 0, j = 1, \ldots, k-1\}, \quad k \geq 3.
\]

This is a \((k - 1)\)-parameter family of parabolic IFS.
Corollary 7.5.  
(i) For Lebesgue-a.e. $A \in U$,  
$$\dim_H(J_A) = \min\{s(A), 1\}.$$  
(ii) For any set $G \subset U_{<1}$ we have  
$$\dim_H\{A \in G : \dim_H(J_A) < s(A)\} \leq \sup_G s(A) + (k - 2).$$  
(iii) For Lebesgue-a.e. $A \in U_{>1}$ the set $J_A$ has positive Lebesgue measure.  

Proof. As in Example 3, we have that if $a_i \leq \frac{1}{2}$ for some $i \leq k - 1$, then $J_A$ contains an interval $a_n$ and there is nothing to prove. If $a_i > \frac{1}{2}$ for all $i$, then (7.7) holds. Then we can apply Proposition 7.2 (with $d = k - 1$, when (7.2) is vacuous), and the statements follow from our theorems.  

References  


Received November 16, 1999. The authors were supported in part by the OTKA foundation grant F019099 and NSF grants DMS 9800786 and DMS 9801583.

Department of Stochastics  
Institute of Mathematics  
Technical University of Budapest  
1521 Budapest P.O.Box 91  
Hungary  
E-mail address: simonk@math.bme.hu

Box 354350, Department of Mathematics  
University of Washington  
Seattle, WA 98195-4350  
E-mail address: solomyak@math.washington.edu

Department of Mathematics  
University of North Texas  
Denton, TX 76203-1430  
E-mail address: urbanski@unt.edu
A RIEMANN SINGULARITIES THEOREM FOR PRYM
THETA DIVISORS, WITH APPLICATIONS

Roy Smith and Robert Varley

Let \((P, \Xi)\) be the naturally polarized model of the Prym variety associated to the étale double cover \(\pi : \tilde{C} \to C\) of smooth connected curves, where \(\Xi \subset P \subset \text{Pic}^{2g-2}(\tilde{C})\), and \(g(C) = g\). If \(L\) is any “nonexceptional” singularity of \(\Xi\), i.e., a point \(L\) on \(\Xi \subset \text{Pic}^{2g-2}(\tilde{C})\) such that \(h^0(\tilde{C}, L) \geq 4\), but which cannot be expressed as \(\pi^*(M)(B)\) for any line bundle \(M\) on \(C\) with \(h^0(C, M) \geq 2\) and effective divisor \(B \geq 0\) on \(\tilde{C}\), then we prove \(\text{mult}_L(\Xi) = \frac{1}{2} h^0(\tilde{C}, L)\). We deduce that if \(C\) is nontetragonal of genus \(g \geq 11\), then double points are dense in \(\text{sing}_{st}\Xi = \{L \in \Xi \subset \text{Pic}^{2g-2}(\tilde{C}) \text{ such that } h^0(\tilde{C}, L) \geq 4\}\). Let \(X = \tilde{\alpha}^{-1}(P) \subset \text{Nm}^{-1}(|\omega_C|)\) where \(\text{Nm} : \tilde{C}^{(2g-2)} \to C^{(2g-2)}\) is the norm map on divisors induced by \(\pi\), and \(\tilde{\alpha} : \tilde{C}^{(2g-2)} \to \text{Pic}^{2g-2}(\tilde{C})\) is the Abel map for \(\tilde{C}\). If \(h : X \to |\omega_C|\) is the restriction of \(\text{Nm}\) to \(X\) and \(\varphi : X \to \Xi\) is the restriction of \(\tilde{\alpha}\) to \(X\), and if \(\dim(\text{sing}\Xi) \leq g - 6\), we identify the bundle \(h^*(\mathcal{O}(1))\) defined by the norm map \(h\), as the line bundle \(T_{\varphi} \otimes \varphi^*(K_{\Xi})\) intrinsic to \(X\), where \(T_{\varphi}\) is the bundle of “tangents along the fibers” of \(\varphi\). Finally we give a proof of the Torelli theorem for cubic threefolds, using the Abel parametrization \(\varphi : X \to \Xi\).

Introduction.

If \(C\) is a smooth curve of genus \(g\), among the most basic tools for the study of the natural theta divisor \(\Theta(C) \subset \text{Pic}^{g-1}(C)\) of the Jacobian of \(C\) are Abel’s and Riemann’s theorems that describe the geometry of the “Abel” map \(\alpha : C^{(g-1)} \to \Theta(C)\) parametrizing \(\Theta\) by the symmetric product of the curve. They say the map \(\alpha\) is birational, and that over a point \(L\) of multiplicity \(\mu\) on \(\Theta\), the fiber \(\alpha^{-1}(L) \cong |L| \cong \mathbb{P}^{\mu-1}\), is smooth and isomorphic to the complete linear system \(|L|\), a projective space of dimension \(\mu - 1\). The essential point here is that (one plus) the dimension of the fiber \(\alpha^{-1}(L)\) computes the multiplicity of the point \(L\) on \(\Theta\). It follows also (see [K]) that the normal bundle to the fiber \(\alpha^{-1}(L)\) in \(C^{(g-1)}\) maps onto the tangent cone to \(\Theta\) at \(L\), and that there is a natural determinantal equation for the tangent cone to \(\Theta\) at \(L\).
In the case of the Prym variety of a connected étale double cover \( \pi : \tilde{C} \to C \) of a smooth curve \( C \) of genus \( g \), the natural theta divisor \( \Xi(C) = (P \cap \Theta(C))_{\text{red}} \subset \text{Pic}^{2g-2}(\tilde{C}) \) is parametrized by the restriction \( \varphi : X \to \Xi \) of the Abel map \( \tilde{\alpha} : \tilde{C}^{(2g-2)} \to \Theta(\tilde{C}) \) for \( \tilde{C} \); to the inverse image \( \tilde{\alpha}^{-1}(P) = X \) of the natural translate \( P \subset \text{Pic}^{2g-2}(\tilde{C}) \) of the Prym variety of \( \pi \) (see Section 1 below for the precise definitions). Consequently there are two natural ways to study the theta divisor \( \Xi \), either as the intersection \( (P \cap \Theta)_{\text{red}} \) or as the image of the Abel map \( \varphi : X \to P \). Using the intersection representation \( 2\Xi = (P \cdot \Theta) \), Mumford in [M1, p. 343] gives a Pfaffian equation for the (projectivized) tangent cone \( \text{Pic}_L\Xi \) of \( \Xi \) at a point \( L \) by restricting Kempf’s equation for \( \text{Pic}_L\Theta \). This equation is valid only when the intersection \( \text{Pic}_L P \cap \text{Pic}_L\Theta \) is proper and hence equal as a set to \( \text{Pic}_L\Xi \). Mumford gave a necessary and sufficient condition for the intersection \( \text{Pic}_L P \cap \text{Pic}_L\Theta \) to be proper, but only when \( h^0(\tilde{C}, L) = 2 \). I.e., [M1, Prop., p. 343], when \( h^0(\tilde{C}, L) = 2 \) the intersection \( \text{Pic}_L P \cap \text{Pic}_L\Theta \) is proper if and only if \( L \) is not of form \( \pi^* (M)(B) \) for any line bundle \( M \) on \( C \) with \( h^0(C, M) \geq 2 \) and divisor \( B \geq 0 \) on \( \tilde{C} \). Combining the intersection representation with the Abel parametrization of \( \Xi \), in the present paper we deduce (Theorem 2.1) that Mumford’s condition for the intersection \( \text{Pic}_L P \cap \text{Pic}_L\Theta \) to be proper is sufficient without any hypothesis on \( h^0(\tilde{C}, L) \). We also give a counterexample (Example 2.18) with \( h^0(\tilde{C}, L) = 4 \), to the necessity of the condition. The Abel parametrization \( \varphi : X \to \Xi \) of the theta divisor for Pryms differs from that for Jacobians in that the fiber of the Abel map over a general point on a Prym theta divisor is isomorphic to \( \mathbb{P}^1 \) rather than \( \mathbb{P}^0 \), and also that the source space \( X \) of the Abel-Prym map is not always smooth. Thus there are two concepts of normal space to a fiber of \( \varphi \), the Zariski normal space and the normal cone. We show in Corollary 2.8 that the intersection \( \text{Pic}_L P \cap \text{Pic}_L\Theta \) is the image of the union of the Zariski normal spaces in \( X \) at points of the fiber \( \varphi^{-1}(L) \), and consequently whenever \( X \) is smooth along \( \varphi^{-1}(L) \), then \( \text{Pic}_L P \cap \text{Pic}_L\Theta \) equals \( \text{Pic}_L\Xi \) as sets. It follows that whenever \( X \) is smooth along \( \varphi^{-1}(L) \), one can compute the multiplicity of \( \Xi \) at \( L \), from the dimension of the fiber \( \varphi^{-1}(L) \). I.e., then \( \text{mult}_L(\Xi) = (1/2) h^0(\tilde{C}, L) = (1/2)(1 + \dim \varphi^{-1}(L)) \).

Finally the smoothness criterion of Beauville and Welters is used to show in Lemma 2.15 that \( X \) is singular precisely over “exceptional” singular points of \( \Xi \), those called “case 1” by Mumford in [M1, p. 344]. (See Section 1.6 for the definition.) Consequently one can use this analog for Prym varieties of the Riemann singularities theorem (RST), to compute the multiplicity of \( \Xi \) at all nonexceptional singular points. In Theorem 3.2 and Corollary 3.3 we prove, by generalizing an argument of Welters, a criterion for the fiber \( \varphi^{-1}(L) \) over a generic point \( L \) of a component of \( \text{sing}\Xi \) to be \( \Xi \subset \mathbb{P}^3 \). Combining this with a result of Debarre, we deduce that if \( C \) is nontetragonal of genus \( g \geq 11 \),
and \( \dim(P) = p = g - 1 \), then on every component \( Z \) of \( \text{sing}\Xi \) of dimension \( \geq p - 6 \), double points of \( \Xi \) are dense, and at every double point \( L \) on \( Z \), the quadric tangent cone \( \mathbb{P}C_L\Xi \) contains the Prym canonical curve \( \varphi_\eta(C) \). Since it is known that \( \dim(\text{sing}\Xi) \geq p - 6 \), this adds further evidence at least when \( g \geq 11 \), for a “modified Donagi’s conjecture”, (see [Do, D1, Ve, LS] and Section 1.7 below). In particular, one can ask whether the Prym canonical model of a doubly covered nontetragonal curve \( C \) of genus \( g \geq 11 \) is the unique spanning curve in the base locus of those quadric tangent cones to \( \Xi \) at all double points of components \( Z \) of \( \text{sing}\Xi \) such that \( \dim(Z) \geq p - 6 \).

Since Debarre has shown that a general \( C \) with \( g \geq 8 \) can be recovered in this way, and since every Prym canonical model of a curve \( C \) with \( g \geq 9 \) and Clifford index \( \geq 3 \) is determined by the quadrics containing it ([LS]), our density result brings the state of knowledge on this problem near that which was provided for Jacobians by the paper [AM] of Andreotti and Mayer. A primary problem remaining open is to prove, say for doubly covered nontetragonal curves \( C \) of genus \( g \geq 11 \), that the quadric tangent cones at stable double points generate the ideal of all quadrics containing \( \varphi_\eta(C) \), an analog of Mark Green’s theorem [Gr]. As a further application of the dimension estimate in Proposition 3.1 we deduce Corollary 3.5(i) a criterion for \( \varphi^{-1}(\text{sing}\Xi) \) to have codimension \( \geq 2 \) in \( \mathbb{X} \), and use this to prove (Theorem 4.2) an intrinsic formula for the line bundle defined by the norm map \( h \) on \( \mathbb{X} \). In Section 5 we apply the Riemann singularities theorem to a proof of the Torelli theorem for a cubic threefold \( W \). The proof assumes the usual presentation of the intermediate Jacobian of \( W \) as the Prym variety for a conic bundle representation of \( W \). The new feature is that it describes the geometry of \( \Xi \) via the Abel parametrization, which exists for all Prym varieties, rather than the parametrization via the Fano surface of \( W \), which is somewhat peculiar to the cubic threefold. At the end of the paper we append an outline of the results.

1. Background on Prym varieties.

1.1. General conventions and notation. In this paper all curves considered are smooth, complete, connected, nonhyperelliptic, and defined over \( \mathbb{C} \). (This last restriction seems irrelevant in Sections 2 and 4 where any algebraically closed field of characteristic \( \neq 2 \) should do, but in Section 3, Corollary 3.4, we use results of Debarre [D1] where the field is assumed to be \( \mathbb{C} \), in Lemma 3.6 we use Bertini’s theorem, and in Section 5, Lemma 5.5, we use a result of [SV1] which depends on the characteristic zero Kawamata Viehweg vanishing theorem.) The primary source for the definition and basic properties of Prym varieties is [M1]. References in textbook form are [LB] and [ACGH]. We also use the fundamental results of [B1], [D1], and [We2].
For any variety $V$ and a point $p$ on $V$, we denote by $\mathbb{P}C_pV$ the projectivized tangent cone of $V$ at $p$, and by $\mathbb{P}T_pV$ the projectivized Zariski tangent space. If $S \subset V$ is a subvariety then $\mathbb{P}NC(S/V)$ denotes the projectivized normal cone of $S$ in $V$.

1.2. The Prym variety $(P, \Xi)$ of a double cover $\pi : \tilde{C} \rightarrow C$. The fundamental object of study is a connected étale double cover $\pi : \tilde{C} \rightarrow C$ of smooth curves, where if $g = g(C)$ is the genus of $C$, then $\tilde{g} = g(\tilde{C}) = 2g - 1$. The map $\pi$ induces a norm map $Nm : \text{Pic}^d(\tilde{C}) \rightarrow \text{Pic}^d(C)$ on line bundles for all $d$, and if $d = 0$, the Prym variety of $\pi : \tilde{C} \rightarrow C$, denoted $P_0(\tilde{C}/C)$ or simply $P_0$, is defined to be that connected component of $Nm^{-1}(0) \subset \text{Pic}^0(\tilde{C})$ which contains 0. To obtain a polarization on $P_0$ consider the translate $P \subset \text{Pic}^{2g-2}(\tilde{C})$ defined as $P = \{L \in \text{Pic}^{2g-2}(\tilde{C}) : Nm(L) = \omega_C\}$, and $h^0(\tilde{C}, L)$ is even. Then the reduced codimension one subvariety $\Xi = \{L \in P : h^0(L) > 0\} \subset P$ defines a principal polarization on $P$ such that as divisors, $P \cdot \Theta = 2\Xi$, where $\Theta = \{L : \deg(L) = 2g - 2 = \tilde{g} - 1, h^0(\tilde{C}, L) > 0\} \subset \text{Pic}^{2g-2}(\tilde{C})$ is the canonical theta divisor on $\text{Pic}^{g-1}(\tilde{C})$. The principally polarized Prym variety defined by $\pi$, is the pair $(P_0, \Xi)$ where $\Xi$ is given only up to translation, or (more often for us) the pair $(P, \Xi)$ where the inclusion $\Xi \subset P$ is canonically defined. If $g = g(\tilde{C})$ and $P$ is the Prym variety of $\pi : \tilde{C} \rightarrow C$, we denote the dimension of $P$ by $p = \dim(P) = g - 1$.

1.3. The divisor variety $X$ defined by $\tilde{C} \rightarrow C$. The most important geometric tool for study of a Jacobian variety is the family of Abel maps. In particular for $\tilde{C}$, the principal such map is the birational surjection $\tilde{\alpha} : \tilde{C}^{(2g-2)} \rightarrow \tilde{\Theta} \subset \text{Pic}^{2g-2}(\tilde{C})$, defined by $\tilde{\alpha}(D) = O(D)$. Since $\Xi = P \cap \tilde{\Theta}$ as sets, it is natural to restrict this map over $P$; we denote the resulting map $\varphi : X \rightarrow \Xi \subset P$, the Abel parametrization of the Prym theta divisor $\Xi$, where $X = \tilde{\alpha}^{-1}(P) \subset \tilde{C}^{(2g-2)}$. The question of irreducibility and smoothness of $X$ has been studied by Welters and Beauville in [We2] and [B1]. When $C$ is nonhyperelliptic, $X$ is a reduced, irreducible, normal, local complete intersection variety, in particular Cohen Macaulay. Moreover by [M1], $\varphi$ is a $\mathbb{P}^1$ bundle over $\Xi_{sm} = \{\text{smooth points of } \Xi\}$, and over each point of $\Xi$ the fiber of $\varphi$ is isomorphic to some $\mathbb{P}^n$ with $n$ odd. Indeed, if $L$ is a point of $\Xi \subset \tilde{\Theta} \subset \text{Pic}^{2g-2}(\tilde{C})$, then $L$ is a line bundle on $\tilde{C}$ and $\varphi^{-1}(L) \cong |L|$, so $\dim \varphi^{-1}(L) = h^0(\tilde{C}, L) - 1$, where by definition of $\Xi$, $h^0(\tilde{C}, L)$ is even and positive. It is possible for the fiber dimension of $\varphi$ to be one also over some “exceptional” singular points of $\Xi$.

1.4. The restricted norm map $h : X \rightarrow |\omega_C|$. In addition to the Abel map $\varphi : X \rightarrow \Xi$, the other important map on $X$ is the restriction to $X$ of the norm map $Nm : \tilde{C}^{(2g-2)} \rightarrow C^{(2g-2)}$ on divisors, denoted $h : X \rightarrow |\omega_C|$. Note that by definition of $X$, $Nm$ maps $X$ onto the canonical linear system
|ω_C| on C. Indeed X is defined as a scheme by Welters [We2] and Beauville [B1] as a connected component of the inverse image of |ω_C| under the norm map. Thus if α : C^{(2g-2)} → Θ(C) is the Abel map for C, since P and |ω_C| inherit their reduced scheme structures as components of the fibers Nm^{-1}(ω_C) and α^{-1}(ω_C) respectively, and since the compositions α ∘ Nm and Nm ∘ α are equal, the scheme structure of X is induced either from X ⊂ Nm^{-1}(|ω_C|) = (α ∘ Nm)^{-1}(ω_C), as a connected component of the fiber over ω_C of the composition C^{(2g-2)} → C^{(2g-2)} → Pic^{2g-2}(C), or from X = α^{-1}(P) ⊂ (Nm ∘ α)^{-1}(ω_C), as a connected component of the fiber over ω_C of the composition C^{(2g-2)} → Pic^{2g-2}(C) → Pic^{2g-2}(C). Thus to study X, one extracts from the diagram below:

\[ \begin{array}{ccc}
\tilde{\alpha} : C^{(2g-2)} & \longrightarrow & \text{Pic}^{2g-2}(\tilde{C}) \\
\downarrow Nm & & \downarrow Nm \\
\alpha : C^{(2g-2)} & \longrightarrow & \text{Pic}^{2g-2}(C) 
\end{array} \]

the following diagram of subvarieties and restrictions:

\[ \begin{array}{ccc}
X & \xrightarrow{\varphi} & \Xi \\
\downarrow h & & \downarrow \\
|ω_C| & \longrightarrow & \{ω_C\} 
\end{array} \]

The map h : X → |ω_C| ∼= ℙ^{g-1} is a finite surjection, hence defines an ample line bundle on X.

**Definition 1.4.1.** Denote by O_X(1) the line bundle h^*(O(1)), where O(1) is the standard ample line bundle on the projective space |ω_C|.

In Theorem 4.2 below we give a formula for the line bundle O_X(1), in terms of data intrinsically defined by X, at least for curves C with dim(singΞ) ≤ g−5, i.e., those C not on Mumford’s list in [M1, p. 344]. Now the canonical model of the curve C is the dual variety of the branch divisor of the map h and the curve C parametrizes the irreducible components D_p (see proof of Theorem 4.2 for the definition of the D_p) of the divisors h^*(H) for hyperplanes H tangent to the branch locus of h in |ω_C|. [Using [SV3] for the irreducibility of the divisors D_p, the arguments in [SV4, pp. 357, 360], generalize exactly]. Since the linear system defining h recovers π : C → C, it is of interest to know when it is complete, i.e., when h^0(X, O_X(1)) = g(C). We conjecture this is true when C is nonhyperelliptic, but this remains open for g ≥ 4 (see [SV4, p. 359] when g = 3).

**1.5. Prym canonical curves.** The double cover π : C → C defines a unique square-trivial line bundle η on C by ker(π^* : Pic^0(C) → Pic^0(\tilde{C})) = \{0, η\}. The linear series ω_C ⊗ η is base point free when C is nonhyperelliptic.
and the image $\varphi_\eta(C)$ of the associated projective map $\varphi_\eta : C \to |\omega_C \otimes \eta|^*$ is called the Prym canonical model of $C$. The line bundle $\omega_C \otimes \eta$ is very ample when $C$ is nontetragonal and also when $C$ is a generic tetragonal curve; see [D1] for a precise analysis of those tetragonal curves for which $\omega_C \otimes \eta$ is not very ample.

1.6. Stable and exceptional singularities. A point $L$ of $\Xi \subset \text{Pic}^{2g-2}(\widetilde{C})$ will be called (cf. [D1], [T1]) a “stable singularity” of $\Xi$ (with respect to the double cover $\pi : \widetilde{C} \to C$) if and only if $h^0(\widetilde{C}, L) \geq 4$, and an “exceptional singularity” of $\Xi$ (again with respect to $\pi$) if and only if $L = \pi^*(M)(B)$, where $M$ is a line bundle on $C$ with $h^0(C, M) \geq 2$ and $B \geq 0$ is an effective divisor on $\widetilde{C}$. When a double cover $\widetilde{C} \to C$ representing $(P, \Xi)$ is given or understood, the set of stable singularities is denoted $\text{sing}_{\text{st}} \Xi$, and the set of exceptional singularities is denoted $\text{sing}_{\text{ex}} \Xi$. According to [M1, p. 343], for every Prym representation of $(P, \Xi)$, we have $\text{sing} \Xi = \text{sing}_{\text{st}} \Xi \cup \text{sing}_{\text{ex}} \Xi$. Thus a Prym representation of $(P, \Xi)$ defines a decomposition of $\text{sing} \Xi$ into two generally overlapping subsets, since in particular any line bundle $L = \pi^*(M)(B)$ on $\Xi$, where $M$ is a line bundle on $C$ with $h^0(C, M) \geq 3$, is both stable and exceptional. For example, the unique singularity on the theta divisor of the intermediate Jacobian of a cubic threefold $W$ (see Section 5 below), is both stable and exceptional, for any Prym representation associated to a general line on $W$. Debarre has shown in [D1] that Prym representations of the same abelian variety $(P, \Xi)$ by different double covers of tetragonal curves, can lead to different decompositions of $\text{sing} \Xi$ into stable and exceptional subsets. For $g(C) \geq 7$, i.e., for $p = \text{dim}(P) \geq 6$, $\text{sing}_{\text{st}} \Xi$ is always nonempty and every irreducible component of $\text{sing}_{\text{st}} \Xi$ has dimension $\geq p - 6$, [D2]. By [M1], on any Prym theta divisor $\Xi$, all components of $\text{sing} \Xi$ of dimension $\geq p - 4$ lie entirely in $\text{sing}_{\text{ex}}(\Xi)$, but for a general curve of any genus $\text{sing}_{\text{ex}}(\Xi)$ is empty [see [LB, p. 389], and Prop. 2.19 below].

1.7. Donagi’s conjecture. In [Do] Donagi made his famous “tetragonal conjecture”, which implies that two smooth connected étale double covers $\widetilde{C}_1 \to C_1, \widetilde{C}_2 \to C_2$ of nontetragonal curves $C_i$, are isomorphic as double covers if and only if they define isomorphic polarized Prym varieties $(P_i, \Xi_i)$. Verra found in [Ve] a lovely counterexample where $C_i$ are generic smooth plane sextics (hence of genus 10). He noted that plane sextics are the only curves with the same Clifford index as tetragonal curves and suggested that consequently these may be the only counterexamples. The conjecture must then be modified [cf. [LS]] to assume at least that $\text{Cliff}(C_i) \geq 3$. One approach to proving the modified Donagi’s conjecture, analogous to Green’s result in [Gr] which refines Andreotti Mayer’s approach for Jacobians, is to try to show that a Prym canonical model $\varphi_\eta(C)$ of a nontetragonal curve is determined by the base locus of the quadric tangent cones at appropriately determined double points of $\Xi$. This approach has several complications.
First only the “stable” double points on Ξ have tangent cones which always contain \( \varphi_\eta(C) \). Secondly since the subset \( \text{sing}_{\text{st}} \Xi \) depends on the double cover, one does not prove the conjecture simply by showing that \( \varphi_\eta(C) \) is determined by the base locus of tangent cones to \( \text{sing}_{\text{st}} \Xi \). For example, although there are generally three doubly covered tetragonal curves with the same Prym variety, it is entirely possible that each double cover is determined by the tangent cones to Ξ at those double points which are stable for that double cover. However, there are good reasons to believe this approach will eventually succeed.

Debarre shows in [D1] that for Prym varieties of doubly covered nontetragonal curves \( C \) of genus \( g \geq 11 \), the locus \( \text{sing}_{\text{st}} \Xi \) is intrinsically defined by Ξ, independently of which double cover is considered to represent \((P, \Xi)\). In particular then \( \text{sing}_{\text{st}} \Xi \) is the union of all irreducible components of \( \text{sing} \Xi \) having dimension \( \geq p - 6 \), where \( p = \dim(P) \). In [LS] it is shown using results of Green and Lazarsfeld that for all doubly covered curves \( C \) with \( \text{Cliff}(C) \geq 3 \) and \( g(C) \geq 9 \), that \( \varphi_\eta(C) \) is determined by the base locus of the quadrics containing it. Debarre shows in [D2] for \( g \geq 7 \), and \( C \) general, that the quadric tangent cones to \( \Xi \) at its double points (all of which are stable when \( C \) is general), generate the ideal of quadrics containing \( \varphi_\eta(C) \).

The prerequisite existence and density result for stable double points on generic \( \Xi \) follows from Welters’ “generic Riemann singularities theorem” for Prym varieties in [We1]. In the present paper we provide another step in this approach to Donagi’s conjecture, by proving a precise Riemann singularities theorem for Prym varieties, Theorem 2.1 below, and deducing in Corollary 3.4 that double points are dense in \( \text{sing}_{\text{st}} \Xi \), for every doubly covered nontetragonal curve \( C \) of genus \( g \geq 11 \). In Corollary 2.22 below we also deduce Welters’ generic RST for Prym varieties from the precise version in Theorem 2.1.

2. A Riemann singularities theorem for Prym varieties.

In this section, assume \( \bar{C} \to C \) is an étale connected double cover of a smooth nonhyperelliptic curve \( C \), \( \Xi \subset \bar{\Theta} \subset \text{Pic}^{2g-2}(\bar{C}) \) is the natural model \( \Xi = (\bar{\Theta} \cap P)_{\text{red}} \) for the theta divisor of the associated Prym variety \( P \subset \text{Pic}^{2g-2}(\bar{C}) \), and \( L \) is a point of \( \Xi \). Thus \( L \) is a line bundle on \( \bar{C} \) with a positive even number of global sections, and \( \text{Nm}(L) = \omega_{\bar{C}} \). In particular \( (1/2)h^0(\bar{C}, L) \) is an integer. Since as divisors \( 2\Xi = (\bar{\Theta} \cdot P) \), it follows that \( \text{mult}_L \Xi \geq (1/2)\text{mult}_L \bar{\Theta} = (1/2)h^0(\bar{C}, L) \), and \( (1/2)h^0(\bar{C}, L) \) is the “expected” multiplicity of \( \Xi \) at \( L \). Our goal is a simple criterion for \( \text{mult}_L \Xi \) to equal this expected multiplicity.
2.0. Terminology. If \( \text{mult}_L \Xi = (1/2)h^0(\tilde{C}, L) \), we say the “Riemann singularity theorem” (RST) holds at \( L \). (In that case and that case only, the Pfaffian described by Mumford in [M1, p. 343], the square root of the restriction of Kempf’s equation for \( P \tilde{C}_L \tilde{\Theta} \), gives an equation for the tangent cone \( P \tilde{C}_L \Xi \) to the Prym theta divisor.) Recall that \( L \) is an “exceptional singularity” of \( \Xi \) if and only if it falls in case 1 of Mumford’s description [M1, p. 344] of singularities of \( \Xi \), i.e., if and only if \( L \) lies in \( \Xi \) and \( L = \pi^*(M)(B) \), where \( M \) is a line bundle on \( C \) with \( h^0(C, M) \geq 2 \) and \( B \geq 0 \) is an effective divisor on \( \tilde{C} \).

**Theorem 2.1.** Assume \( \tilde{C} \to C \) is an étale connected double cover of a smooth nonhyperelliptic curve \( C \) of genus \( g \geq 3 \), and \( L \) a point of \( \Xi \subset P \subset \text{Pic}^{2g-2}(\tilde{C}) \). If \( L \) is not an exceptional singularity of \( \Xi \), then \( \text{mult}_L \Xi = (1/2)h^0(\tilde{C}, L) \); i.e., RST holds at every nonexceptional \( L \) on \( \Xi \).

**Corollary 2.2.** If \( C \) is not tetragonal, \( g = g(C) \geq 11 \), and \( \tilde{C} \to C \) is any étale connected double cover, then RST holds at a general point \( L \) of every component of the locus \( \text{sing}_{st} \Xi \) of stable singularities of \( \Xi \), i.e., at a general point of every component of \( \Xi \) of dimension \( \geq g-7 \).

**Remarks 2.3.** (i) Mumford [M1, p. 343] originally proved Theorem 2.1 and its converse when \( h^0(\tilde{C}, L) = 2 \). In [Sh, Lemma 5.7, p. 121] Shokurov generalizes this argument to give a sufficient criterion for the RST condition to fail as follows: If \( H^0(C, L) \) contains a subspace of dimension greater than \( (1/2)h^0(\tilde{C}, L) \) which is isotropic for the form \( \langle s, t \rangle = s \otimes \iota^*(t) - t \otimes \iota^*(s) \) [M1, p. 343], where \( \iota : \tilde{C} \to \tilde{C} \) is the involution associated to the double cover \( \pi : \tilde{C} \to C \), then \( \text{mult}_L \Xi > (1/2)h^0(\tilde{C}, L) \). He applies this to show if \( C \) is a general bielliptic curve, then \( \Xi \) has too many triple points for \( (P, \Xi) \) to be the Jacobian of a curve.

(ii) The converse of Theorem 2.1 can fail when \( h^0(\tilde{C}, L) > 2 \), as we will show below by giving an example of an exceptional singularity \( L \) at which the RST does hold, i.e., one with \( \text{mult}_L \Xi = (1/2)h^0(\tilde{C}, L) \).

(iii) Theorem 2.1 implies (Corollary 2.22 below) Welters’ theorem [We1] that RST holds at every point of \( \Xi \) when \( C \) is a general curve, using only the classical Gieseker Petri theorem [G], [ACGH, p. 215], which implies that there are no exceptional singularities on \( \Xi \) when \( C \) is general.

(iv) We will apply Corollary 2.2 to prove (in Corollary 3.4 below) that if \( C \) is nontetragonal and \( g \geq 11 \), then a general point \( L \) of any component of the locus \( \text{sing}_{st} \Xi \) of stable singularities of \( \Xi \) is a double point. This should be a fundamental initial step in any attempt to generalize the method of Andreatti-Mayer [AM] and Green [Gr] to prove a suitable form of the conjecture of Donagi [Do, Ve, LS], e.g., that the Prym map is injective on the set of doubly covered smooth nontetragonal curves with \( g(C) = g \geq 11 \).
Theorem 2.1 is true also when the curve $C$ is hyperelliptic. In fact if $g(C) \leq 5$, or if $C$ is either hyperelliptic or trigonal, the theorem is immediate since then by \([M1, R, AM]\) either $\dim P = p \leq 4$, or $(P, \Xi)$ is a Jacobian, so every component of $\text{sing} \Xi$ has dimension $\geq p - 4$. Then by \([M1, \text{Lemma}, p. 345]\) the only nonexceptional points of $\Xi$ are smooth points, and at smooth points the conclusion of the theorem follows immediately from the equation $2\Xi = \tilde{\Theta} \cdot P$, \([M1, \text{Cor., p. 342}]\). Pryms of generic doubly covered plane quintic curves are either Jacobians of genus five curves or intermediate Jacobians of cubic threefolds, and then all singularities on $\Xi$ are exceptional as well.

**Proof of Corollary 2.2.** Debarre has shown \([D1, \text{Th. 3.1(i), p. 548}]\) that with the hypotheses of Corollary 2.2 every component of $\text{sing}_{\text{ex}} \Xi$ has lower dimension than any component of $\text{sing}_{\text{st}} \Xi$, so Corollary 2.2 follows immediately from Theorem 2.1.

**Proof of Theorem 2.1.** The first observation is that the problem is purely set theoretic.

**Lemma 2.4.** With the hypotheses of Theorem 2.1, the following statements are equivalent.

1. $\text{mult}_L \Xi = (1/2)h^0(\tilde{C}, L)$.
2. $\mathbb{P}C_L \Xi = \mathbb{P}C_L \tilde{\Theta} \cap \mathbb{P}T_L P$ as sets.
3. $2[\mathbb{P}C_L \Xi] = [\mathbb{P}C_L \tilde{\Theta}] \cdot [\mathbb{P}T_L P]$ as cycles.
4. $\mathbb{P}C_L \tilde{\Theta} \not\supset \mathbb{P}T_L P$.

**Sketch of Proof.** If $\tilde{\vartheta}$ is a Taylor series at $L$ for the theta function of $\tilde{\Theta}$, then $\tilde{\vartheta}$ restricts on $T_L P$ to $\xi^2$, the square of the Taylor expansion at $L$ of a theta function for $\Xi$. Consequently, the lowest order term of $\tilde{\vartheta}$ which does not vanish identically on $T_L P$ equals the square of the lowest nonvanishing term of $\xi$, i.e., equals the square of an equation for the tangent cone of $\Xi$ at $L$. In particular the leading term $\tilde{\vartheta}_h$ of $\tilde{\vartheta}$ defines $\mathbb{P}C_L \Xi$ as a set if and only if $\tilde{\vartheta}_h$ does not vanish identically on $T_L P$. I.e., if $\tilde{\vartheta}_h$ is the lowest nonvanishing term of $\tilde{\vartheta}$ on $T_L \text{Pic}^{2g-2}(\tilde{C})$, hence an equation for the tangent cone of $\tilde{\Theta}$ at $L$, then $\tilde{\vartheta}_h|T_L P = (\tilde{\vartheta}_h + \cdots) = \xi^2 = (\xi_{h/2} + \cdots)^2$. Hence $\mathbb{P}T_L \tilde{\Theta} \not\supset \mathbb{P}T_L P$ iff $\tilde{\vartheta}_h|T_L P = (\xi_{h/2})^2$ is not identically zero, iff $\tilde{\vartheta}_h|T_L P = (\xi_{h/2})^2$ is the square of an equation for the tangent cone of $\Xi$, iff $\xi_{h/2}$ is the first nonvanishing term of $\xi$, iff $\text{mult}_L \Xi = h/2$. Since by the classical Riemann singularities theorem for $\tilde{\Theta}$ we have $h = h^0(\tilde{C}, L)$, the lemma follows.

We can now summarize the proof of Theorem 2.1 as follows: If $\varphi : X \to \Xi$ is the Abel parametrization of the Prym theta divisor by the special variety $X$ of divisors on $\tilde{C}$, we show first in Corollary 2.9 that $\mathbb{P}C_L \Xi = \mathbb{P}C_L \tilde{\Theta} \cap \mathbb{P}T_L P$ holds as sets whenever $X$ is smooth at every point of the fiber $\varphi^{-1}(L)$. Then
we complete the proof by showing in Lemma 2.15 that $X$ is smooth along 
$\varphi^{-1}(L)$ if and only if $L$ is not an exceptional singularity of $\Xi$.

**Tangent spaces to the divisor variety** $X \subset \bar{C}^{(2g-2)}$. Recall that for
each point $L$ of $P$, the inclusion $P \subset \text{Pic}^{2g-2}(\bar{C})$ induces an inclusion of
projective tangent spaces $|\omega_C \otimes \eta|^* = \mathbb{P}T_L P \subset \mathbb{P}T_L \text{Pic}^{2g-2}(\bar{C}) = |\omega_C|^*$,
and that we want to determine the intersection in $|\omega_C|^*$ of the subspace
$\mathbb{P}T_L P$ with the projectivized tangent cone $\mathbb{P}C_L \bar{\Theta}$. From the Riemann Kempf
singularities theorem [K, Thm. 1, p. 178] the cone $\mathbb{P}C_L \bar{\Theta}$ is the union of
the projectivized images, under the differential of the Abel map, of the
tangent spaces to the divisor variety $X$ tangent to $(\bar{\Theta})$. Therefore
we need to describe the Zariski tangent space to $X$ at a point $\bar{D}$ of $|L|$.

**Lemma 2.5.** If $\bar{C} \rightarrow C$ is any smooth connected étale double cover, $C$
nonhyperelliptic, $X \subset \bar{C}^{(2g-2)}$ is the special variety of divisors on $\bar{C}$, and $\bar{\alpha}_*$
is the differential of the Abel map $\bar{\alpha} : \bar{C}^{(2g-2)} \rightarrow \text{Pic}^{2g-2}(\bar{C})$ for $\bar{C}$,
then the Zariski tangent space to $X$ at $\bar{D}$ is given by:

$$
(2.5.1) \quad T_{\bar{D}}X = (\bar{\alpha}_* \bar{D})^{-1}(T_L P).
$$

**Proof.** The scheme structure of $X$ may be defined by pulling back that of $P$, $X = \bar{\alpha}^{-1}(P)$. Thus the Zariski tangent space to $X$ is also a pull back from that of $P$, i.e., $T_{\bar{D}}X = \bar{\alpha}_*^{-1}(T_L P)$.

Since the cone $\mathbb{P}C_L \bar{\Theta}$ is ruled by the image spaces $\mathbb{P}\bar{\alpha}_*(T_{\bar{D}} \bar{C}^{(2g-2)}) = 
\langle \bar{D} \rangle$ = the span of the divisor $\bar{D}$ in the canonical space $|\omega_{\bar{C}}|^*$ of the curve $\bar{C}$, in order to intersect $\mathbb{P}T_L P$ with $\mathbb{P}C_L \bar{\Theta}$, a natural first step is to intersect $\mathbb{P}T_L P$ with each ruling $\langle \bar{D} \rangle$.

**Lemma 2.6.** The intersection $\mathbb{P}T_L P \cap \langle \bar{D} \rangle$ equals the projectivized image, in
$\mathbb{P}T_L P \cong |\omega \otimes \eta|^*$, of the Zariski tangent space $T_{\bar{D}}X$ under the derivative $\varphi_* \bar{D}$
of the restricted Abel map $\varphi : X \rightarrow \Xi$. I.e., $\mathbb{P}\varphi_* T_{\bar{D}} X = ((\mathbb{P}T_L P) \cap \langle \bar{D} \rangle) \subset (\mathbb{P}T_L P \cap \mathbb{P}C_L \bar{\Theta})$.

**Proof.** Since the map $\bar{\alpha}_* \bar{D} : \mathbb{P}T_{\bar{D}} \bar{C}^{(2g-2)} \rightarrow \langle \bar{D} \rangle$ is surjective by the Riemann
Kempf theorem (see also [MM]), its restriction to $(\bar{\alpha}_* \bar{D})^{-1}(\langle \bar{D} \rangle) = (\bar{\alpha}_* \bar{D})^{-1}(\mathbb{P}T_L P) = T_{\bar{D}}X$ (by (2.5.1)), surjects onto $\langle \bar{D} \rangle \cap \mathbb{P}T_L P$. Since $\varphi_* \bar{D}$ is the restriction to $T_{\bar{D}}X$ of $\bar{\alpha}_* \bar{D}$, thus $\mathbb{P}\varphi_* T_{\bar{D}} X = \mathbb{P}T_L P \cap \langle \bar{D} \rangle$ as claimed.

**Corollary 2.7.** For any point $\bar{D}$ of $X$, $\dim(\mathbb{P}T_L P \cap \langle \bar{D} \rangle) = \dim(\mathbb{P}\bar{\alpha}_* T_{\bar{D}}X)$
$= \dim T_{\bar{D}}X - \dim |\bar{D}| - 1$. 


Proof. This follows from the rank formula for a linear map. I.e., the linear map \( \varphi_{*,\bar{D}} \) has domain \( T_{\bar{D}}X \), projectivized image = \( \mathbb{P}T_L P \cap \langle \bar{D} \rangle \), and we claim the kernel equals \( T_{\bar{D}}|\bar{D}| \). Indeed, since all fibers of the abel map \( \bar{\alpha} \) on \( \bar{C}^{(2g-2)} \) are nonsingular, and \( \bar{\alpha}^{-1}(\mathcal{O}(\bar{D})) = |\bar{D}| \), the kernel of \( \bar{\alpha}_{*,\bar{D}} \) equals \( T_{\bar{D}}|\bar{D}| \) which has the same dimension as \( |\bar{D}| \), and since \( |\bar{D}| \subset X \), we also have kernel \( (\varphi_{*,\bar{D}}) = T_{\bar{D}}|\bar{D}| \). □

Corollary 2.8. As sets, the intersection \( \mathbb{P}C_L \Theta \cap \mathbb{P}T_L P \) equals the union of the images \( \mathbb{P}\varphi_*(T_{\bar{D}}X) \) of all the Zariski tangent spaces to \( X \) at points \( \bar{D} \) of \( |L| = \varphi^{-1}(L) \).

Proof. This is immediate from Lemma 2.6 and the Riemann Kempf theorem. □

Thus to determine when the intersection \( \mathbb{P}C_L \Theta \cap \mathbb{P}T_L P \) equals the tangent cone \( \mathbb{P}C_L \Xi \) as sets, we only need to determine when that tangent cone is the set theoretic image of the Zariski tangent spaces along the fiber \( \varphi^{-1}(L) \). For a proper map between smooth varieties, if the scheme theoretic fiber over a point \( L \) of the target variety is also smooth, then the normal bundle to the fiber surjects onto the tangent cone to the image variety at \( L \), [K, Lemma p. 179], [MM, p. 230]. Since the scheme theoretic fibers of \( \varphi \) are equal to the corresponding fibers of the Abel map \( \bar{\alpha} \), they are always smooth, and we get the following abstract version of the RST for Prym varieties.

Corollary 2.9. With the hypotheses of Theorem 2.1, if \( X \) is smooth at every point of the fiber \( \varphi^{-1}(L) \), then \( \mathbb{P}C_L \Xi = \mathbb{P}C_L \Theta \cap \mathbb{P}T_L P \) as sets.

Proof. The projective tangent cone \( \mathbb{P}C_L \Xi \) is the exceptional fiber over \( L \) of the blowup of \( \Xi \) at \( L \), and the projective normal cone in \( X \) to \( \varphi^{-1}(L) = |L| \) is the exceptional fiber of the blowup of \( X \) along \( |L| \). Since \( \Xi = \varphi(X) \) and \( \varphi : X \to \Xi \subset \text{Pic}^{2g-2}(\bar{C}) \) is proper, the map induced by \( \varphi \) on these blowups is surjective. In particular the exceptional fiber over \( |L| \) surjects onto the exceptional fiber over \( L \), i.e., the projective normal cone in \( X \) to \( |L| \) surjects onto the projective tangent cone \( \mathbb{P}C_L \Xi \), whether \( X \) is smooth or not. Since the scheme theoretic fibers \( \varphi^{-1}(L) = \bar{\alpha}^{-1}(L) \) are equal, and the fibers \( \bar{\alpha}^{-1}(L) \) are always smooth by the Mattuck Mayer version of the Riemann Roch theorem [MM], the fibers \( \varphi^{-1}(L) \) are also smooth. Thus whenever \( X \) is smooth at \( \bar{D} \), the projective normal space \( \mathbb{P}N_{\bar{D}}(|L|/X) \) is a fiber of the projective normal cone \( \mathbb{P}NC(|L|/X) \) and the induced map is defined there by the derivative \( \bar{\alpha}_* \). Since the tangent space \( T_{\bar{D}}X \) and the normal space \( N_{\bar{D}}(|L|/X) = T_{\bar{D}}X/T_{\bar{D}}(|L|) = T_{\bar{D}}X/\text{ker} \bar{\alpha}_* \), have the same image under \( \bar{\alpha}_* \), Corollary 2.9 follows from Corollary 2.8. □

We deduce the following abstract version of Mumford’s result:
Corollary 2.10. If \( L \) is a point of \( \Xi \) such that \( h^0(\tilde{C}, L) = 2 \), then \( \Xi \) is singular at \( L \) iff the RST theorem fails for \( \Xi \) at \( L \), iff \( L = \mathcal{O}(\tilde{D}) = \varphi(\tilde{D}) \) is the image of some singular point \( \tilde{D} \) of \( X \).

Proof. If \( h^0(\tilde{C}, L) = 2 \), then the RST holds at \( L \) iff \( \text{mult}_L \Xi = 1 \), iff \( \Xi \) is smooth at \( L \). If \( X \) is smooth at every point \( \tilde{D} \) in \( \varphi^{-1}(L) \), then Corollary 2.9 implies the RST holds at \( L \). If \( X \) is singular at \( \tilde{D} \), since \( \dim X = g - 1 \), then \( \dim(T_{\tilde{D}} X) \geq \dim(X) + 1 \geq g \). If \( L = \mathcal{O}(\tilde{D}) \) and \( h^0(\tilde{C}, L) = 2 \), then \( \dim(\tilde{D}) = 1 = \dim(\ker(\varphi_*)) \), hence by Corollary 2.7 \( \dim \mathbb{P} \varphi_*(T_{\tilde{D}} X) \geq g - 2 = p - 1 = \dim \mathbb{P} T_L P \). Since by Lemma 2.6, \( \mathbb{P} \varphi_*(T_{\tilde{D}} X) = \mathbb{P} T_L P \cap (\tilde{D}) \), it follows that \( \dim \mathbb{P} T_L P = \dim(\mathbb{P} T_L P \cap (\tilde{D})) \). Thus \( \mathbb{P} T_L P = (\mathbb{P} T_L P \cap (\tilde{D})) \subset (\mathbb{P} T_L P \cap \mathcal{P} C_\ell \Theta) \subset \mathcal{P} C_\ell \Theta \). Hence RST fails at \( L = \mathcal{O}(\tilde{D}) \), by Lemma 2.4 (iv).

Detecting singularities of \( X \). To complete the proof of Theorem 2.1 we will relate the smoothness of \( X \) to the existence of exceptional singularities on \( \Xi \), in particular we show that \( L \) is an exceptional singularity of \( \Xi \) if and only if \( X \) is singular at some point \( \tilde{D} \) of \( \varphi^{-1}(L) \). We will use formula (2.5.1) for the tangent space to \( X \) to deduce a smoothness criterion for \( X \), and then relate it to Beauville’s formulation of Welters’ criterion. First of all, to measure when \( X \) is singular at \( \tilde{D} \) we need to compute the dimension of the tangent space \( T_{\tilde{D}} X \). Denote \( H^0(C, \omega_C) \) by \( \Omega_C \) and \( H^0(\tilde{C}, \omega_{\tilde{C}}) \) by \( \Omega_{\tilde{C}} \).

Lemma 2.11. For any point \( \tilde{D} \) of \( X \), \( \dim T_{\tilde{D}} X = g - 2 + \dim \{ \omega \text{ in } \Omega_C \text{ such that } (\pi^*(\omega)) \geq \tilde{D} \} \).

Proof. By formula (2.5.1) \( T_{\tilde{D}} X = (\tilde{\alpha}_s)^{-1}(T_L P) \) is the tangent space to \( X \) at \( \tilde{D} \). Hence \( T_{\tilde{D}} X \) is defined as a subspace of \( T_{\tilde{D}} \tilde{C}(2g - 2) \) by the pullback of those linear equations in \( T_L \mathbb{P} \text{Pic}^{2g - 2}(\tilde{C}) \) which vanish on \( T_L P \) where \( L = \mathcal{O}(\tilde{D}) \). Now \( T_L \mathbb{P} \text{Pic}^{2g - 2}(\tilde{C}) \cong \Omega_{\tilde{C}} \) and the subspace of equations vanishing on \( T_L P \) corresponds to the subspace \( \pi^*(\Omega_C) = (\Omega_{\tilde{C}})^+ \subset \Omega_{\tilde{C}} \). Hence the codimension of \( T_{\tilde{D}} X \) in \( T_{\tilde{D}} \tilde{C}(2g - 2) \) equals the number of equations in \( \pi^*(\Omega_C) \) minus the number which pull back trivially to \( T_{\tilde{D}} \tilde{C}(2g - 2) \), i.e., it equals \( g - \dim \{ \omega \text{ in } \Omega_C : \pi^*(\omega) \text{ vanishes on } \tilde{\alpha}_s T_{\tilde{D}}(\tilde{C}(2g - 2)) \} \). Since \( \mathbb{P} \tilde{\alpha}_s T_{\tilde{D}}(\tilde{C}(2g - 2)) = (\tilde{D}) \) = the span of the divisor \( \tilde{D} \) on the canonical model of \( \tilde{C} \) in \( \mathbb{P}^*(\Omega_{\tilde{C}}) \cong \mathbb{P} T_L \text{Pic}^{2g - 2}(\tilde{C}) \), the linear form \( \pi^*(\omega) \) vanishes on \( \tilde{\alpha}_s T_{\tilde{D}}(\tilde{C}(2g - 2)) \) if and only if \( (\pi^*(\omega)) \geq \tilde{D} \). The following sequence thus summarizes the calculation.

\[
0 \rightarrow \{ \omega \text{ in } \Omega_C : (\pi^*(\omega)) \geq \tilde{D} \} \rightarrow (\Omega_{\tilde{C}})^+ \rightarrow T_{\tilde{D}}^*(\tilde{C}(2g - 2)) \rightarrow T_{\tilde{D}}^*(X) \rightarrow 0.
\]

Thus \( \dim T_{\tilde{D}}(X) = \dim T_{\tilde{D}}^*(X) = (2g - 2) - g + \dim \{ \omega \text{ in } \Omega_C \text{ such that } (\pi^*(\omega)) \geq \tilde{D} \} = g - 2 + \dim \{ \omega \text{ in } \Omega_C \text{ such that } (\pi^*(\omega)) \geq \tilde{D} \} \).
Corollary 2.12 (smoothness criterion). $X$ is smooth at $\tilde{D}$ if and only if the only differentials $\omega$ on $C$ such that $\pi^*(\omega)$ vanishes on $\tilde{D}$ are the multiples of $\omega_0 = \text{the differential vanishing on } \text{Nm}(\tilde{D}) = D_0$.

Proof. $X$ is smooth iff $\dim T_{\tilde{D}} X = \dim X = g - 1$, and by Lemma 2.11, this is equivalent to $\dim \{ \omega \in \Omega_C \text{ such that } (\pi^*(\omega)) \geq \tilde{D} \} = 1$. □

Next we relate this to Beauville’s formulation [B1] of Welters’ criterion [We2] for smoothness of $X$ at $\tilde{D}$.

Lemma 2.13. $X$ is singular at $\tilde{D}$ iff there exists an effective divisor $A \geq 0$ on $C$ such that $h^0(C, A) \geq 2$ and $\pi^*(A) \leq \tilde{D}$.

Proof. Let $\tilde{D} = p_1 + p'_1 + \cdots + p_r + p'_r + q_1 + \cdots + q_s$, where each pair $\{p_i, p'_i\}$ is a conjugate pair, and the set $p_1, \ldots, p_r, q_1, \ldots, q_s$ contains no conjugate pairs. Then for any divisor $A$ on $C$, $\pi^*(A) \leq \tilde{D}$ iff $A \leq \overline{p}_1 + \cdots + \overline{p}_r$, where $\overline{p} = \pi(p)$. Moreover for any differential $\omega$ on $C$, $\pi^*(\omega) \geq \tilde{D}$ iff $(\omega) \geq \overline{p}_1 + \cdots + \overline{p}_r + \overline{q}_1 + \cdots + \overline{q}_s$. Now if $X$ is singular at $\tilde{D}$, by Corollary 2.12 there are two independent differentials $\omega_1$, $\omega_2$ on $C$ such that $\pi^*(\omega_i) \geq \tilde{D}$, and if we define $A = \overline{p}_1 + \cdots + \overline{p}_r$, and $B = \overline{q}_1 + \cdots + \overline{q}_s$, then for $i = 1, 2$ we have $(\omega_i) \geq A + B$, hence $h^0(K - A - B) \geq 2$. Since also $\pi_* (\tilde{D}) = 2A + B$ is a canonical divisor, $h^0(A) = h^0(K - A - B) \geq 2$, and $\pi^*(A) \leq \tilde{D}$, so that the Beauville - Welters criterion is satisfied. Conversely, if there is an $A \geq 0$ such that $h^0(A) \geq 2$ and $\pi^*(A) \leq \tilde{D}$, then since $A \leq \overline{p}_1 + \cdots + \overline{p}_r$, the same two properties hold for $\overline{p}_1 + \cdots + \overline{p}_r$, so we may as well assume $A = \overline{p}_1 + \cdots + \overline{p}_r$. Then again, $h^0(K - A - B) = h^0(A) \geq 2$, so there are at least 2 independent differentials $\omega$ such that $(\omega) \geq A + B$, and hence such that $\pi^*(\omega) \geq \tilde{D}$, whence by Corollary 2.12 $X$ is singular at $\tilde{D}$. □

This yields the following alternate dimension formula for $T_{\tilde{D}} X$.

Corollary 2.14. At any point $\tilde{D}$ of $X$, if $A \geq 0$ is the largest effective divisor on $C$ such that $\pi^*(A) \leq \tilde{D}$, then $\dim T_{\tilde{D}} X = g - 2 + h^0(A)$.

Proof. In the notation of the previous proof, Serre duality yields $h^0(A) = h^0(K - A - B) = \dim \{ \omega \in \Omega_C \text{ such that } (\pi^*(\omega)) \geq \tilde{D} \}$. □

The usefulness of the B-W formulation of singularity of $X$ at $\tilde{D}$, is its close connection with the concept of exceptional singularities.

Lemma 2.15. A point $L$ on $\Xi$ is an exceptional singularity iff $L = \mathcal{O}(\tilde{D}) = \varphi(\tilde{D})$ for some singular point $\tilde{D}$ on $X$. I.e., $\text{sing}_{\text{ex}}(\Xi) = \varphi(\text{sing}X)$.

Proof. If $L = \mathcal{O}(\tilde{D})$ for some $\tilde{D}$ at which $X$ is singular, then by Lemma 2.13, $\tilde{D} = \pi^*(A)(B)$ where $h^0(A) \geq 2$ and $B \geq 0$, and then taking $M = \mathcal{O}(A)$, $L$ is exceptional. Conversely, if $L = \pi^*(M)(B)$ for some line bundle $M$ on $C$
with \( h^0(M) \geq 2 \) and \( B \geq 0 \), and if \( A \) is any divisor in \(|M|\), then \( \pi^*(A)(B) = \bar{D} \) belongs to \(|L| = \varphi^{-1}(L) \subset X \), and \( X \) is singular at \( \bar{D} \geq \pi^*(A) \), by Lemma 2.13.

Now Corollary 2.9, Lemma 2.4 and Lemma 2.15 imply Theorem 2.1.

**Corollary 2.16.** If \( L \) is a point of \( \Xi \) which is not an exceptional singularity, then the RST holds at \( L \), i.e., \( \mathbb{P}C_L \Xi = \mathbb{P}C_L \bar{\Theta} \cap \mathbb{P}T_L \mathbb{P} \) as sets, and hence \( \text{mult}_L \Xi = (1/2) h^0(C, L) \).

**Proof.** If \( L \) is a point of \( \Xi \) which is not an exceptional singularity, then \( X \) is smooth at every point \( \varphi^{-1}(L) \), so RST holds at \( L \) by Corollary 2.9.

And Lemma 2.4.

In particular we recover Mumford’s result in its original form.

**Corollary 2.17.** If \( L \) is a point of \( \Xi \) such that \( h^0(\bar{\Theta}, L) = 2 \), then \( \Xi \) is singular at \( L \) iff \( L \) is an exceptional singularity. (In particular, “exceptional singularities” are really singular points of \( \Xi \).

**Proof.** This follows from Corollary 2.10 and Lemma 2.15.

The converse of Corollary 2.16 can fail when \( h^0(\bar{\Theta}, L) \geq 4 \).

**Example 2.18.** Let \( C \) be a nonhyperelliptic genus 5 curve with two vanishing even theta nulls \( M_1, M_2 \) with \( h^0(M_i) = 2 \), and let the line bundle \( \eta \) associated to the double cover be defined by their difference \( \eta = M_1 - M_2 \).

Then \( L = \pi^*(M_1) \) implies \( \text{Nm}(L) = 2M_1 = \omega_C \), and \( h^0(L) = h^0(M_1) + h^0(M_1 + \eta) = h^0(M_1) + h^0(M_2) = 4 \), so \( L \) is a stable and exceptional singularity on \( \Xi \). However by [V, p. 948, ll. 1-3], \( L \) is then a vanishing even theta null on \( (P, \Xi) \) so \( \text{mult}_L \Xi = \text{either 2 or 4} \). Since \( C \) is nonhyperelliptic, hence by [M1, p. 344] \( \text{sing} \Xi \) is zero dimensional, \( P \) is indecomposable so \( \text{mult}_L \Xi \leq 3 \) by [SV1, p. 319]. Hence \( \text{mult}_L \Xi = 2 = h^0(L)/2 \), and \( L \) is both a stable and exceptional double point on \( \Xi \) at which RST holds.

**Gieseker’s theorem and exceptional singularities.** We recall the following proof from [LB], modifying it slightly to conform to our definition of exceptional singularity.

**Proposition 2.19.** If \( C \) is a general curve of genus \( \geq 2 \), then for every double cover \( \bar{C} \to C \), there are no exceptional singularities on \( \Xi \).

**Proof [LB, Remark (6.7) p. 389].** If \( C \) has a double cover with an exceptional singularity \( L \) on \( \Xi \), then by definition \( L = \pi^*(M)(B) \), for some line bundle \( M \) on \( C \) with \( h^0(M) \geq 2 \) and some divisor \( B \geq 0 \) on \( \bar{C} \). Then \( \text{Nm}(L) = (2M)(\text{Nm}(B)) = \omega_{\bar{C}} \), where \( \text{Nm}(B) \geq 0 \). Hence \( h^0(\omega_{\bar{C}} - 2M) = h^0(\text{Nm}(B)) \geq 1 \). We can deduce that \( C \) is special in moduli. For then we can choose a 2 dimensional subspace \( W \subset H^0(C, M) \) and consider...
the cup product map \( \mu : W \otimes H^0(C, K - M) \to H^0(C, K) \). If \( E \) is the base locus of the pencil \([W]\), the base point free pencil trick [ACGH, p. 126] implies the kernel of \( \mu \) is isomorphic to \( H^0(K - 2M + E) \). Since \( h^0(K - 2M + E) \geq h^0(K - 2M) \geq 1 \), the cup product map \( \mu \) above is not injective, thus neither is Petri’s map \( \mu_0 : H^0(M) \otimes H^0(C, K - M) \to H^0(C, K) \), of which \( \mu \) is a restriction. Then by Gieseker’s theorem, [G], [ACGH, Thm (1.7), p. 215], the curve \( C \) is not general. \( \square \)

Remarks 2.20. The apparent contradiction between the two statements:
(i) that for \( g(C) \geq 2 \) there are in general no exceptional singularities on \( \Xi \), and (ii) the theorem of Mumford that for \( g(C) \leq 4 \), all singularities on \( \Xi \) are exceptional, is of course resolved by the fact that in this range a general \( \Xi \) has no singularities at all.

Proposition 2.19 gives a proof of the following result, whose statement was communicated privately to us by Debarre.

Corollary 2.21. For any double cover of a general curve \( C \) of genus \( g \geq 2 \), the special variety of divisors \( X \) is smooth.

Proof. By Lemma 2.15, \( \text{sing}X \subset \varphi^{-1}(\text{sing}_{\text{ex}}\Xi) \), and by Proposition 2.19, for general \( C \), \( \text{sing}_{\text{ex}}\Xi = \emptyset \). \( \square \)

Corollary 2.22 ([We1]). If \( C \) is a general curve of genus \( g \geq 2 \), then for any connected \( \acute{\text{e}} \text{tale} \) double cover \( \pi : \tilde{C} \to C \), RST holds everywhere on \( \Xi \).

Proof. Since for any double cover of a general curve \( C \), \( \Xi \) has no exceptional singularities, RST holds everywhere on \( \Xi \). \( \square \)

3. On the density of double points in \( \text{sing}_{\text{st}}\Xi \).

As always, assume \( C \) is a smooth nonhyperelliptic curve and \( \pi : \tilde{C} \to C \) a connected \( \acute{\text{e}} \text{tale} \) double cover. For potential use in the Andreotti-Mayer-Green approach to Donagi’s conjecture, we want to give a criterion for the existence of as many stable double points on \( \Xi \) as can be hoped for. We will show in Corollary 3.4 below that if \( C \) is nontetragonal and of genus \( g \geq 11 \), then for all double covers of \( C \), double points are dense in every component of \( \text{sing}_{\text{st}}\Xi \). By Corollary 2.2 it would suffice to show the existence of points \( L \) with \( h^0(\tilde{C}, L) = 4 \), or equivalently with \( h^0(\tilde{C}, L) \leq 4 \) on every component of \( \text{sing}_{\text{st}}\Xi \). Note that if \( C \) is nonhyperelliptic, and \( 3 \leq g(C) \leq 6 \), then \( X \) is irreducible and \( 2 \leq \dim(X) \leq 5 \). Hence for any \( L \) on \( \Xi \), we have \( h^0(\tilde{C}, L) - 1 = \dim \varphi^{-1}(L) \leq \dim(X) - 1 \leq 4 \), so that \( h^0(\tilde{C}, L) \leq 5 \), and since \( h^0(\tilde{C}, L) \) is even, in fact then \( h^0(\tilde{C}, L) \leq 4 \) for all points \( L \) on \( \Xi \). Hence giving a criterion for \( h^0(\tilde{C}, L) \leq 4 \) to hold at a general point \( L \) of a component of \( \text{sing}\Xi \) is a challenge only when \( g(C) \geq 7 \). We are not in fact able to rule out the possible existence of small components of \( \text{sing}\Xi \) on which
$h^0$ is always $\ge 6$, but we can obtain the estimate $h^0(\tilde{C}, L) \le 4$ at general points of relatively large components of $\text{sing}\Xi$. To do this we globalize an argument of Welters [We1, Lemma 3.2, p. 681] for changing arbitrary points $L$ of $\text{sing}_{\text{st}}\Xi$ into ones with $h^0(\tilde{C}, L) = 4$, using the “parity trick” of Mumford [M2, bottom of p. 186]. I.e., if $N\text{m}(L) = \omega_C, h^0(L) \ge 1$, and $p$ is not a base point of $|L|$, then $N\text{m}(L(p' - p)) = \omega_C$ and $h^0(L(p' - p)) = h^0(L) - 1$.

Applying this principle twice changes a point $L$ of $\Xi$ with $h^0(L) \ge 4$ into another point $L'$ of $\Xi$ with $h^0(L') = h^0(L) - 2$. As just described, this trick gives no information on whether the new point $L'$ lies on the same component of $\text{sing}\Xi$ as the original point $L$. To show every component of $\text{sing}_{\text{st}}\Xi$ contains a point with $h^0 = 4$ we use the following global version of the parity trick (whose hypotheses are vacuous for $g \le 4$).

**Proposition 3.1.** Assume $C$ is smooth, nonhyperelliptic, of genus $g \ge 5$, and $\pi : \tilde{C} \to C$ any étale connected double cover. Let $Z \subset \varphi^{-1}(\text{sing}\Xi)$ be any irreducible component of $\varphi^{-1}(\text{sing}\Xi)$ on which the generic fiber of $\varphi$ is $\equiv \mathbb{P}^r$, with $r \ge 3$. Then there exists a closed irreducible subvariety $Z' \subset X$ such that $\dim(Z') = \dim(Z)$, and $|D'| \equiv \mathbb{P}^{r-2}$ for $D'$ general on $Z'$. In particular $\dim(\varphi(Z')) \ge \dim(\varphi(Z)) + 2$.

Assuming Proposition 3.1, we deduce the following results.

**Theorem 3.2.** If $C$ is smooth, not hyperelliptic, $g(C) = g \ge 3, \tilde{C} \to C$ is an étale connected double cover, $\varphi : X \to \Xi$ is the Abel map, and $Z$ an irreducible component of $\varphi^{-1}(\text{sing}\Xi)$ such that $\dim(\varphi(Z)) \ge \dim(\text{sing}\Xi) - 1$, then $h^0(\tilde{C}, L) \le 4$ at a general point $L$ on $\varphi(Z)$.

**Proof.** We have shown in the remarks just above Proposition 3.1 that $h^0(\tilde{C}, L) \le 4$ is true everywhere on $\Xi$ if $g(C) \le 6$. Assuming $g \ge 7$ and that the theorem is false, there is an irreducible component $Z$ of $\varphi^{-1}(\text{sing}\Xi)$ such that $\dim(\varphi(Z)) \ge \dim(\text{sing}\Xi) - 1$, and at a general point $L$ on $\varphi(Z)$ we have $h^0(\tilde{C}, L) \ge 6$. Then $|L| \equiv \mathbb{P}^r$ where $r \ge 5$, whence $|D'| \equiv \mathbb{P}^{r-2}$, with $r - 2 \ge 3$, where $D'$ is a generic point of the variety $Z'$ constructed in Proposition 3.1. Then $\varphi(Z') \subset \text{sing}_{\text{st}}\Xi$, but $\dim(Z') \ge \dim(\text{sing}\Xi) + 1$, a contradiction.

**Corollary 3.3.** Assume $C$ is smooth, nonhyperelliptic, $g(C) = g \ge 6$ and $\dim(\text{sing}\Xi) \le p - 5 = g - 6$, i.e., $C$ not on Mumford’s list in [M1, Thm., p. 344]. Then for $W \subset \text{sing}_{\text{st}}\Xi$ any component of stable singularities, and a general point $L$ on $W$, we have $h^0(\tilde{C}, L) = 4$.

**Proof.** If $\tilde{C} \to C$ is an étale connected double cover such that $\dim(\text{sing}\Xi) \le p - 5$, and $W \subset \text{sing}_{\text{st}}\Xi$ is any component of stable singularities, then $\dim W \ge p - 6 \ge \dim(\text{sing}\Xi) - 1$. Thus by Theorem 3.2, for a general point $L$ on $W$, we have $h^0(\tilde{C}, L) \le 4$. Since also $h^0(\tilde{C}, L) \ge 4$ by definition of stable singularities, it follows that $h^0(\tilde{C}, L) = 4$. □
Corollary 3.4. If $C$ is nontetragonal and $g(C) = g \geq 11$, and $W \subset \mathrm{sing}_{\text{st}} \Xi$ is any component of stable singularities, then at a general point $L$ on $W$, $\mathrm{mult}_L \Xi = 2$.

Proof. With these hypotheses, $\dim(W) \geq p - 6$, and $\dim(\mathrm{sing}_{\text{ex}} \Xi) \leq p - 7$ by [D1, Th. 3.1(i), pp. 547-8]. Hence a general point $L$ on $W$ is not exceptional, so RST holds at $L$ by Theorem 2.1. Since $C$ is not on Mumford’s list in [M1, p. 344], also $\dim(W) \leq \dim(\mathrm{sing} \Xi) \leq p - 5$. Thus by Corollary 3.3, $\mathrm{mult}_L \Xi = 2$. □

Corollary 3.5. Assume $C$ is a smooth nonhyperelliptic curve and $\pi : \tilde{C} \to C$ a connected étale double cover.

(i) If $g(C) \geq 5$, and $\dim(\mathrm{sing} \Xi) \leq p - 5$, then $\dim \varphi^{-1}(\mathrm{sing} \Xi) \leq p - 2$.

(ii) If $g \geq 6$, and $\dim(\mathrm{sing} \Xi) = p - 6$, then $\dim \varphi^{-1}(\mathrm{sing} \Xi) \leq p - 3$.

Proof. (i) If $g \geq 5$, $\dim(\mathrm{sing} \Xi) \leq p - 5$, and if there were a component $Z$ of $\varphi^{-1}(\mathrm{sing} \Xi)$ of dimension $p - 1$, then the generic fiber dimension of $\varphi$ on $Z$ must be $\geq 4$, hence $\geq 5$, (since all fibers are odd dimensional). But by Proposition 3.1 there would be a subvariety $Z'$ of $X$ of the same dimension as $Z$, such that for $D'$ general on $Z'$, we have $|D'| \cong \mathbb{P}^{r-2}$. Then $r - 2 \geq 3$ implies that $\varphi(Z') \subset \mathrm{sing} \Xi$ also, but $\dim Z' = \dim Z$ implies that $Z'$ is also a component of $\varphi^{-1}(\mathrm{sing} \Xi)$, a contradiction.

(ii) The same proof works again. □

Proposition 3.1 will be proved in Lemmas 3.6 through 3.10.

Lemma 3.6. If $L$ is a line bundle on a curve $\tilde{C}$ such that $\dim |L| \geq 2$, then there exists a divisor $D$ in $|L|$ of form $D = E + p + q$, with $p$ and $q$ each occurring simply in $D$, and such that $p$ is not a base point of $|L|$ and $q$ is not a base point of $|L - p|$.

Proof. If $\dim |L| = r \geq 2$, and $B_1$ is the base divisor of $|L|$, choose (by Bertini) a divisor $D_1$ in $|L - B_1|$ such that (i) $D_1$ consists of distinct points and (ii) $\text{supp}(D_1) \cap \text{supp}(B_1) = \emptyset$, and let $p$ be any point of $D_1$. Then $p$ is not a base point of $|L|$ so $\dim |L - p| = r - 1 \geq 1$. Since $p$ does not belong to the divisor $B_1 + D_1 - p$ of $|L - p|$, then $p$ does not belong to the base divisor $B_2$ of $|L - p|$. Then choose a divisor $D_2$ in $|L - p - B_2|$ consisting of distinct points and such that $\text{supp}(D_2) \cap \text{supp}(B_2 + p) = \emptyset$, and let $q$ be any point of $D_2$. Then $D = p + B_2 + D_2 = p + q + E$ satisfies the requirements of the Lemma. □

Lemma 3.7. Let $\tilde{C}$ be a curve, let $d \geq 2$ be an integer, and define $\mathcal{D}$ as follows: $\mathcal{D} = \{(p, q, D) : D \geq p + q\} \subset \tilde{C} \times \tilde{C} \times \tilde{C}^{(d)}$. Then the projection $\mathcal{D} \to \tilde{C}^{(d)}$ is a finite map of degree $d(d - 1)$ étale at $(p, q, D)$ if $p$ and $q$ occur simply in $D$. 
Proof. Since $\tilde{C}$ is assumed complete the map is proper and quasi finite, hence finite, of the stated degree. If $p, q$ occur simply in $D = E + p + q$, the addition map $\tilde{C} \times \tilde{C} \times \tilde{C}^{(d-2)} \rightarrow C^{(d)}$ is étale at $(p, q, E)$, so it suffices to show the map $\mathcal{D} \rightarrow \tilde{C} \times \tilde{C} \times \tilde{C}^{(d-2)}$ taking $(x, y, F)$ to $(x, y, F - x - y)$ is étale at $(p, q, D)$. But the map $\tilde{C} \times \tilde{C} \times \tilde{C}^{(d-2)} \rightarrow \mathcal{D}$ taking $(x, y, H)$ to $(x, y, H + x + y)$ is a local analytic inverse from an analytic neighborhood of $(p, q, E)$ to an analytic neighborhood of $(p, q, D)$. (Since the varieties are smooth any local analytic bijection is a local analytic isomorphism.)  

Lemma 3.8. Let $\tilde{C}$ and $\mathcal{D}$ be as in Lemma 3.7 and assume further that $\tilde{C}$ has a fixed point free involution $\iota$. If $Z \subset \mathcal{D}$ is any subvariety, define $Z' = \{(p', q', D')\} \text{ for all } (p, q, D) \in Z = \text{the “flip” of } Z$, where $p' = \iota(p), q' = \iota(q)$, and $D' = D - p - q + p' + q'$. Then $\dim(Z) = \dim(Z')$.

Proof. Since the map $\tilde{C} \times \tilde{C} \times \tilde{C}^{(d-2)} \rightarrow \mathcal{D}$ taking $(x, y, H)$ to $(x, y, H + x + y)$ is an analytic bijection, it preserves the dimension of subvarieties, so it suffices to check that in $\tilde{C} \times \tilde{C} \times \tilde{C}^{(d-2)}$ the map taking $(x, y, H)$ to $(x', y', H)$ preserves the dimension of subvarieties. Since the map is a regular involution, hence an isomorphism, it does indeed preserve dimension.

Lemma 3.9. Let $\pi: \tilde{C} \rightarrow C$ be any connected étale double cover of a smooth curve $C$, with associated Prym theta divisor $\Xi$, and Abel map $\varphi: X \rightarrow \Xi$. If $Z$ is an irreducible component of $\varphi^{-1}(\text{sing } \Xi)$, and $L$ in $\varphi(Z)$ is a general point of the image of $Z$, then $|L| = \varphi^{-1}(L) \subset Z_{sm}$, and $Z$ is the only component of $\varphi^{-1}(\text{sing } \Xi)$ which dominates $\varphi(Z)$.

Proof. Let $U \subset \varphi(Z)_{sm}$ be the (irreducible) open subset of $\varphi(Z)_{sm} \subset \text{sing } \Xi$ on which $|L|$ is minimal. Then over $U$, the map $\varphi^{-1}(U) \rightarrow U$ is a locally trivial projective bundle over a smooth irreducible base, hence $\varphi^{-1}(U)$ is a smooth irreducible subset of $\varphi^{-1}(\text{sing } \Xi)$ containing an open, dense, subset of $Z$. Thus $Z \subset \text{cl}(\varphi^{-1}(U))$, and since $Z$ is a maximal irreducible subset of $\varphi^{-1}(\text{sing } \Xi)$, we must have $Z = \text{cl}(\varphi^{-1}(U))$. This proves both statements of the lemma. I.e., for $L$ in $U$, $\varphi^{-1}(L) = |L| \subset \varphi^{-1}(U) \subset Z_{sm}$. Moreover we have shown that any component of $\varphi^{-1}(\text{sing } \Xi)$ dominating $\varphi(Z)$ equals $\text{cl}(\varphi^{-1}(U))$. Indeed this proof shows that for every subvariety $W \subset \text{sing } \Xi$, exactly one component of $\varphi^{-1}(W)$ dominates $W$.

Now we can deduce Proposition 3.1.

Lemma 3.10. Let $Z \subset \varphi^{-1}(\text{sing } \Xi)$ be any component of $\varphi^{-1}(\text{sing } \Xi)$ on which the generic fiber of $\varphi$ is $\cong \mathbb{P}^r$, $r \geq 3$. Then there exists $Z' \subset X$ such that $Z'$ is irreducible, $\dim Z' = \dim Z$, and such that for $D'$ general on $Z'$, we have $|D'| \cong \mathbb{P}^{r-2}$, and $\dim(\varphi(Z')) \geq \dim(\varphi(Z)) + 2$.

Proof. Choose $L$ a general point of $\varphi(Z)$ and using Lemma 3.6, choose $D_1$ in $|L| \subset Z_{sm}$ of form $p_1 + q_1 + E_1 = D_1$, where $p_1$ and $q_1$ occur simply in
$|\mathcal{L}|$, $p_1$ is not a base point of $|\mathcal{L}|$ and $q_1$ not a base point of $|\mathcal{L} - p_1|$. Define $\tilde{Z} = \{(p, q, D) : D \geq p + q, \text{ and } D \text{ belongs to } Z\} \subset \tilde{X} = \{(p, q, D) : D \geq p + q, \text{ and } D \text{ belongs to } X\}$. Then $\tilde{Z} \to Z$ is finite and étale at $(p_1, q_1, D_1)$. Thus since $D_1$ is in $|\mathcal{L}| \subset Z_{sm}$, $\tilde{Z}$ is also smooth at $(p_1, q_1, D_1)$. Hence there is a unique component $\tilde{Z}_1$ of $\tilde{Z}$ containing the point $(p_1, q_1, D_1)$. Now “flip” $\tilde{Z}_1$ as in Lemma 3.8, to $\tilde{Z}'_1 = \{(p', q', D')\}$ for all $(p, q, D)$ in $\tilde{Z}_1$ and let $Z' = \text{image of } \tilde{Z}'_1$ under projection to $X$. Since by Mumford’s parity trick, [M2, p. 188, step III], replacing two points changes the parity of $D$ twice, hence leaves it even, $\tilde{Z}_1$ is contained in $\tilde{X}$. Hence the flipped set $Z'$ lies in $X$. Then $D'_1 \in Z'$ and $|D'_1| \cong \mathbb{P}^{r-2}$, by the choice of $p_1$ and $q_1$ in $D_1$. So since no divisor in $\tilde{Z}_1$ can have its dimension lowered by more than 2 through flipping 2 points, thus for $D'$ generic in $Z'$, we have $|D'| \cong \mathbb{P}^{r-2}$. Since the fibers of $\varphi$ are contained in complete linear series, hence the restricted Abel map $\varphi : Z' \to \Xi$ has generic fiber dimension $\leq r - 2$. Thus $\dim(\varphi(Z')) \geq \dim(\varphi(Z)) + 2$. □

Remarks 3.11. (i) When $C$ is nonhyperelliptic, the conclusion of Corollary 3.3 that $h^0(\tilde{C}, L) = 4$ for a general stable singularity $L$ on $\Xi$, holds vacuously when $3 \leq g(C) \leq 4$ since then $\text{sing}_{\text{st}} \Xi = \emptyset$. The equation $h^0(\tilde{C}, L) = 4$ holds for all points $L$ on $\text{sing}_{\text{st}} \Xi$ when $g(C) = 5$ by the remarks above Proposition 3.1. Welters’ argument [We1, Lemma 3.2, p. 681], for the existence of at least one point $L$ on $\text{sing}_{\text{st}} \Xi$ with $h^0(\tilde{C}, L) = 4$ whenever $\text{sing}_{\text{st}} \Xi \neq \emptyset$, already implies the conclusion of Corollary 3.3 for any double cover such that $\text{sing}_{\text{st}} \Xi$ is irreducible.

(ii) Since it is known (see Proposition 5.1 below) that the tangent cone to $\Xi$ at a stable double point contains the Prym canonical model $\varphi_\eta(C)$ of $C$, Corollary 3.4 provides as many such quadrics as possible for nontetragonal curves with $g(C) \geq 11$. I.e., then for all $L$ in a dense open subset of $\text{sing}_{\text{st}} \Xi, \mathbb{P}C_L(\Xi)$ is a quadric such that $\varphi_\eta(C) \subset \mathbb{P}C_L(\Xi)$. Considering the results of [LS], a primary open question concerning Donagi’s conjecture then is whether in this case these quadrics generate the space of all quadrics containing $\varphi_\eta(C)$.

For tetragonal curves, Debarre [D1] has shown that the Prym variety of a generic tetragonal curve of genus $g \geq 13$ arises as Prym variety of exactly three doubly covered curves $C_i$, all tetragonal. Further, $\text{sing} \Xi$ has dimension $p - 6$ and has 3 components of that dimension, and the generic point of each component is nonexceptional for the representation of $P$ as a tetragonal Prym associated to exactly two of the $C_i$. Hence Corollary 3.3 implies that the generic singular point on any one of the three components is a double point, and it follows that the base locus of the quadric tangent cones to any one of these components must contain the Prym canonical models of the two curves $C_i$ for which this is a stable, nonexceptional component. It is conceivable that the quadric tangent cones at double points of the union
of two of these components determines the unique curve \( C_i \) for which both these components are stable.

(iii) The conclusion of Corollary 3.5(i) that \( \dim \varphi^{-1}(\text{sing} \Xi) \leq p - 2 \), holds also for all nonhyperelliptic curves \( C \) with \( g = 3, 4 \); for \( g = 3 \), it holds since \( \text{sing} \Xi = \emptyset \) and \( \dim X = p = 2 \), and for \( g(C) = 4 \) it holds by irreducibility of \( X \), since \( \dim X = 3 \) and all fibers of \( \varphi \) are odd dimensional. For \( g = 5 \), Corollary 3.5(i) holds since then \( \dim(\text{sing} \Xi) \leq p - 5 \) implies \( \text{sing} \Xi = \emptyset \), while \( \dim X = 4 \). The hypothesis is necessary here however since by Example 2.18, there is a doubly covered nonhyperelliptic curve \( C \) with \( g(C) = 5 \), \( \dim(\text{sing} \Xi) = 0 = p - 4 \), and \( h^0(L) = 4 \). Thus \( \dim \varphi^{-1}(L) = 3 = p - 1 \).

In Corollary 3.5(ii), the hypotheses cannot hold for \( g \leq 5 \) and the conclusion can fail as we have seen. When \( g = 6 \), the hypothesis is necessary since a plane quintic curve \( C \) with an odd double cover \( P, \Xi \) over \( C \) is an example of \( \Xi \) with \( \dim(\text{sing} \Xi) = 0 = p - 5 \), and \( \varphi^{-1}(\text{sing} \Xi) \cong \mathbb{P}^3 \) has dimension \( 3 = p - 2 \).

(iv) Corollary 3.5 is useful for comparing line bundles on \( X \) with pullbacks of line bundles from \( \Xi \). This will be applied in Section 4 to describe the fundamental line bundle \( \mathcal{O}_X(1) \) associated to the divisor variety \( X \). An open question concerning the relation of \( X \) to the Prym Torelli problem is to compute \( h^0(X, \mathcal{O}_X(1)) \).

(v) We do not know, even when \( \dim(\text{sing} \Xi) \leq p - 5 \), whether any components \( Z \) of \( \varphi^{-1}(\text{sing} \Xi) \) exist that do not dominate components of \( \text{sing} \Xi \). In particular we do not know whether there exist any components \( Z \) of \( \varphi^{-1}(\text{sing} \Xi) \) on which the generic fiber dimension of \( \varphi \) is \( \geq 5 \).

4. A formula for the line bundle \( \mathcal{O}_X(1) \) defined by the norm map \( h : X \to \lvert \omega_C \rvert \).

Recall that if \( P \) is the Prym variety associated to a smooth connected étale double cover \( \tilde{C} \to C \), and \( g = g(C) \), then \( p = \dim(P) = \dim(X) = g - 1 \), where \( \varphi : X \to \Xi \) is the restriction of the Abel map \( \tilde{\alpha} : \tilde{C}^{(2g-2)} \to \text{Pic}^{2g-2}(\tilde{C}) \) over \( \Xi \subset P \subset \text{Pic}^{2g-2}(\tilde{C}) \), and \( \dim(\Xi) = p - 1 = g - 2 \). The restriction to \( X \) of the norm map \( \text{Nm} : \tilde{C}^{(2g-2)} \to C^{(2g-2)} \) is denoted \( h : X \to \lvert \omega_C \rvert \), and the associated line bundle \( h^*(\mathcal{O}_{\lvert \omega_C \rvert}(1)) \) is denoted \( \mathcal{O}_X(1) \). We will show with mild genericity hypotheses that this line bundle is obtained from the pullback \( \varphi^*(K_\Xi) \) of the canonical bundle on \( \Xi \) by twisting with the “tangent bundle along the fibers of \( \varphi \)”. We must first give a definition of this relative tangent sheaf for our present situation in which \( \varphi : X \to \Xi \) is not necessarily a \( \mathbb{P}^1 \)-bundle over \( \text{sing}(\Xi) \).
Definition 4.1. Given $\varphi : X \to \Xi$ as above, define $T_\varphi$ on all of $X$ to be the coherent sheaf $T_\varphi = \text{Hom}(\Omega^1_\varphi, O_X) =$ the dual $O_X$-module of the relative Kähler differentials $\Omega^1_\varphi$.

Note that on the open set $U = \varphi^{-1}(\Xi_{sm})$, the restriction $U \to \Xi_{sm}$ is a Zariski locally trivial $\mathbb{P}^1$ bundle, hence $T_\varphi$ is the intuitive relative tangent bundle at least on $U$; in particular its restriction to $U$ is a subbundle of $T_X$ whose restriction to each fiber of $\varphi : U \to \Xi_{sm}$ is the tangent bundle to the fiber.

Theorem 4.2. Assume $C$ is nonhyperelliptic, $\pi : \tilde{C} \to C$ any étale connected double cover and $(P, \Xi)$ the Prym variety. If $g(C) = 3$ or 4, or if $g(C) \geq 5$ and $\text{dim}(\text{sing}(\Xi)) \leq g - 6$, then $O_X(1) \cong T_\varphi \otimes \varphi^*(K_{\Xi})$, where $O_X(1)$ is the line bundle associated to the norm map $h : X \to |\omega_C|$.

Corollary 4.3. (i) Under the hypotheses of Theorem 4.2, the line bundle $O_X(1)$ is intrinsically defined on $X$, i.e., $O_X(1)$ is determined by $X$ as an abstract variety.

(ii) Under the hypotheses of Theorem 4.2, the sheaf $T_\varphi = \text{Hom}(\Omega^1_\varphi, O_X)$ is the unique line bundle on $X$ which on $X - \varphi^{-1}(\text{sing}(\Xi))$ equals the bundle of tangents along the fibers of $\varphi$; in particular the hypotheses of Theorem 4.2 imply that $T_\varphi$ is a line bundle on all of $X$.

Proof of Corollary 4.3(i). First we will show the map $\varphi$ is determined intrinsically by $X$.

Claim. Two points of $X$ lie in the same fiber of $\varphi$ if and only if they can be joined by a smooth rational curve $\lambda$ on $X$.

Since the fibers are projective spaces any two points in the same fiber are joined by a curve isomorphic to $\mathbb{P}^1$. Moreover since $P$ contains no rational curves, every smooth rational curve on $X$ is collapsed to a point by $\varphi$, hence lies in some fiber of $\varphi$. Thus two points which are joined by a smooth rational curve do lie in the same fiber of $\varphi$. QED for the Claim.

(Note that $\varphi$ can be regarded as the extremal contraction $\text{cont}_R$ defined by any smooth rational curve on $X$. I.e., if $\lambda$ is a general fiber of $\varphi$ on $X$, then $\varphi$ induces an exact sequence in homology $0 \to \mathbb{R}[\lambda] \to H_2(X, \mathbb{R}) \to H_2(\Xi, \mathbb{R}) \to 0$, by Leray s.s.)

Thus the fibers of $\varphi$ are characterized by $X$. Since $\Xi = \varphi(X)$ is normal, we claim $\Xi$ is characterized as a scheme by the fibers of $\varphi$ in $X$. First $\Xi$ has the quotient topology induced by $\varphi$, since $\varphi$ is proper, so $\Xi$ is determined as a topological space. Then since $\Xi$ is normal, $O_\Xi = \varphi_*(O_X)$, so the regular functions on open subsets $U$ of $\Xi$ are functions on $\varphi^{-1}(U)$ which are constant on the fibers of $\varphi$. Hence $O_\Xi$ and the fibers of $\varphi$ are determined by $X$. Since $X$ determines $\varphi : X \to \Xi$, by Theorem 4.2 $X$ determines $O_X(1)$.  

□
Proof of Corollary 4.3(ii). Since $\mathcal{O}_X(1) \cong T_{\varphi} \otimes \varphi^*(K_{\Xi})$, $T_{\varphi} = \mathcal{O}_X(1) \otimes (\varphi^*(K_{\Xi}))^*$ is the tensor product of two line bundles. QED Corollary 4.3.

Proof of Theorem 4.2. If $g = 3$, the Prym is a 2 dimensional Jacobian and $\varphi : X \to \Xi$ is a $\mathbb{P}^1$ bundle over a smooth genus 2 curve, and in this case the formula has been proved in [SV4, p. 358]. If $g \geq 4$, we claim it suffices, by a “Hartogs” argument, to show that $\mathcal{O}_X(1)$ and $T_{\varphi} \otimes \varphi^*(K_{\Xi})$ are isomorphic on the open subset $U = \varphi^{-1}(\Xi_{sm})$ of $X$.

Definition 4.4. We say a sheaf $\mathcal{F}$ on an irreducible scheme $X$ has the “Hartogs property” if its sections extend uniquely across closed sets of codimension $\geq 2$. I.e., if for every closed subset $Z \subset X$ all of whose components are of codimension $\geq 2$ in $X$, and every open set $V \subset X$, the restriction $H^0(V, \mathcal{F}) \to H^0(V - (Z \cap V), \mathcal{F})$ is an isomorphism.

Lemma 4.5. If $X$ is an irreducible Cohen Macaulay scheme and $\mathcal{F}$ is a coherent sheaf of $\mathcal{O}_X$ modules, then the sheaf $\mathcal{F}^* = \mathcal{H}om(\mathcal{F}, \mathcal{O}_X)$ has the Hartogs property. (In particular $\mathcal{O}_X$ itself, all locally free $\mathcal{O}_X$ modules, and all “reflexive” $\mathcal{O}_X$ modules, have the Hartogs property on a Cohen Macaulay variety.)

Proof. This follows from some properties of depth, which we recall.

(i) If $\mathcal{F}$ is a coherent sheaf on a scheme $X$ and $Z \subset X$ is a closed subset, then local sections of $\mathcal{F}$ extend uniquely across $Z$ if and only if $\mathcal{F}$ has depth $\geq 2$ along $Z$, ([Gro, Prop. 1.11, pp. 11-12, Thm. 3.8, p. 44] or see [SV5, Prop. 18, p. 391], for a summary statement).

(ii) If $X$ is an algebraic scheme, and $Z \subset X$ a closed subset, such that $\mathcal{O}_X$ has depth $\geq 2$ along $Z$, then for any coherent sheaf $\mathcal{F}$ of $\mathcal{O}_X$ modules, the coherent sheaf $\mathcal{F}^* = \mathcal{H}om(\mathcal{F}, \mathcal{O}_X)$ also has depth $\geq 2$ along $Z$, ([SV5, Lemma 22, p. 392; similar to lemma, p. 21, of [S]]).

(iii) If $X$ is an irreducible noetherian Cohen Macaulay scheme, and $Z \subset X$ a closed subset, then $\mathcal{O}_X$ has depth $\geq k$ along $Z$ if and only if every irreducible component of $Z$ has codimension $\geq k$ in $X$ [H, p. 184]. QED for Lemma 4.5.

Since when $C$ is nonhyperelliptic $X$ is an irreducible normal local complete intersection, in particular Cohen Macaulay, and since $\mathcal{O}_X(1)$ and $\varphi^*(K_{\Xi})$ are line bundles on $X$, it follows from Lemma 4.5 that both $\mathcal{O}_X(1)$ and $T_{\varphi} \otimes \varphi^*(K_{\Xi})$ have the Hartogs property on $X$. If $g = 4$, then $(P, \Xi)$ is a 3 dimensional Jacobian hence $\Xi$ is singular only when $(P, \Xi)$ is a hyperelliptic Jacobian and then $\Xi$ has one singular point. Then since $X$ is irreducible of dimension 3 and every fiber of $\varphi$ is an odd dimensional projective space, $\varphi^{-1}(\text{sing}\Xi) \cong \mathbb{P}^1$ hence has codimension two in $X$. Now using Corollary 3.5(i), under the hypotheses of Theorem 4.2, also if $g \geq 4$ then
$Z = \varphi^{-1}(\text{sing}(\Xi))$ has codimension at least 2 in $X$. Hence if we have an isomorphism between $\mathcal{O}_X(1) \otimes (\varphi^*(K\Xi))^*$ and $\mathcal{T}_\varphi$ over $X - Z = U = \varphi^{-1}(\Xi_{sm})$, the isomorphism extends uniquely to an isomorphism over all of $X$.

Thus, from now on we will primarily consider $U$ and the map $\varphi : U \rightarrow \Xi_{sm}$. We claim that on $U$, $\mathcal{O}_X(1) \otimes (\mathcal{T}_\varphi)^*$ is the pullback of a line bundle on $\Xi_{sm}$. To see this, take any point $z$ in $\Xi_{sm}$ and consider the preimage $\varphi^{-1}(z) \cong \mathbb{P}^1$; we will check that the restrictions of $\mathcal{O}_X(1)$ and $\mathcal{T}_\varphi$ to $\varphi^{-1}(z)$ are both line bundles of degree two. For $\mathcal{O}_X(1)|\varphi^{-1}(z)$, consider the map $h : X \rightarrow |\omega_C|$ followed by the injective linear map of projective spaces $\pi^* : |\omega_C| \rightarrow |\omega_C|$; then the degree of this composition will equal the degree of the restriction of $h$ to $\varphi^{-1}(z)$, i.e., the degree of $\mathcal{O}_X(1)|\varphi^{-1}(z)$. The composition $X \rightarrow |\omega_C|$ is given by $D \mapsto D + \iota^*(D)$, where $\iota : C \rightarrow \tilde{C}$ is the involution. On a line $\varphi^{-1}(z) = \{D_z\} = \mathbb{P}H^0(\tilde{C}, L_z)$, this map is induced by the map $H^0(\tilde{C}, L_z) \rightarrow H^0(\tilde{C}, \omega_{\tilde{C}})$ on sections: $s \mapsto s \otimes \iota^*(s)$, where $s \in H^0(\tilde{C}, L_z)$, $\iota^*(s) \in H^0(\tilde{C}, \iota^*(L_z))$, and $L_z \otimes \iota^*(L_z) \cong \pi^*(\text{Nm}(L_z)) \cong \pi^*(\omega_C) \cong \omega_{\tilde{C}}$; thus the map $\varphi^{-1}(z) \cong \mathbb{P}H^0(\tilde{C}, L_z) \rightarrow \mathbb{P}H^0(\tilde{C}, \omega_{\tilde{C}}) = |\omega_{\tilde{C}}|$ is homogeneous of degree 2, hence is given on $\varphi^{-1}(z) \cong \mathbb{P}^1$ by sections of $\mathcal{O}_{\mathbb{P}^1}(2)$. On the other hand, the restriction $(\mathcal{T}_\varphi)|\varphi^{-1}(z)$ is the tangent bundle of $\varphi^{-1}(z) \cong \mathbb{P}^1$ which has degree 2. Thus both $\mathcal{O}_X(1)$ and $\mathcal{T}_\varphi$ restrict to the line bundle $\mathcal{O}(2)$ on each fiber $\varphi^{-1}(z) \cong \mathbb{P}^1$ (since $z \in \Xi_{sm}$) and hence the line bundle $\mathcal{O}_X(1) \otimes (\mathcal{T}_\varphi)^*$ on $U$ is trivial on each fiber of $\varphi$.

It follows that there exists a line bundle, say $\mathcal{M}$, on $\Xi_{sm}$ such that $\mathcal{O}_X(1) \otimes (\mathcal{T}_\varphi)^* \cong \varphi^*(\mathcal{M})$. Indeed, if we set $\mathcal{M} = \varphi_*(\mathcal{O}_X(1) \otimes (\mathcal{T}_\varphi)^*)$, then $\mathcal{M}$ is a line bundle on $\Xi_{sm}$ by Grauert’s theorem [H, Cor. 12.9, p. 288], since $\mathcal{F} = \mathcal{O}_X(1) \otimes (\mathcal{T}_\varphi)^*$ is flat over $\Xi_{sm}$ and for each $z \in \Xi_{sm}$, $H^0(\varphi^{-1}(z), (\mathcal{O}_X(1) \otimes (\mathcal{T}_\varphi)^*))|\varphi^{-1}(z)) = H^0(\varphi^{-1}(z), \mathcal{O}_{\varphi^{-1}(z)}) = 1$. Then the natural homomorphism of line bundles on $U$, $\varphi^*(\mathcal{M}) = \varphi^*(\varphi_*(\mathcal{O}_X(1) \otimes (\mathcal{T}_\varphi)^*)) \rightarrow \mathcal{O}_X(1) \otimes (\mathcal{T}_\varphi)^*$ is an isomorphism since it is evidently an isomorphism on each fiber.

It remains to show that the line bundle $\mathcal{M}$ on $\Xi_{sm}$ is isomorphic to $K\Xi_{sm}$. For this, we will show how to express divisors in both series $|\mathcal{O}_X(1)|$ and $|K\Xi|$ in terms of the “standard divisors” $\{D_p\}$ on $X$. Recall that for any point $p$ on $\tilde{C}$, the divisor $D_p = \{\text{those } D \in X \text{ such that } D \geq p\}$. Then for all points $\overline{p}$ in $C$, if $H_{\overline{p}} \subset |\omega_C|$ is the hyperplane of $|\omega_C| \cong (\mathbb{P}^{g-1})^*$ corresponding to the point $\varphi_\omega(\overline{p})$ of $|\omega_C|^{*}$ on the canonical curve $\varphi_\omega(C) \subset |\omega_C|^{*} = \mathbb{P}^{g-1}$, and if $\pi^{-1}(\overline{p}) = \{p, p'\}$, then $h^{-1}(H_{\overline{p}}) = D_p + D_{p'}$. Thus $\mathcal{O}_X(1) \cong \mathcal{O}_X(D_p + D_{p'})$. Now consider a general point $p \in \tilde{C}$; then one knows ([SV2]) the following: $D_p \subset X$ is irreducible, maps birationally onto $\Xi$, and the formula $\varphi(D_p \cap D_{p'}) = \Gamma_{\overline{p}}$ holds, where $\Gamma_{\overline{p}}$ is the Gauss divisor on $\Xi$ defined by the Prym canonical image $\varphi_\omega(\overline{p})$ of the point $\overline{p} = \pi(p) \in C$. Since $\varphi(D_p \cap D_{p'}) = \Gamma_{\overline{p}} \in |\mathcal{O}_\Xi(\Xi)|$, and $\mathcal{O}_\Xi(\Xi) \cong K\Xi$ by
adjunction, we see that the divisor \( \varphi(D_p \cap D_{p'}) \) on \( \Xi \) is a canonical divisor, i.e., \( \mathcal{O}_\Xi(\varphi(D_p \cap D_{p'}) \cong K_\Xi \). Let \( F \subset \Xi \) be the locus of points over which \( \varphi : D_p \rightarrow \Xi \) is not an isomorphism; since \( D_p \rightarrow \Xi \) is birational and \( \Xi \) is normal, \( \text{codim}_\Xi(F) \geq 2 \).

Now \( D_p \subset X \) provides a section of the \( \mathbb{P}^1 \)-bundle \( \varphi : U \rightarrow \Xi_{sm} \) over \( V \subset \Xi_{sm} \), and therefore \( N(D_p/U)(D_p \cap \varphi^{-1}(V)) \cong T_p(D_p \cap \varphi^{-1}(V)) \). I.e., at a point \( x \) of \( D_p \) in \( \varphi^{-1}(V) \), \( D_p \) is transverse to the fiber \( \varphi^{-1}(z) \), where \( z = \varphi(x) \), so the tangent space \( T_x(\varphi^{-1}(z)) \subset T_x(U) \) maps isomorphically onto the normal space \( N_x(D_p) = T_x(U)/T_x(D_p) \). (Note that although the bundles \( T_\varphi \) and \( \mathcal{O}_X(D_p) \) are different both on \( X \) and on the open set \( U \subset X \), indeed they have different restrictions to fibers of \( \varphi \), they have the same restrictions to the section \( (D_p \cap \varphi^{-1}(V)) \) over \( V \).) The normal bundle \( N(D_p/U) \) is the restriction \( (D_p \cap U) \) of \( \mathcal{O}_{D_p}(D_p) = \mathcal{O}_X(D_p)|D_p \), so in \( \varphi^{-1}(V) \), \( (T_\varphi)D_p \cong \mathcal{O}_{D_p}(D_p) \) on \( \text{codim}_\Xi(F) \geq 2 \).

The normal bundle \( N(D_p/U) \) is the restriction \( (D_p \cap U) \) of \( \mathcal{O}_{D_p}(D_p) = \mathcal{O}_X(D_p)|D_p \), so in \( \varphi^{-1}(V) \), \( (T_\varphi)D_p \cong \mathcal{O}_{D_p}(D_p) \) on \( \text{codim}_\Xi(F) \geq 2 \).

Then \( \text{dim}(\text{sing}\Xi) = g - 7 \), at least for \( g(C) \geq 11 \). Since Debarre [D1] gave a list of those tetragonal curves with \( \text{dim}(\text{sing}\Xi) = g - 6 \), this would give a good account of the dimension of \( \text{sing}\Xi \).

(i) The formula, \( \mathcal{O}_X(1) \cong T_\varphi \otimes \varphi^*(K_\Xi) \), gives a simple way to think of \( \mathcal{O}_X(1) \) in terms of the canonical bundle \( K_X \). Namely, consider in general a \( \mathbb{P}^1 \)-bundle \( \varphi : X \rightarrow \Xi \) over a variety \( \Xi \) of general type; then \( K_X \) would have \( (\text{additive notation}) \) the form \( \Omega_\varphi^1 + \varphi^*(K_\Xi) \) and the relative canonical bundle \( \Omega_\varphi^1 \) is negative on the fibers of \( \varphi \), so if one “changes the sign of \( K_X \) along the fibers” (i.e., replaces \( \Omega_\varphi^1 + \varphi^*(K_\Xi) \) by \( T_\varphi + \varphi^*(K_\Xi) \)), then one obtains an ample line bundle on \( X \) intrinsic to the \( \mathbb{P}^1 \)-bundle structure \( \varphi : X \rightarrow \Xi \). The proof of Theorem 4.2 given here generalizes the one in [SV4], and is our original proof of the \( \mathcal{O}_X(1) \) formula. The formula...
relating reducible divisors in \(|\mathcal{O}_X(1)|\) to Gauss divisors on \(\Xi\), needed for this generalization, is in [SV2].

(iii) It is possible to prove the result of Corollary 4.3(i), that the line bundle \(\mathcal{O}_X(1)\), is intrinsically defined by \(X\), without any hypotheses on \(\text{codim}(\varphi^{-1}(\text{sing}\Xi))\). In this generality, the line bundle \(K_X\) is more convenient to work with than the coherent sheaf \(\mathcal{H}om(\mathcal{O}_X, \mathcal{O}_X)\), and then (see [SV3]) we can compute \(K_X = 2\varphi^*(K_\Xi) - \mathcal{O}_X(1)\), which yields the formula \(\mathcal{O}_X(1) = 2\varphi^*(K_\Xi) - K_X\). Since \(\varphi\) is determined by \(X\) in general (another proof of this is also in [SV3]) this implies that \(X\) always determines the line bundle \(\mathcal{O}_X(1)\). Note that if \(\varphi : X \rightarrow \Xi\) were a \(\mathbb{P}^1\) bundle, so that we had \(K_X = \Omega^1_\varphi + \varphi^*(K_\Xi)\) and \(T_\varphi = \mathcal{H}om(\Omega^1_\varphi, \mathcal{O}_X)\), these sheaves would be line bundles, and this new version of the formula would be equivalent to the one proved above in Theorem 4.2. In particular, the formulas are always equivalent over \(X - \varphi^{-1}(\text{sing}\Xi)\) so that when \(\varphi^{-1}(\text{sing}\Xi)\) has codimension \(\geq 2\) in \(X\), by using the depth argument above to extend such an isomorphism, the more general formula in [SV3] gives another proof of Theorem 4.2.

(iv) Recently Izadi and Pauly (see [IP]) have given a proof of a formula for \(\mathcal{O}_{X^-}(1)\), where \(X^-\) is the “odd” half of the divisor variety for the Prym, analogous to the one in Theorem 4.2 above. The proof in [SV3] of the formula for \(\mathcal{O}_X(1)\) applies to \(X^-\) as well, hence gives a version of their formula. I.e., in that case we get again \(\mathcal{O}_{X^-}(1) = (\varphi^*(K_\Xi))^2 \otimes (K_{X^-})^*,\) but since \(\varphi : X^- \rightarrow P^-\) is birational onto a (smooth) abelian variety, a canonical divisor of \(X^-\) is the pullback of a canonical divisor of \(P^-\) plus a divisor \(E\) whose support is the exceptional divisor of \(\varphi : X^- \rightarrow P^-\), and thus one gets \(\mathcal{O}_X(1) = (\varphi^*(K_\Xi))^2 \otimes (\mathcal{O}_{X^-}(E))^*\). The proof in [IP] (Lemma 2.2, p. 6, [IP]) identifies the divisor \(E\) more precisely.

5. A proof of the Torelli theorem for cubic threefolds.

If \(W\) is a smooth cubic threefold, then associated to a general line \(\lambda\) on \(W\) there is a conic bundle representation of \(W\) and consequently a Prym representation of the intermediate Jacobian \((J(W), \Theta(W))\) as a Prym variety \((P, \Xi)\) associated to an “odd” double cover of a smooth plane quintic \(C\) [M1, pp. 347-8], [CG, App.], [B2], [T2]. Moreover \(\Xi\) has a unique singular point, a triple point at which the projective tangent cone is \(W\). This unpublished result of Mumford is treated particularly clearly in [B2]. Prym theory is used there to establish that there is only one singular point \(L\) and then the theory of the Abel Jacobi map on the Fano surface \(F\) of \(W\), in particular the “tangent bundle theorem” and the parametrization of \(\Xi\) by \(F \times F\), is used to deduce the multiplicity and the structure of the tangent cone of \(\Xi\) at \(L\). The following argument computes the structure of \(\Xi\) at its unique singular point, including the multiplicity and tangent cone, using the Abel parametrization \(\varphi : X \rightarrow \Xi\) (which exists for all Prym varieties), and the
explicit form of the Prym canonical map defined by a conic bundle structure
on a cubic threefold.

First we give a criterion for the tangent cone at a point of $\Xi$ to contain
the Prym canonical model $\varphi_\eta(C) \subset |\omega_C \otimes \eta|^*$ of $C$.

**Proposition 5.1.** Let $\pi : \widetilde{C} \to C$ be any connected étale double cover of a
smooth nonhyperelliptic curve $C$ of genus $g \geq 4$, $L$ any stable singular point
of $\Xi$ (possibly also exceptional), and $\text{mult}_L \Xi = r$. If $(1/2)h^0(\widetilde{C}, L) = r$,
i.e., if RST holds at $L$, or if $L$ is base point free and $(1/2)h^0(C, L) \leq r \leq h^0(\widetilde{C}, L) - 1$, then as sets $\varphi_\eta(C) \subset \mathbb{P}C_L(\Xi)$.

**Lemma 5.2.** If $p$ and $p'$ are conjugate points on the canonical model of $\widetilde{C}$
in $|\omega_{\widetilde{C}}|^*$, i.e., if the double cover $\pi : \widetilde{C} \to C$ maps $\pi(p) = \pi(p') = \overline{p}$, and if
$L_{p,p'}$ is the line in $|\omega_{\widetilde{C}}|^*$ joining $p$ to $p'$, and $\varphi_\eta : C \to \varphi_\eta(C) \subset |\omega_C \otimes \eta|^*$ is
the Prym canonical map, then $\varphi_\eta(\overline{p}) = L_{p,p'} \cap |\omega_C \otimes \eta|^*$.

**Proof of Lemma 5.2.** See [T1], p. 957, line 11, or [SV2], proof of part 1 of
main theorem, claim 1. \hfill $\Box$

**Proof of Proposition 5.1.** Lemma 5.2 implies $\varphi_\eta(C) \subset \text{Sec}(\varphi_{\kappa}(\widetilde{C})) \cap \mathbb{P}T_L P$.
Moreover $2\Xi = \widetilde{\Theta} \cdot P$ implies $\mathbb{P}C_L(\Xi) = \{\widetilde{\theta}_{2r} = 0\} \cap \mathbb{P}T_L P$ when $r = \text{mult}_L \Xi$.
Hence we have only to show that, under the hypotheses of Proposition 5.1,
the inclusion $\text{Sec}(\varphi_{\kappa}(\widetilde{C})) \cap \mathbb{P}T_L P \subset \{\widetilde{\theta}_{2r} = 0\} \cap \mathbb{P}T_L P$ holds. I.e., we are
assuming either $4 \leq \text{mult}_L \widetilde{\Theta} = 2r$, or $L$ is base point free and $4 \leq \text{mult}_L \Theta \leq 2r \leq 2\text{mult}_L \Theta - 2$. Hence, in the first case by [K, p. 183] and in the second
case by [ACGH, Thm. 1.6(ii), p. 232], we have $\text{Sec}(\varphi_{\kappa}(\widetilde{C})) \subset \{\widetilde{\theta}_{2r} = 0\}$, thus as desired we get $\varphi_\eta(C) \subset \text{Sec}(\varphi_{\kappa}(\widetilde{C})) \cap \mathbb{P}T_L P \subset \{\widetilde{\theta}_{2r} = 0\} \cap \mathbb{P}T_L P = \mathbb{P}C_L(\Xi)$.

**Remark 5.3.** Since RST holds at every stable double point of $\Xi$, if $C$ is
nontetragonal and $g(C) = g \geq 11$, then in light of Proposition 5.1 and
Corollary 3.4, for all $L$ in a dense open subset of $\text{sing}_s \Xi$, $\mathbb{P}C_L(\Xi)$ is a quadric
such that $\varphi_\eta(C) \subset \mathbb{P}C_L(\Xi)$.

**The Torelli theorem.** We assume the following facts about the double
cover representing $J(W)$ as a Prym variety, [B2], [B3]. For a general line $\lambda$
on $W$, the family $C$ of triangles on $W$ having $\lambda$ as one side, is a smooth curve
$C$ doubly covered by the smooth connected curve $\widetilde{C}$ of lines on $W$ distinct
from $\lambda$ but incident to $\lambda$. Associating each triangle to the plane it spans
embeds $C$ as a quintic curve in the $\mathbb{P}^2$ of planes containing $\lambda$ in $\mathbb{P}^4$, the double
cover $\widetilde{C} \to C$ is “odd” in the sense that $h^0(C, H \otimes \eta)$ is odd, where $H = \mathcal{O}(1)$
defines the plane embedding of $C$ and $\eta$ defines the double cover, and the
Prym variety $(P, \Xi)$ associated to the double cover $\widetilde{C} \to C$ is isomorphic to
$(J(W), \Theta(W))$. The Prym canonical map $\varphi_{\omega \otimes \eta} : C \to \mathbb{P}^4$ takes a point of
Lemma 5.4. There is exactly one exceptional singularity on \( \Xi \), \( L = \pi^*(\mathcal{O}_C(1)) \), which is also stable and at which \( h^0(\overline{C}, L) = 4 \).

Proof (cf. [B2]). An exceptional singularity on \( \Xi \) is a line bundle \( L \) of the form \( L = \pi^*(M)(B) \) where \( M \) is a line bundle on \( C \) with \( h^0(M) \geq 2 \) and \( B \geq 0 \) is an effective divisor on \( \overline{C} \), and where \( h^0(\overline{C}, L) \) is even and \( \text{Nm}(L) = \omega_C \). In particular \( M \) is an effective line bundle on \( C \) with \( \text{deg}(\omega_C - 2M) = \mathcal{O}(\pi(B)) \) also effective, hence \( \text{deg}(M) \leq 5 \). Since \( C \) is a smooth plane quintic, \( C \) is neither hyperelliptic nor trigonal so \( \text{deg}(M) = 4 \) or 5, and the canonical series \( \mathcal{K}_C = \mathcal{O}_C(2) \) is cut out on \( C \) by conics. If \( \text{deg}(M) = 4 \) then \( h^0(C, M) \leq 2 \) by Clifford so \( h^0(M) = 2 \). Since \( C \) is not trigonal \( M \) has no base point and thus there is a divisor \( D \) of 4 distinct points in \( |M| \). By RRT, \( h^0(C, K - M) = 3 \) and hence there are 3 independent conics passing through \( D \). Then the 4 points of \( D \) fail by one to impose independent conditions on conics so all 4 points of \( D \) lie on a line and \( |O(1) - D| \) is effective, i.e., \( M = \mathcal{O}(1)(-p) \) for some point \( p \) on \( C \). If \( \text{deg}(M) = 5 \), then \( M \) has at most one base point and again there is a divisor \( D \) in \( |M| \) with 5 distinct points and \( 2 \leq h^0(M) \leq 3 \) by Clifford, implies \( h^0(K - M) \geq 2 \). Again at least 4 points of \( M \) lie on a line and we have either \( M = \mathcal{O}(1) \) or \( M = \mathcal{O}(1)(p - \overline{q}) \) where \( p, \overline{q} \) are on \( C \). Thus if \( H = \mathcal{O}_C(1) \) the only possibilities for \( M \) are \( H, H - p \) or \( H + p - \overline{q} \), corresponding to series of form \( g_5^2, g_4^1 \), or \( g_5^1 \) on \( C \). The case \( M = H \) gives the one actual exceptional singularity on \( \Xi, L = \pi^*(H) \). This has \( h^0(\overline{C}, L) = 4 \), since \( h^0(\overline{C}, L) = h^0(\overline{C}, H) + h^0(\overline{C}, H \otimes \eta) \), \( h^0(C, H) = 3 \), and \( h^0(C, H \otimes \eta) \) is odd and less than 3, hence \( = 1 \). If \( M = H + p - \overline{q} \), with \( \overline{p} \neq \overline{q} \), then \( \text{deg}(M) = 5 \) so \( L = \pi^*(M) \) and \( \text{Nm}(L) = 2M = K \), so \( M \) is a theta characteristic. But for \( M = H + p - \overline{q} \) to be a theta characteristic, we must have \( K = 2H - 2p + 2\overline{q} = K - 2p + 2\overline{q} \), hence \( -2p + 2\overline{q} = 0 \), and our plane quintic would be hyperelliptic, a contradiction.

If \( M = H - p \) then we would have \( L = \pi^*(H - p)(B) \) and \( \pi_*(\pi^*(H - p)(B)) = 2(H - p)(\pi_* B) = K - 2p + \pi_*(B) = K \). Thus we would need \( B = 2p \) or \( 2p' \) or \( p + p' \). If \( B = p + p' \) then \( L = \pi^*(H) \), the singular point we already have. If \( B = 2p \), then the parity is opposite to that of \( \pi^*(H) \) by Mumford’s parity trick [M2, p. 186] so \( h^0(C, L) \) is not even. Hence there are no exceptional singularities on \( \Xi \) other than the one coming from \( L = \pi^*(H) \). \( \square \)

Lemma 5.5. The point \( L = \pi^*(H) \) is a triple point on \( \Xi \) such that \( \varphi_\eta(C) \subset \mathbb{P}C_L(\Xi) \).
Proof. Since both $H$ and $H \otimes \eta$ are odd theta characteristics on $C$ and $h^0(C, M) = 3$, it follows by [V, p. 948] that $L$ is a singular odd theta characteristic on $P$, in particular $\Xi$ has odd multiplicity $\geq 3$ at $L$. By Lemma 5.4 above and the Lemma on p. 345 of [M1], dim(sing$\Xi$) = 0, so the Prym variety $(P, \Xi)$ is not a polarized product of elliptic curves, hence by [SV1, p. 319], $\text{mult}_L \Xi \leq 4$, hence $\text{mult}_L \Xi = 3$. That $\varphi_\eta(C) \subset \mathbb{P}C_L(\Xi)$ then follows from Proposition 5.1 since $h^0(T_L P, \Xi) = 4$, and $L = \pi^*(H)$ is base point free since $H$ is.

Lemma 5.6. There are no nonexceptional singularities on $\Xi$.

Proof. By Theorem 2.1 the RST holds at every nonexceptional singularity of $\Xi$. Since the source space $X$ of the Abel map $\varphi : X \to \Xi$ is 5 dimensional and irreducible, the largest possible fiber of $\varphi$ is $\mathbb{P}^d$, so by Theorem 2.1 all nonexceptional singularities of $\Xi$ are stable double points. Thus the tangent cone at any such point is a quadric containing the Prym canonical curve $\varphi_\eta(C)$. The same argument proves this for every Prym representation of $J(W)$, i.e., for every choice of general line $\lambda$ on $W$, hence the tangent quadric at every nonexceptional singular point contains the Prym canonical model of every plane quintic $C_\lambda$ associated to every general line $\lambda$ on $W$. Since the Prym canonical model $\varphi_\eta(C_\lambda)$ is the locus of vertices of residual pairs of lines in all triangles lying on $W$ and having $\lambda$ as one side [B3, Remarque 6.27], the union of these Prym canonical curves is dense in $W$. Since the tangent cone at a double point cannot contain the smooth cubic hypersurface $W$, there are no double points on $\Xi$, and $L = \pi^*(H)$ is in fact the only singular point on $\Xi$.

Lemma 5.7. The theta divisor $\Xi$ has a unique singular point $L$, at which $\mathbb{P}C_L \Xi = W$.

Proof. We know the triple point $L = \pi^*(O_C(1))$ is the unique singular point on $\Xi$, and by the argument of Lemma 5.6 that $\mathbb{P}C_L \Xi$ contains the union of the Prym canonical curves $\varphi_\eta(C_\lambda)$ for every general line $\lambda$ on $W$. Hence $\mathbb{P}C_L \Xi \supset W$, and since these are both cubic hypersurfaces and $W$ is smooth, we conclude $\mathbb{P}C_L \Xi = W$.

This proves the Torelli theorem for $W$.

6. Outline of the RST and its corollaries.

Theorem.

(1) For all $L$ in $\Xi$, we have $\cup_{|L|} \mathbb{P} \varphi_\ast (T_D X) = \mathbb{P}C_L \tilde{\Theta} \cap \mathbb{P}T_L P$.

(2) For all $L$ such that $|L| \subset X_{sm}$, we have $\cup_{|L|} \mathbb{P} \varphi_\ast (T_D X) = \mathbb{P}C_L \Xi$.

(3) Sing$_{ex} \Xi = \varphi(\text{sing} X)$, hence for all $L$ in $\Xi - \text{sing}_e \Xi$, $|L| \subset X_{sm}$.

(4) For all $L$ in $\Xi - \text{sing}_e \Xi$, we have $\mathbb{P}C_L \Xi = \mathbb{P}C_L \tilde{\Theta} \cap \mathbb{P}T_L P$, and thus $\text{mult}_L \Xi = (1/2) h^0(C, L)$, i.e., “RST holds” at every $L$ in $\Xi - \text{sing}_e \Xi$. 

If \( g(C) \geq 11 \) and \( C \) not tetragonal, then for any double cover of \( C \), \( \operatorname{sing}_{2,\text{st}}(\Xi) \) is dense in \( \operatorname{sing}_{\text{st}}(\Xi) \).

If \( 2 \leq (1/2)h^0(\tilde{C}, L) = \operatorname{mult}_L \Xi \), i.e., if RST holds at \( L \) in \( \operatorname{sing} \Xi \), or if \( L \) is base point free and if \( \operatorname{mult}_L \Xi = r \), then \( \varphi^\eta(C) \subset \operatorname{Sec}(\varphi_K^{\eta}(\tilde{C})) \cap \mathbb{P}T_L P \subset \{ \partial_{2r} = 0 \} \cap \mathbb{P}T_L P = \mathbb{P}C_L(\Xi) \), as sets.

Cor: If \( g(C) \geq 11 \) and \( C \) not tetragonal, then for any double cover of \( C \), \( \varphi^\eta(C) \subset \mathbb{P}C_L(\Xi) \) for all \( L \) in a dense open subset of \( \operatorname{sing}_{\text{st}}(\Xi) \).

Cor: If \( W \) is a smooth cubic threefold and \((P, \Xi)\) the Prym variety associated to the odd cover of the discriminant plane quintic \( C \) for the conic bundle structure on \( W \) defined by any general line on \( W \), then \( \varphi^\eta(C) \subset \mathbb{P}C_L(\Xi) \), where \( L \) is the unique singular point on \( \Xi \). Since the union of these Prym canonical models is dense in \( W \), \( W \subset \mathbb{P}C_L(\Xi) \), and since by \([\text{SV1}, V]\) \( L \) is a triple point, \( W = \mathbb{P}C_L(\Xi) \).

Cor: The restricted norm map \( h : X \to |\omega_C| \) is defined by a linear subsystem of \( |O_X(1)| \), where \( O_X(1) \cong \mathcal{T}_\varphi \otimes \varphi^*(K_\Xi) \), and \( \mathcal{T}_\varphi \) is the bundle of tangents “along the fibers” of \( \varphi \), in case \( \dim(\operatorname{sing}(\Xi)) \leq p - 5 \).

References


A. Tjurin, *The geometry of the Poincaré theta divisor of a Prym variety*, Math. USSR Izvestija, **9**(5), (1975), 951-986;


A. Tjurin, *The geometry of the Fano surface of a nonsingular cubic $F \subset \mathbb{P}^4$ and Torelli theorems for Fano surfaces and cubics*, Math. USSR. Izvestija, **5** (1971), 517-546, MR 44 #2757, Zbl 252.14004.


Received December 30, 1999 and revised June 13, 2000.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF GEORGIA
ATHENS, GA 30602
E-mail address: roy@math.uga.edu

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF GEORGIA
ATHENS, GA 30602
E-mail address: rvarley@math.uga.edu
ACKNOWLEDGEMENTS

The editors gratefully acknowledge the service of the following persons who have been consulted concerning the preparation of volumes one hundred ninety-six through one hundred ninety-nine of the Pacific Journal of Mathematics.

CONTENTS

Volume 201, no. 1 and no. 2

Janusz J. Charatonik and Janusz R. Prajs: On local connectedness of absolute retracts 83

Elliot Benjamin: On imaginary quadratic number fields with 2-class group of rank 4 and infinite 2-class field tower 257

Albert Boggess: CR extension for $L^p$ CR functions on a quadric submanifold of $C^n$ 1

Thomas Branson and A. Rod Gover: Conformally invariant non-local operators 19

Mark Brittenham: Free genus one knots with large volume 61

Emilio Bujalance, Grzegorz Gromadzki and Peter Turbek: On non-orientable Riemann surfaces with $2p$ or $2p + 2$ automorphisms 267

Judith Epstein, Dmitry Fuchs and Maike Meyer: Chekanov–Eliashberg invariants and transverse approximations of Legendrian knots 89

Huitao Feng and Enli Guo: Novikov-type inequalities for vector fields with non-isolated zero points 107

Dmitry Fuchs with Judith Epstein and Maike Meyer 89

Najoua Gamara and Ridha Yacoub: CR Yamabe conjecture – the conformally flat case 121

Yasuhiro Gotoh: On composition operators which preserve BMO 289

A. Rod Gover with Thomas Branson 19

Eric L. Grinberg and Eric Todd Quinto: Analytic continuation of convex bodies and Funk’s characterization of the sphere 309

Grzegorz Gromadzki with Emilio Bujalance and Peter Turbek 267

Enli Guo with Huitao Feng 107

Ya Ryong Heo: An endpoint estimate for some maximal operators associated to submanifolds of low codimension 323

Alexander S. Kleshchev and Daniel K. Nakano: On comparing the cohomology of general linear and symmetric groups 339

Vijay Kodiyalam and V.S. Sunder: Flatness and fusion coefficients 177

Bernhard Lamel: Holomorphic maps of real submanifolds in complex spaces of different dimensions 357
Gary M. Lieberman: Regularity of solutions of obstacle problems for elliptic equations with oblique boundary conditions 389

Maike Meyer with Judith Epstein and Dmitry Fuchs 89

Daniel K. Nakano with Alexander S. Kleshchev 339

Sergey Neshveyev and Erling Størmer: Entropy in type I algebras 421

Kyewon Koh Park and Ayşe A. Şahin: Even Kakutani equivalence via $\vec{\alpha}$ and $\vec{\beta}$ equivalence in $\mathbb{Z}^2$ 205

Janusz R. Prajs with Janusz J. Charatonik 83

Yuanwei Qi and Mingxin Wang: The self-similar profiles of generalized KPZ equation 223

Eric Todd Quinto with Eric L. Grinberg 309

B. Rodrigues and W. Veys: Holomorphy of Igusa’s and topological zeta functions for homogeneous polynomials 429

Ayşe A. Şahin with Kyewon Koh Park 205

K. Simon, B. Solomyak and M. Urbański: Hausdorff dimension of limit sets for parabolic IFS with overlaps 441

Roy Smith and Robert Varley: A Riemann singularities theorem for Prym theta divisors, with applications 479

B. Solomyak with K. Simon and M. Urbański 441

Erling Størmer with Sergey Neshveyev 421

Robert S. Strichartz: The Laplacian on the Sierpinski gasket via the method of averages 241

V.S. Sunder with Vijay Kodiyalam 177

Peter Turbek with Emilio Bujalance and Grzegorz Gromadzki 267

M. Urbański with K. Simon and B. Solomyak 441

Robert Varley with Roy Smith 479

W. Veys with B. Rodrigues 429

Mingxin Wang with Yuanwei Qi 223

Ridha Yacoub with Najoua Gamara 121
Guidelines for Authors

Authors may submit manuscripts at pjm.math.berkeley.edu/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use \LaTeX, but papers in other varieties of \TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as \LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of Bib\TeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text (“the curve looks like this:”). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by means of comments such as “Place Figure 1 here”. The same considerations apply to tables, which should be used sparingly.

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.

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.
On imaginary quadratic number fields with 2-class group of rank 4 and infinite 2-class field tower
Elliot Benjamin

On non-orientable Riemann surfaces with $2p$ or $2p+2$ automorphisms
Emilio Bujalance, Grzegorz Gromadzki and Peter Turbek

On composition operators which preserve BMO
Yasuhiro Gotô

Analytic continuation of convex bodies and Funk’s characterization of the sphere
Eric L. Grinberg and Eric Todd Quinto

An endpoint estimate for some maximal operators associated to submanifolds of low codimension
Ya Ryong Heo

On comparing the cohomology of general linear and symmetric groups
Alexander S. Kleshchev and Daniel K. Nakano

Holomorphic maps of real submanifolds in complex spaces of different dimensions
Bernhard Lamel

Regularity of solutions of obstacle problems for elliptic equations with oblique boundary conditions
Gary M. Lieberman

Entropy in type I algebras
Sergey Neshveyev and Erling Størmer

Holomorphy of Igusa’s and topological zeta functions for homogeneous polynomials
B. Rodrigues and W. Veys

Hausdorff dimension of limit sets for parabolic IFS with overlaps
K. Simon, B. Solomyak and M. Urbański

A Riemann singularities theorem for Prym theta divisors, with applications
Roy Smith and Robert Varley

ACKNOWLEDGEMENTS

Volume 201 No. 2 December 2001