CONSTRUCTION DE LA TOUR DES 2-CORPS DE
CLASSES DE HILBERT DE CERTAINS CORPS
BIQUADRATIQUES

Abdelmalek Azizi

Let \( p \) and \( q \) be prime numbers such that \( p \equiv 1 \mod 8 \), \( q \equiv -1 \mod 4 \) and \( (\frac{p}{q}) = -1 \), \( d = pq \), \( k = \mathbb{Q}(\sqrt{d}, i) \), \( k_2^{(1)} \) be the 2-Hilbert class field of \( k \), \( k_2^{(2)} \) be the 2-Hilbert class field of \( k_2^{(1)} \) and \( G_2 \) be the Galois group of \( k_2^{(2)}/k \). The 2-part \( C_{k,2} \) of the class group of \( k \) is of type \((2,2)\), so \( k_2^{(1)} \) contains three extensions \( K_i/k \), \( i = 1, 2, 3 \). Our goal is to determine the group \( C_{k,2} \), to study the problem of capitulation of the 2-classes of \( k \) in \( K_i \), \( i = 1, 2, 3 \) and to construct the 2-class field tower of \( k \).

Résumé.

Soient \( p \) et \( q \) deux nombres premiers tels que \( p \equiv 1 \mod 8 \), \( q \equiv -1 \mod 4 \) et \( (\frac{p}{q}) = -1 \), \( d = pq \), \( i = \sqrt{-1} \), \( k = \mathbb{Q}(\sqrt{d}, i) \), \( k_2^{(1)} \) le 2-corps de classes de Hilbert de \( k \), \( k_2^{(2)} \) le 2-corps de classes de Hilbert de \( k_2^{(1)} \) et \( G_2 \) le groupe de Galois de \( k_2^{(2)}/k \). La 2-partie \( C_{k,2} \), du groupe de classes de \( k \) est de type \((2,2)\), par suite \( k_2^{(1)} \) contient trois extensions \( K_i/k \), \( i = 1, 2, 3 \). On s’intéresse à déterminer le groupe \( C_{k,2} \), à étudier la capitulation des 2-classes de \( k \) dans \( K_i \), \( i = 1, 2, 3 \) et à la construction de la tour du 2-corps de classes de Hilbert de \( k \).

1. Introduction.

Soient \( k \) un corps de nombres de degré fini sur \( \mathbb{Q} \), \( F \) une extension non ramifiée de \( k \) et \( p \) un nombre premier. L'extension \( k^{(1)} \) de \( k \), abélienne maximale et non-ramifiée pour tous les idéaux premiers finis et infinis , est dite corps de classes de Hilbert de \( k \). De même l'extension \( k_2^{(1)} \) de \( k \) dont le degré est une puissance de \( p \), abélienne maximale et non-ramifiée pour tous les idéaux premiers finis et infinis est dite \( p \)-corps de classes de Hilbert de \( k \).

La recherche des idéaux de \( k \) qui capitulent dans \( F \) (deviennent principaux dans \( F \) ), a été l’objet d’étude d’un grand nombre de mathématiciens. En effet, Kronecker était parmi les premiers à avoir abordé des problèmes de capitulation dans le cas des corps quadratiques imaginaires. Dans le cas
où $F$ est égal au corps de classes de Hilbert $k^{(1)}$ de $k$. D. Hilbert avait conjecturé que toutes les classes de $k$ capitulent dans $k^{(1)}$ (théorème de l'idéal principal). La preuve de ce dernier théorème a été réduite par E. Artin à un problème de la théorie des groupes, et c’est Ph. Furtwängler qui l’avait achevée.

Le cas où $F/k$ est une extension cyclique et $[F:k] = p$, un nombre premier, a été traité par Hilbert. Sa réponse est le sujet du Théorème 94 qui affirme qu’il y a au moins une classe non-triviale dans $k$ qui capitule dans $F$. De plus, Hilbert avait trouvé le résultat suivant:

Soient $\sigma$ un générateur du groupe de Galois de $F/k$, $N$ la norme de $F/k$, $U_0$ le groupe des unités de $k$, $U$ le groupe des unités de $F$ et $U^*$ le sous-groupe des unités de $U$ dont la norme, relative à l’extension $F/k$, est égale à 1. Alors le groupe des classes de $k$ qui capitulent dans $F$ est isomorphe au groupe quotient $U^*/U^{1-\sigma} = H^1(U)$, le groupe cohomologique de $U$ de dimension 1.

À l’aide de ce théorème et de plusieurs résultats sur les groupes cohomologiques des unités, on montre le théorème suivant:

**Théorème 1.** Soit $F/k$ une extension cyclique de degré un nombre premier, alors le nombre des classes qui capitulent dans $F/k$ est égal à $[F:k][U_0 : N(U)]$.

On trouve une preuve de ce théorème dans un papier de Heider et Schmithals [11].

Plusieurs résultats ont été établis; en particulier on a:

Soit $k$ tel que $C_{k,2}$, la 2-partie du groupe des classes $C_k$ de $k$, est isomorphe à $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, $k_2^{(2)}$ le 2-corps de classes de Hilbert de $k_2^{(1)}$ et $G_2$ le groupe de Galois de $k_2^{(2)}/k$. On sait par la théorie des corps de classes que $\text{Gal}(k_2^{(1)}/k) \simeq C_{k,2}$, par suite $\text{Gal}(k_2^{(1)}/k) \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. Alors $k_2^{(1)}$ contient trois extensions quadratiques de $k$ dénotées par $K_1$, $K_2$ et $K_3$.

D’après Kisilevsky [12] on a:

**Théorème 2.** Soient $k$ tel que $C_{k,2} \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ et $G_2$ le groupe de Galois de $k_2^{(2)}/k$; alors on a trois types de capitulation:

Type 1. *Les quatre classes de $C_{k,2}$ capitulent dans chacune des extensions $K_i/k$, $i = 1, 2, 3$. Ceci est possible si et seulement si $k_2^{(2)} = k_2^{(1)}$.*

Type 2. *Les quatre classes de $C_{k,2}$ capitulent toutes seulement dans une extension parmi les trois extensions $K_i/k$, $i = 1, 2, 3$. Dans ce cas le groupe $G_2$ est diédral.*

Type 3. *Seulement deux classes capitulent dans chacune des extensions $K_i/k$, $i = 1, 2, 3$. Dans ce cas le groupe $G_2$ est semidiédral ou quaternionique.*
CONSTRUCTION DE LA TOUR DE HILBERT

Soit $G'_2$ le groupe dérivé de $G_2$, si $C_{k,2} \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, alors on a de même $G'_2/G'_2 \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, ce qui implique que $G'_2$ est cyclique. Comme $G'_2 \simeq \text{Gal}(k_2^{(2)}/k_2)$ est cyclique, alors la tour des 2-corps de classes de Hilbert de $k$ s'arrête en $k_2^{(2)}$. Donc pour construire la tour des 2-corps de classes de Hilbert, il suffit de construire $k_2^{(1)}$ et $k_2^{(2)}$. On trouve plusieurs travaux sur la construction des 2-corps de classes de Hilbert; en particulier on trouve les travaux de H. Cohn dans [9] et dans [10], qui a construit le 2-corps de classes de Hilbert de $\mathbb{Q}(\sqrt{-p})$ où $p$ est un premier tel que $p \equiv 1 \mod 4$.

Dans toute la suite on désigne par $p$ et $q$ deux nombres premiers tels que $p \equiv 1 \mod 8$, $q \equiv -1 \mod 4$ et $(\frac{p}{q}) = -1$, $d = pq$, $k = \mathbb{Q}(\sqrt{d}, i)$, $k_2^{(1)}$ le 2-corps de classes de Hilbert de $k$, $k_2^{(2)}$ le 2-corps de classes de Hilbert de $k_2^{(1)}$ et $G_2$ le groupe de Galois de $k_2^{(2)}/k$. D'après Azizi [2], on a $C_{k,2} \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. Donc $k_2^{(1)}$ contient trois extensions quadratiques de $k$, $K_1$, $K_2$ et $K_3$. Notre but est de déterminer $C_{k,2}$, d'étudier la capitulation dans les trois extensions $K_i/k$, $i = 1, 2, 3$ et de construire la tour des 2-corps de classes de Hilbert de $k$. En particulier on a le résultat principal suivant:

**Théorème 3.** Soient $\epsilon$ l'unité fondamentale de $\mathbb{Q}(\sqrt{p})$, $L = \mathbb{Q}(\sqrt{-p})$ et $L_2^{(1)}$ le 2-corps de classes de Hilbert de $L$. Alors il existe deux entiers $a$, $b \in \mathbb{N}$ tels que $p = a^2 + 16b^2$. Soient $\pi_1 = a + 4bi$, $\pi_2 = a - 4bi$, $\mathcal{H}_1$ et $\mathcal{H}_2$ les idéaux premiers au-dessus de $\pi_1$ et $\pi_2$ dans $k$. Alors la 2-partie du groupe de classes de $k$ est engendré par les classes de $\mathcal{H}_1$ et $\mathcal{H}_2$ et les trois extensions quadratiques non-ramifiées sur $k$ sont: $k^{(*)} = k(\sqrt{p})$, $K_1 = k(\sqrt{\pi_1})$ et $K_2 = k(\sqrt{\pi_2})$. De plus si $k_2^{(2)} \neq k_2^{(1)}$, alors seules la classe de $\mathcal{H}_1$ et son carré capitulent dans $K_1$ et il en est de même pour $K_2$, c'est-à-dire seules la classe de $\mathcal{H}_2$ et son carré capitulent dans $K_2$. De plus on a $k_2^{(1)} = k(\sqrt{p})(\sqrt{\epsilon})$ et $k_2^{(2)} = k^{(*)}L_2^{(1)}$.

2. Capitulation dans le corps de genres de $k$.

Soient $p$ et $q$ deux nombres premiers, $d = pq$ et $k = \mathbb{Q}(\sqrt{d}, i)$. Dans toute la suite on supposera que $p \equiv 1 \mod 8$, $q \equiv -1 \mod 4$ et $(\frac{p}{q}) = -1$. Soient $k_2^{(1)}$ le 2-corps de classes de Hilbert de $k$ et $k^{(*)}$ le corps des genres de $k$ (c'est l'extension maximale non-ramifiée pour tous les idéaux premiers, finis et infinis, et qui est abélienne sur $\mathbb{Q}$).
Diagramme 1.

Le corps des genres de $k$ est $k^{(s)} = \mathbb{Q}(\sqrt{p}, \sqrt{q}, i)$. D’après Azizi [6], on a les deux résultats suivants:

**Théorème 4.** Soient $k = \mathbb{Q}(\sqrt{pq}, i)$ avec $p$ et $q$ deux nombres premiers tels que $q \equiv -1 \mod 4, p \equiv 1 \mod 8, (\frac{p}{q}) = -1$, $k^{(s)} = \mathbb{Q}(\sqrt{p}, \sqrt{q}, i)$ le corps des genres de $k$, $k_2^{(1)}$ le 2-corps de classes de Hilbert de $k$, $k_2^{(2)}$ le 2-corps de classes de Hilbert de $k_2^{(1)}$ et $C_{k,2}$ la 2-partie du groupe des classes au sens large de $k$. Alors on a:

1. toutes les classes de $C_{k,2}$ capitulent dans $k^{(s)}$.
2. $k_2^{(2)} \neq k_2^{(1)} \iff 4|h(k^{(s)}) \iff p = x^2 + 32y^2$.

**Corollaire 5.** Soit $k = \mathbb{Q}(\sqrt{pq}, i)$ avec $p$ et $q$ deux nombres premiers tels que $q \equiv -1 \mod 4, p \equiv 1 \mod 8, (\frac{p}{q}) = -1$. Soit $G_2$ le groupe de Galois de $k_2^{(2)}/k$. Alors le groupe $G_2$ est de type $(2, 2)$ ou bien diéidal d’ordre $2^m$. De plus, $G_2$ est diéidal si et seulement si $p = x^2 + 32y^2$ avec $x$ et $y$ deux entiers naturels.

D’après le Théorème 4, toutes les classes de $C_{k,2}$ capitulent dans $k^{(s)}$ et d’après le Théorème 2 on a les deux possibilités suivantes:

i) si $k_2^{(2)} = k_2^{(1)}$, alors toutes les classes de $C_{k,2}$ capitulent dans $K_1$ et dans $K_2$.

ii) si $k_2^{(2)} \neq k_2^{(1)}$, alors seulement une classe non-triviale de $C_{k,2}$ capitule dans $K_1$ et il en est de même pour $K_2$. 
Dans toute la suite on va déterminer le groupe \( C_{k,2} \) et mettre le point sur les classes qui capitulent dans \( K_1 \) et celles qui capitulent dans \( K_2 \).

3. Capitulation de type classe et construction de la tour des 2-corps de classes de Hilbert.

Soient \( d = pq \) avec \( p \) et \( q \) deux premiers tels que \( p \equiv 1 \mod 8 \), \( q \equiv -1 \mod 4 \) et \( \left( \frac{p}{q} \right) = -1 \), \( k = \mathbb{Q}(\sqrt{d}, i) \), \( C_{k,2} \) le 2-groupe des classes de \( k \), \( E_k \) le groupe des unités de \( k \), \( k_2^{(1)} \) le 2-corps de classes de Hilbert de \( k \), \( k^{(s)} \) le corps des genres de \( k \), \( H \) le groupe de Galois de \( k_2^{(1)}/k \) et \( \hat{H} \) le groupe des caractères de \( H \). Soit \( L \) une extension quadratique de \( k \). Alors il existe un nombre \( \alpha \) de \( k \) tel que \( L = k(\sqrt{\alpha}) \). Si \( L/k \) est non-ramifié aux idéaux premiers finis, alors on peut choisir \( \alpha \) premier avec 2 tel que:

i) il existe un idéal \( H \) tel que \( H^2 = (\alpha) \);

ii) il existe \( x \in k \) tel que \( \alpha = x^2 \mod 4 \).

Cette dernière condition est satisfaite si et seulement si \( \alpha = x^2 + 4r/s \), où \( r \) et \( s \) sont des entiers de \( k \) et \( s \) est premier avec 2.

Soient \( \overline{R_k} \) l’ensemble de tous les nombres \( \alpha \) de \( k \) vérifiant les deux conditions précédentes, \( R_k = \overline{R_k}/\overline{R_k} \cap (k^*)^2 \) et \( U_k \) l’ensemble des unités de \( k \) appartenant à \( \overline{R_k} \).

Définition 6. On dit que \( k \) est de type classe si et seulement si \( U_k = E_k^2 \). Dans le cas contraire on dit que \( k \) est de type unité.

On définit un homomorphisme \( \varphi \) de \( R_k \) dans \( C_{k,2} \) de la façon suivante: À une classe de \( R_k \) d’un nombre \( \alpha \) on fait correspondre la classe de l’idéal \( H \) tel que \( H^2 = (\alpha) \). Alors on a Ker \( (\varphi) = U_k/E_k^2 \) et \( k \) est de type classe si et seulement si Im \( (\varphi) \) est égal au sous-groupe du classes au sens restreint, engendré par les éléments d’ordre deux.

Proposition 7. Soient \( k^{(s)} = k(\sqrt{p}) \) le corps des genres de \( k \) et \( H_0 \) l’idéal de \( k \) tel que \( H_0^2 = (p) \). Si \( k \) est de type classe, alors la classe de l’idéal \( H_0 \) est d’ordre 2. Si \( (\pi_1) \) et \( (\pi_2) \) sont les deux idéaux premiers au-dessus de \( p \) dans \( \mathbb{Q}(i) \) et \( H_1 \) (resp. \( H_2 \)) est un idéal premier au-dessus de \( \pi_1 \) (resp. \( \pi_2 \)) dans \( k/\mathbb{Q}(i) \), alors les classes de \( H_1 \) et de \( H_2 \) engendrent \( C_{k,2} \) et on a \( H_1H_2 = H_0 \).

Preuve. Comme \( p \equiv 1 \mod 4 \), alors il existe deux nombres \( \pi_1 \) et \( \pi_2 \) de \( \mathbb{Q}(i) \) tels que \( \pi_1\pi_2 = p \). De plus, puisque \( p \) est ramifié dans \( k/\mathbb{Q}(i) \), alors il existe deux idéaux de \( k \), \( H_1 \) et \( H_2 \) tels que \( H_1^2 = (\pi_1), H_2^2 = (\pi_2) \) et \( (H_1H_2)^2 = (p) = H_0^2 \). D’où \( H_1H_2 = H_0 \). Si \( k \) est de type classe , la classe de l’idéal \( H_0 \) dans \( C_{k,2} \) est d’ordre 2, car sinon, il existe un \( \beta \in k \) tel que \( H_0 = (\beta) \) et \( (\beta^2) = (p) \). Il s’ensuit que \( p = \beta^2 \epsilon \) pour une certaine unité \( \epsilon \) de \( k \). Comme \( k(\sqrt{p}) = k(\sqrt{\epsilon}) \) est non-ramifié pour tous les idéaux premiers, alors \( \epsilon \in U_k \) et \( \epsilon \notin E_k^2 \), ce qui est contraire au fait que \( k \) est de type classe.
Il vient que les classes de \( H_1 \) et \( H_2 \) sont d’ordre 2 et comme leur produit est d’ordre 2, alors leurs classes engendrent \( C_{k,2} \).

**Proposition 8.** Soient \( a \) un entier composé, impair et sans facteurs carrés, \( k = \mathbb{Q}(\sqrt{a}, i) \), \( p \) un nombre premier et \( \mathcal{H} \) un idéal de \( k \) tel que \( \mathcal{H}^2 = (p) \). Alors on a:

i) Si l’unité fondamentale de \( \mathbb{Q}(\sqrt{a}) \) est de norme \(-1\), alors \( \mathcal{H} \) est d’ordre 2 dans \( C_{k,2} \).

ii) Si l’unité fondamentale de \( \mathbb{Q}(\sqrt{a}) \), \( \epsilon_0 = s + t\sqrt{a} \), est de norme 1 on a:

a) Si \( \{\epsilon_0\} \) est un SFU de \( k \), alors \( \mathcal{H} \) est principal si et seulement si \( 2p(s \pm 1) \) ou \( p(s \pm 1) \) est un carré dans \( \mathbb{N} \).

b) Sinon, l'idéal \( \mathcal{H} \) est d'ordre 2 dans \( C_{k,2} \).

**Preuve.** Soient \( p \) un nombre premier et \( \mathcal{H} \) un idéal de \( k \) tel que \( \mathcal{H}^2 = (p) \). On suppose que \( \mathcal{H} \) est principal. Il existe \( \beta \in k \) et \( \epsilon \) une unité de \( k \) tels que \( \beta^2 = p\epsilon \). Nous déterminons les conditions pour que \( p\epsilon \) soit un carré dans \( k \).

Soit \( \epsilon_0 = s + t\sqrt{a} \) l’unité fondamentale de \( \mathbb{Q}(\sqrt{a}) \). Alors un SFU de \( k \) est \( \{\epsilon_0\} \) ou \( \{\sqrt{i\epsilon_0}\} \).

Cas où \( \{\epsilon_0\} \) est un SFU de \( k \):

On se ramène aux cas où \( \epsilon \) est égal à \( i, \epsilon_0 \) ou à \( i\epsilon_0 \).

* Si \( \epsilon = \epsilon_0 \), alors il existe \( (\beta_1, \beta_2) \in \mathbb{Q}(\sqrt{a})^2 \) tel que \( p\epsilon_0 = \beta^2 = (\beta_1 + \beta_2i)^2 = \beta_1^2 - \beta_2^2 + 2\beta_1\beta_2i \). D'où

\[
\begin{align*}
\beta_1\beta_2 &= 0 \\
\beta_1^2 - \beta_2^2 &= p\epsilon_0
\end{align*}
\]

On pose \( \beta_1 = x + y\sqrt{a} \). Alors on a:

\[
\begin{align*}
x^2 + ay^2 &= ps \\
2xy &= pt
\end{align*}
\]

Le nombre \( \Delta \) est le discriminant de l’équation du deuxième degré \( 4x^4 - 4pxs^2 + ap^2t^2 = 0 \) pour l’indéterminée \( x^2 \). On désigne cette dernière équation par (1).

- Si \( \epsilon_0 \) est de norme \(-1\) alors \( \Delta \) est négatif. Donc il n’y a pas de solutions pour l’équation (1).

- Si \( \epsilon_0 \) est de norme 1, alors \( x^2 = \frac{2ps \pm 2p}{4} \). Par suite, il y a une solution pour l’équation (1) si et seulement si \( 2p(s \pm 1) \) est un carré dans \( \mathbb{N} \).

* Soit \( \epsilon = i\epsilon_0 \). De la même façon, on se ramène à \( 2p\epsilon_0 = \gamma^2 \) où \( \gamma \in \mathbb{Q}(\sqrt{a}) \) et on trouve des résultats similaires:

- Si \( \epsilon_0 \) est de norme \(-1\), il n’y a pas de solutions.

- Sinon, il y a une solution si et seulement si \( p(s \pm 1) \) est un carré dans \( \mathbb{N} \).

* Soit $\epsilon = i$. On a $pi = \beta^2 \iff p = 2\beta_1^2$ où $\beta_1$ est la partie réelle de $\beta$. Or ceci implique que $\sqrt{2p} \in \mathbb{Q}(\sqrt{a})$, ce qui n’est pas notre cas.

Cas où $\{\sqrt{i\epsilon_0}\}$ est un SFU de $k$:

Soit l’équation $pe = \beta^2$. Alors on se ramène aux cas: $p^2\epsilon^2 = \beta^4$ et $\epsilon^2 = \pm i\epsilon_0$, $\epsilon = i\epsilon_0$ ou $\epsilon = \epsilon_0$.

- Si $\epsilon^2 = \pm i\epsilon_0$, on a $\beta = \beta_1 + i\beta_2$ et $\pm ip^2\epsilon_0 = (\beta_1^2 - \beta_2^2)^2 - 4(\beta_1\beta_2)^2 + 4\beta_1\beta_2(\beta_1^2 - \beta_2^2)i$. D’où $(\beta_1^2 - \beta_2^2)^2 - 4(\beta_1\beta_2)^2 = (\beta_1^2 - \beta_2^2 - 2\beta_1\beta_2)(\beta_1^2 - \beta_2^2 + 2\beta_1\beta_2) = 0$ et donc on a $\beta_1 - \beta_2)^2 = 2\beta_2^2$ ou $(\beta_1 + \beta_2)^2 = 2\beta_2^2$, ce qui entraîne que $\sqrt{2} \in \mathbb{Q}(\sqrt{a})$. Mais ceci n’est pas notre cas.

- Si $\epsilon = i\epsilon_0$, alors $p\epsilon_0 = \beta^2$ implique que $p$ est un carré dans $\mathbb{Q}(\sqrt{a})$. Ceci n’est pas notre cas.

- Si $\epsilon = \epsilon_0$, on se ramène au cas $\epsilon = i$ (car $i\epsilon_0$ est un carré dans $k$).

**Proposition 9.** Soit $k = \mathbb{Q}(\sqrt{d}, i)$ où $d = pq$ avec $p \equiv 1 \mod 8$, $q \equiv -1 \mod 4$, $(\frac{p}{q}) = -1$. Alors $k$ est de type classe.

**Preuve.** D’après Aziziz [6], $\{\sqrt{i\epsilon_0}\}$ est un SFU de $k$, donc d’après la proposition précédente, l’idéal $\mathcal{H}$ tel que $\mathcal{H}^2 = (p)$ est d’ordre 2 dans $C_{k,2}$. Par suite $k^{(1)}$ n’est pas de la forme $k(\sqrt{\epsilon})$ où $\epsilon$ est une unité de $k$. On suppose que $k$ est de type unité. Il existe une unité $\epsilon$ de $k$ telle que $k(\sqrt{\epsilon})$ soit non-ramifié sur $k$. En particulier, il existe $x$ dans $k$, $r$ et $s$ deux entiers de $k$ tels que $s$ est premier avec 2 et $\epsilon = x^2 + 4\mathbb{Z}$. On désigne par $\epsilon_1 = \sqrt{i\epsilon_0}$ et $\sigma$ l’automorphisme de $k$ défini par $\sigma(i) = i$ et $\sigma(\sqrt{d}) = -\sqrt{d}$. L’unité $\epsilon$ ne peut pas être égale à $i$, car sinon, $k(\sqrt{2})$ sera non-ramifié sur $k$ et par suite $k_2^{(1)}$ sera égal à $k(\sqrt{2}, \sqrt{p})$ et sera abélien sur $Q$. D’où $k_2^{(1)} = k^{(1)}$, ce qui n’est pas le cas. Par conséquent, $\epsilon = \pm i\epsilon_1$ pour un certain entier $m$. Il vient ensuite que $\sigma(\epsilon) = \pm im\sqrt{i\epsilon_0}$ où $\epsilon_0$ est le conjugué de $\epsilon_0$ et $\sigma(\epsilon) = \sigma(x)^2 + \frac{\sigma(r)}{\sigma(s)}$, où $\sigma(r)$ et $\sigma(s)$ restent des entiers de $k$ et $\sigma(s)$ reste premier avec 2. D’où $k(\sqrt{\sigma(\epsilon)})$ est non-ramifié sur $k$. D’autre part, $\epsilon\sigma(\epsilon) = im\sqrt{i\epsilon_0} \times (\pm im) \sqrt{i\epsilon_0} = \pm i$ n’est pas un carré dans $k$. Donc $k(\sqrt{\sigma(\epsilon)}) \neq k(\sqrt{\sigma(\epsilon)})$ et $k(\sqrt{\sigma(\epsilon)})$ est non-ramifié sur $k$, ce qui n’est pas possible d’après le cas $\epsilon = i$. On en déduit que $k$ n’est pas de type unité. Donc $k$ est de type classe. 

**Théorème 10.** Soit $k = \mathbb{Q}(\sqrt{pq}, i)$ avec $p$ et $q$ deux nombres premiers tels que $q \equiv -1 \mod 4$, $p \equiv 1 \mod 8$, $(\frac{p}{q}) = -1$. Soit $\epsilon$ l’unité fondamentale de $\mathbb{Q}(\sqrt{p})$. Alors $k_2^{(1)} = k(\sqrt{p})(\sqrt{i})$. 

Preuve. Montrons que $k_2^{(1)} = k(\sqrt{p})(\sqrt{\epsilon})$. On sait d’après Cohn [9], que si $p \equiv 1 \mod 8$, alors $k_1 = \mathbb{Q}(\sqrt{p}, i)(\sqrt{\epsilon})$ est une extension cyclique non-ramifiée sur $\mathbb{Q}(\sqrt{-p})$. Soit $G_0$ le groupe de Galois de $k_1/Q$ et $\epsilon'$ le conjugué de $\epsilon$, alors $G_0$ est engendré par les automorphismes $\sigma$ et $\tau_1$ définis par:

<table>
<thead>
<tr>
<th></th>
<th>$\sqrt{p}$</th>
<th>$i$</th>
<th>$\sqrt{\epsilon}$</th>
<th>$\sqrt{\epsilon'}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>$-\sqrt{p}$</td>
<td>$-i$</td>
<td>$\sqrt{\epsilon'}$</td>
<td>$-\sqrt{\epsilon}$</td>
</tr>
<tr>
<td>$\tau_1$</td>
<td>$-\sqrt{p}$</td>
<td>$i$</td>
<td>$\sqrt{\epsilon}$</td>
<td>$\sqrt{\epsilon'}$</td>
</tr>
</tbody>
</table>

De plus on a que $\sigma^4 = \tau_1^2 = (\sigma \tau_1)^2 = 1$. Comme $L = k(\sqrt{p})(\sqrt{\epsilon}) = k_1 Q(\sqrt{\epsilon})$; on peut prolonger $\sigma$ et $\tau_1$ par l’identité à $L$. De même, si on désigne par $\tau_2$ l’automorphisme défini sur $\mathbb{Q}(\sqrt{\epsilon})$ par $\tau_2(\sqrt{\epsilon}) = -\sqrt{\epsilon}$, alors on peut prolonger $\tau_2$ par l’identité à $L$. Par suite, le groupe de Galois de $L/Q$ est engendré par $\sigma$, $\tau_1$ et $\tau_2$. On va déterminer le groupe de Galois de $L/k$. Il est clair que $\sigma^2$ et $\tau_1 \tau_2$ laissent fixe $k$. D’autre part $\sigma^2(\sqrt{p}) = \sqrt{p}$, $\tau_1 \tau_2(\sqrt{p}) = -\sqrt{p}$, $\sigma^2(\sqrt{\epsilon}) = -\sqrt{\epsilon}$, $\tau_1 \tau_2(\sqrt{\epsilon}) = \sqrt{\epsilon}$, $(\sigma^2)^2 = (\tau_1 \tau_2)^2 = 1$ et $\sigma^2 \tau_1 \tau_2 \neq 1$. Donc le groupe engendré par $\sigma^2$ et $\tau_1 \tau_2$ laisse fixe $k$ et il est de type $(2, 2)$. De plus, on a que $k(\sqrt{p})$ est non-ramifié sur $k$ et comme $k_1$ est non-ramifié sur $Q(\sqrt{p}, i)$, alors $L$ est non-ramifié sur $Q(\sqrt{\epsilon}) Q(\sqrt{p}, i) = k(\sqrt{p})$. D’où $L$ est non-ramifié sur $k$ et le groupe de Galois de $L/k$ est de type $(2, 2)$. Par conséquent $L = k_2^{(1)}$. \(\square\)

**Théorème 11.** Soit $k = \mathbb{Q}(\sqrt{pq}, i)$ avec $p$ et $q$ deux nombres premiers tels que $q \equiv -1 \mod 4, p \equiv 1 \mod 8, (\frac{p}{q}) = -1$. Alors il existe deux entiers $a, b \in \mathbb{N}$ tels que $p = a^2 + 16b^2$. Soient $\pi_1 = a + 4bi$, $\pi_2 = a - 4bi$, $H_1$ et $H_2$ les idéaux premiers au-dessus de $\pi_1$ et $\pi_2$ dans $k$. Alors la 2-partie du groupe de classes de $k$ est engendré par les classes de $H_1$ et $H_2$ et les trois extensions quadratiques non-ramifiées sur $k$ sont: $k^{(1)} = k(\sqrt{p}), K_1 = k(\sqrt{\pi_1}) et K_2 = k(\sqrt{\pi_2})$. De plus si $k_2^{(2)} \neq k_2^{(1)}$, alors seules la classe de $H_1$ et son carré capitulent dans $K_1$ et il en est de même pour $K_2$, c’est-à-dire seules la classe de $H_2$ et son carré capitulent dans $K_2$.

**Preuve.** On sait, d’après Barruccand et Cohn [7], que si $p \equiv 1 \mod 8$, alors il existe deux entiers $a$ et $b$ tels que $p = a^2 + 16b^2$. Il est clair que $a$ est impair et donc $a \equiv \pm 1 \mod 4$. On pose $\pi_1 = a + 4bi$ et $\pi_2 = a - 4bi$. Comme $-1 = i^2$, alors $\pi_1 \equiv x^2 \mod 4 et \pi_2 \equiv x^2 \mod 4$ sont résolubles. Les nombres $\pi_1$ et $\pi_2$ sont des premiers ramifiés dans $k/Q(i)$. Par suite, il existe deux idéaux de $k$, $H_1$ et $H_2$ tels que $H_1^2 = (\pi_1)$ et $H_2^2 = (\pi_2)$. Par conséquent, $K_1/k$ et $K_2/k$ sont non-ramifiées. Comme $k$ est de type classe, alors la 2-partie du groupe des classes de $k$ est engendrée par les classes de $H_1$ et $H_2$. D’autre part, $H_1^2 = (\sqrt{\pi_1})^2$ entraîne que l’idéal engendré par $H_1$ dans $K_1 = k(\sqrt{\pi_1})$ est égal à l’idéal $(\sqrt{\pi_1})$. Ceci veut dire que $H_1$ capitule
dans $k_1$. Il en est de même pour $\mathcal{H}_2$ dans $k_2$. Ainsi, si $k_2^{(2)} \neq k_2^{(1)}$, comme toutes les classes de $C_2$ capitulent dans $k^{(*)}$, alors seulement la classe de $\mathcal{H}_1$ et son carré capitulent dans $K_1$ et seulement la classe de $\mathcal{H}_2$ et son carré capitulent dans $K_2$.

**Théorème 12.** Soit $k = \mathbb{Q}(\sqrt{pq}, i)$ avec $p$ et $q$ deux nombres premiers tels que $q \equiv -1 \mod 4$, $p \equiv 1 \mod 8$, $(\frac{p}{q}) = -1$. Soient $L = \mathbb{Q}(\sqrt{-p})$ et $L^{(1)}_2$ le 2-corps de classes de Hilbert de $L$, alors $k^{(2)}_2 = k^{(*)}L^{(1)}_2$.

**Preuve.** D’après le théorème précédent, $K_1$ et $K_2$ sont conjugués. Par suite $\text{Gal}(k^{(2)}_2/K_2)$ et $\text{Gal}(k^{(2)}_2/K_1)$ sont conjugués et les 2-groupes de classes de $K_1$ et $K_2$ ont la même structure. D’autre part, d’après le Théorème 2, si $k^{(2)}_2 \neq k^{(1)}_2$, alors les 2-groupes de classes des corps $k^{(*)} = k(\sqrt{p}), K_1 = k(\sqrt{\sqrt{p}})$ et $K_2 = k(\sqrt{\sqrt{2}})$ sont cycliques ou bien un seul corps parmi ces derniers corps est de 2-groupe de classes cyclique. Ainsi le 2-groupe de classes de $k^{(*)}$ est cyclique. Par suite $k^{(2)}_2 = (k^{(*)})^{(1)}_2$. En calculant la 2-partie du nombre de classes de $k^{(*)}$ on voit que si $L^{(1)}_2$ est le 2-corps de classes de Hilbert de $L$, alors $k^{(2)}_2 = k^{(*)}L^{(1)}_2$.

**Remarque 13.** Toute l’étude faite dans ce paragraphe pour le cas $d = pq$ est aussi valable pour le cas $d = p_1p_2$ où $p_1$ et $p_2$ sont deux nombres premiers tels que $p_1 \equiv 1 \mod 8$, $p_2 \equiv 5 \mod 8$ et $C_{k,2}$, le 2-groupe des classes de $k$, est de type $(2,2)$ (pour plus de détails pour ce cas voir [4]).

**References**


Received July 30, 2001 and revised November 28, 2001.

DÉPARTEMENT DE MATHEMATIQUES
Faculté des Sciences
Université Mohammed 1
Oujda
Maroc
E-mail address: azizi@sciences.univ-oujda.ac.ma
Let $X$ be an irreducible smooth projective curve over an algebraically closed field $k$ of characteristic $p$, with $p > 5$. Let $G$ be a connected reductive algebraic group over $k$. Let $H$ be a Levi factor of some parabolic subgroup of $G$ and $\chi$ a character of $H$. Given a reduction $E_H$ of the structure group of a $G$-bundle $E_G$ to $H$, let $E_\chi$ be the line bundle over $X$ associated to $E_H$ for the character $\chi$. If $G$ does not contain any $\text{SL}(n)/\text{Z}$ as a simple factor, where $\text{Z}$ is a subgroup of the center of $\text{SL}(n)$, we prove that a $G$-bundle $E_G$ over $X$ admits a connection if and only if for every such triple $(H, \chi, E_H)$, the degree of the line bundle $E_\chi$ is a multiple of $p$. If $G$ has a factor of the form $\text{SL}(n)/\text{Z}$, then this result is valid if $n$ is not a multiple of $p$. If $G$ is a classical group but not of the form $\text{SL}(n)/\text{Z}$, then this criterion for the existence of connection remains valid even if $p \geq 3$.

1. Introduction.

Let $X$ be an irreducible smooth projective curve over an algebraically closed field $k$. Take a vector bundle $E$ over $X$. A subbundle $V$ of $E$ is called a direct summand if the quotient homomorphism $E \longrightarrow E/V$ splits. For $k = \mathbb{C}$, a theorem of Andre Weil says that $E$ admits a connection if and only if every direct summand of $E$ is of degree zero [10].

Let $G$ be a connected reductive algebraic group over $k$. Let $E_G$ be a principal $G$-bundle over $X$. Our aim is to give a criterion for the existence of a connection on $E_G$. Note that since the dimension of $X$ is one, the curvature of a connection on $E_G$ must vanish. In other words, any connection on $E_G$ is automatically flat.

Let $H \subseteq G$ be a Levi factor of a parabolic subgroup of $G$. Let $E_H \subseteq E_G$ be a reduction of structure group of $E_G$ to $H$. Take a character

$$\chi : H \longrightarrow k^*$$

of $H$ and consider the line bundle $E_\chi := (E_H \times k)/H$ associated to $E_H$ for the character $\chi$. A connection on $E_G$ induces a connection on $E_H$ (Proposition 2.2), which, in turn, induces a connection on the line bundle $E_\chi$. 

11
It is well-known that a line bundle $\xi$ over $X$ admits a connection if and only if the degree of $\xi$ is a multiple of the characteristic of $k$ (see also Corollary 2.1). Therefore, if $E_G$ admits a connection then the degree of any line bundle $E_\chi$ of the above type must be a multiple of the characteristic of $k$.

Let $p$ denote the characteristic of $k$. We will assume that $p > 5$.

Let $Z(\text{SL}(n)) \subset \text{SL}(n)$ be the center. First assume that $G$ does not have a simple factor of the form $\text{SL}(n)/Z$, where $Z \subseteq Z(\text{SL}(n))$. We prove that if the degree of any $E_\chi$ of the above type is a multiple of $p$, then $E_G$ admits a connection (Theorem 2.3).

If $G$ contains a simple factor of the form $\text{SL}(n)/Z$, where $Z \subseteq Z(\text{SL}(n))$, and $p$ does not divide $n$, then $E_G$ admits a connection if and only if the degree of any line bundle $E_\chi$ of the above type is a multiple of $p$.

If $G$ is a classical group but not $\text{SL}(n)/Z$, where $Z \subseteq Z(\text{SL}(n))$, then the condition $p > 5$ can be relaxed to $p > 2$.

2. The Atiyah bundle.

Let $H$ be an algebraic group over $k$. Take a principal $H$-bundle $E_H$ over $X$. The projection of the total space of $E_H$ to $X$ will be denoted by $\psi$. For any open subset $U$ of $X$, consider the space of $H$-invariant vector fields on $\psi^{-1}(U)$ for the natural action of $H$ on the fibers of $\psi$. This gives rise to a vector bundle $\text{At}(E_H)$ on $X$ known as the Atiyah bundle.

Let $\mathfrak{h}$ denote the Lie algebra of $H$. Consider the adjoint action of $H$ on $\mathfrak{h}$. The associated vector bundle $(E_H \times \mathfrak{h})/H$, known as the adjoint bundle, will be denoted by $\text{ad}(E_H)$. Note that $\text{ad}(E_H)$ corresponds to the sheaf of $H$-invariant vertical vector fields on $E_H$. Therefore, we have an exact sequence

\begin{equation}
0 \to \text{ad}(E_H) \to \text{At}(E_H) \to TX \to 0
\end{equation}

of vector bundles over $X$. This sequence is known as the Atiyah exact sequence.

A connection on $E_H$ is a splitting of the exact sequence (2.1) [1], [9]. See Section 5 of [7] for connections on vector bundles in positive characteristics.

Note that both the sheaves $\text{At}(E_H)$ and $TX$ are equipped with a Lie algebra structure induced by the Lie bracket operation of vector fields. Given a splitting

$$\sigma : TX \to \text{At}(E_H)$$

of the Atiyah exact sequence, consider the homomorphism

$$\overline{\sigma} : TX \otimes TX \to \text{ad}(E_H)$$

defined by $s \otimes t \mapsto [\sigma(s), \sigma(t)] - \sigma([s, t])$, where $s$ and $t$ are local sections of $TX$, which is known as the curvature. Since $p \neq 2$, $\overline{\sigma}$ is skew-symmetric and $\dim X = 1$, we have $\overline{\sigma} = 0$. In other words, any connection on $X$ is flat.
Set $H = \text{GL}(n)$. So, using the standard representation of $\text{GL}(n)$, $E_H$ corresponds to a rank $n$ vector bundle $V$ over $X$. The exact sequence (2.1) becomes

\begin{equation}
0 \rightarrow \text{End}(V) \rightarrow \text{At}(V) \rightarrow TX \rightarrow 0,
\end{equation}

where $\text{At}(V)$ is the subbundle of the sheaf of differential operators $\text{Diff}^1_X(V, V)$ defined by the condition that the image by the symbol homomorphism $\text{Diff}^1_X(V, V) \rightarrow TX \otimes \text{End}(V)$ is contained in the subbundle $TX \otimes \text{Id}_V$.

Consider the extension class $\tau \in H^1(X, K_X \otimes \text{End}(V))$ for the exact sequence (2.2), where $K_X$ is the canonical bundle of $X$. Using the trace homomorphism $\text{tr} : \text{End}(V) \rightarrow \mathcal{O}_X$, we have $\text{tr}(\tau) \in H^1(X, K_X) = k$ (where $k$ is the base field), where the identification $H^1(X, K_X) = k$ is the one given by Serre duality.

Let $d \in \mathbb{Z}$ be the degree $V$, which is same as the degree of the line bundle $\wedge^n V$. The image of $d$ in $k$ by the obvious homomorphism $\mathbb{Z} \rightarrow k$ coincides with $\text{tr}(\tau)$.

Consequently, if a $\text{GL}(n)$-bundle admits a connection, then the degree of the corresponding rank $n$ vector bundle is a multiple of $p$, the characteristic of $k$.

This observation and the above identity $d = \text{tr}(\tau)$ together have the following corollary:

**Corollary 2.1.** A line bundle $\xi$ over $X$ admits a connection if and only if the degree of $\xi$ is a multiple of $p$ (possibly zero).

The above corollary is well-known [7, p. 190, Theorem 5.1] (in [7] this Theorem 5.1 is attributed to P. Cartier), [8].

As in the introduction, let $G$ be a connected reductive algebraic group over $k$. Let $P$ be a parabolic subgroup of $G$. Let $R_u(P)$ denote the unipotent radical of $P$. The quotient group $P/R_u(P)$ is called the Levi factor of $P$ [4]. The projection $P \rightarrow P/R_u(P)$ splits in the sense that there is a connected closed reductive subgroup $H$ of $P$ which projects isomorphically to $P/R_u(P)$. However, there may be more than one such subgroup. We will call a subgroup $H$ of $P$ with this property a Levi factor of $P$.

Take a $G$-bundle $E_G$ over $X$. Suppose

$$\sigma : X \rightarrow E_G/H$$

be a reduction of structure group of $E_G$ to a Levi factor $H$. So, the inverse image $q^{-1}(\sigma(X))$, where $q : E_G \rightarrow E_G/H$ is the obvious quotient map, is an $H$-bundle. This $H$-bundle will be denoted by $E_H$. 
Fix a character $\chi : H \longrightarrow k^*$ of $H$. Consider the quotient

$$E_\chi := (E_H \times k)/H$$

for the diagonal action of $H$, where $H$ acts on $k$ through $\chi$, which is a line bundle over $X$. We recall that the diagonal action of any $g \in H$ sends a point $(z, t) \in E_H \times k$ to $(zg, \chi(g^{-1})t)$.

**Proposition 2.2.** If $E_G$ admits a connection, then the degree of the line bundle $E_\chi$ is a multiple of $p$.

**Proof.** Any connection on $E_H$ induces a connection on $E_\chi$. Therefore, in view of Corollary 2.1 it suffices to show that any connection on $E_G$ induces a connection on $E_H$.

Let $\mathfrak{g}$ (respectively, $\mathfrak{h}$) denote the Lie algebra of $G$ (respectively, $H$). Since $H$ is a Levi factor, there exists a $H$-equivariant splitting

$$f : \mathfrak{g} \longrightarrow \mathfrak{h} \quad (2.3)$$

of the inclusion homomorphism of $\mathfrak{h}$ in $\mathfrak{g}$. Indeed, if $\mathfrak{p}$ and $\mathfrak{q}$ are two opposite parabolics containing $\mathfrak{h}$ as the common Levi factor, then the direct sum of the radicals of $\mathfrak{p}$ and $\mathfrak{q}$ is a $H$-invariant complement of $\mathfrak{h}$.

We recall that a connection on $E_G$ is a $\mathfrak{g}$-valued 1-form $\omega$ on $E_G$ satisfying the two conditions:

1) For any $v \in \mathfrak{g}$, the evaluation of $\omega$ on the vector field corresponding to $v$ coincides with the constant function $v$;
2) the form $\omega$ is equivariant for the action of $G$ on $E_G$ and the adjoint action of $G$ on its Lie algebra $\mathfrak{g}$.

The kernel of such a form $\omega$ defines a splitting of the Atiyah exact sequence (2.1). To explain this, let $\psi$ denote the projection of $E_G$ to $X$. Given a tangent vector $v \in T_x X$, where $x \in X$, and a point $z \in \psi^{-1}(x) \subset E_G$, there is a unique tangent vector $w \in T_z E_G$ projecting to $v$ that is contained in the kernel of the form $\omega$. This way we get a section $v'$ of $T_E_G$ over $\psi^{-1}(x)$. This section $v'$ is clearly $G$-invariant. In other words, $v'$ gives an element $v''$ of the fiber $\text{At}(E_G)_x$. Sending any $v$ to $v''$ we obtain a splitting of the Atiyah exact sequence (2.1). Conversely, given a splitting $\sigma : TX \longrightarrow \text{At}(E_G)$ of the Atiyah exact sequence, it is easy to see that there is a unique one-form $\omega$ satisfying the above two conditions such that the kernel of $\omega$ is the image of $\sigma$.

Given a connection on $E_G$ defined by a one-form $\omega$, let $\omega'$ denote the restriction of $\omega$ to $E_H \subseteq E_G$. Now consider the $\mathfrak{h}$-valued one-form

$$\overline{\omega} := f \circ \omega'$$

on $E_H$, where $f$ is defined in (2.3).

It is easy to check that the form $\overline{\omega}$ satisfies the two conditions needed to define a connection on $E_H$. Consequently, existence of a connection on $E_G$
ensures the existence of a connection on $E_H$. This completes the proof of the proposition. □

As before, let $Z(\text{SL}(n))$ denote the center of $\text{SL}(n)$.

Following is the main result proved here:

**Theorem 2.3.** Let $p > 5$ and assume that $G$ does not contain $\text{SL}(n)/Z$ as a simple factor, where $Z \subseteq Z(\text{SL}(n))$. A $G$-bundle $E_G$ over $X$ admits a connection if and only if for every pair $(H, \chi)$, where $\chi$ is a character of the Levi factor $H$ of some parabolic subgroup, the degree of the line bundle $E_\chi$ is a multiple of $p$. If there is a subgroup $Z \subseteq Z(\text{SL}(n))$ such that $G$ contains $\text{SL}(n)/Z$ as a simple factor, then same criterion is valid if $p > 5$ and $p$ does not divide $n$.

Since $H$ is reductive, Proposition 2.2 says that if $E_G$ admits a connection, then the line bundle $E_\chi$ is a multiple of $p$. We will complete the proof of the theorem in Section 4. In the next section we will show that it suffices to prove for simple groups.

### 3. Reduction to the case of simple groups.

Let $Z(G) \subset G$ denote the reduced center of $G$. Let

$$G' := G/Z(G)$$

be the quotient. Consider the commutator $[G, G]$, and let

$$Z = G/[G, G]$$

be the quotient. So $G'$ is a semisimple quotient of $G$ and $Z$ is an abelian quotient of $G$.

For a principal $G$-bundle $E_G$ on $X$, let $E_{G'}$ (respectively, $E_Z$) denote the principal $G'$-bundle (respectively, principal $Z$-bundle) obtained by extending the structure group of $E_G$ using the obvious projection of $G$ to $G'$ (respectively, $Z$).

**Lemma 3.1.** The $G$-bundle $E_G$ admits a connection if and only if both $E_{G'}$ and $E_Z$ admit connection.

**Proof.** Since $E_{G'}$ and $E_Z$ are extensions of structure group of $E_G$, any connection on $E_G$ induces connection on $E_{G'}$ and $E_Z$.

Note that the fiber product $E_{G'} \times_X E_Z$ is a principal $(G' \times Z)$-bundle. Let

$$\rho : G \rightarrow G' \times Z$$

be the diagonal homomorphism induced by the projections of $G$ to $G'$ and $Z$. Since the kernel of $\rho$ is finite and it induces an isomorphism of Lie algebras, the natural map $E_G \rightarrow E_{G'} \times_X E_Z$ is an étale covering map. Consequently, the Atiyah exact sequence for $E_G$ and $E_{G'} \times_X E_Z$ coincide.
It is easy to see that if

\[ 0 \to \text{ad} \left( E_{G'} \right) \to \mathcal{A} \xrightarrow{f_1} TX \to 0 \]

and

\[ 0 \to \text{ad}(E_Z) \to \mathcal{B} \xrightarrow{f_2} TX \to 0 \]

are the Atiyah exact sequences for \( E_{G'} \) and \( E_Z \) respectively, and \( p \) (respectively, \( q \)) is the obvious projection of \( \mathcal{A} \oplus \mathcal{B} \) to \( \mathcal{A} \) (respectively, \( \mathcal{B} \)), then the exact sequence

\[ 0 \to \text{ad}(E_{G'}) \oplus \text{ad}(E_Z) \to \ker(f_1 \circ p - f_2 \circ q) \subset \mathcal{A} \oplus \mathcal{B} \to TX \to 0 \]

obtained by combining the above two exact sequences is the Atiyah exact sequence for \( E_{G'} \times X E_Z \). From this it follows that if the Atiyah exact sequences for \( E_{G'} \) and \( E_Z \) split, then the Atiyah exact sequence for \( E_{G'} \times X E_Z \) also splits. Indeed, if

\[ \sigma_1 : TX \to \mathcal{A} \]

and \( \sigma_2 : TX \to \mathcal{B} \) are splittings of Atiyah exact sequences for \( E_{G'} \) and \( E_Z \) respectively, then the diagonal homomorphism

\[ (\sigma_1, \sigma_2) : TX \to \mathcal{A} \oplus \mathcal{B} \]

is the splitting of the Atiyah exact sequence for \( E_{G'} \times X E_Z \). Therefore, if both \( E_{G'} \) and \( E_Z \) admit connections then the \((G' \times Z)\)-bundle \( E_{G'} \times X E_Z \) also admits a connection. This completes the proof of the lemma. \( \square \)

The group \( Z \) is a product of copies of \( k^* \), and \( Z \) has exactly one parabolic subgroup which is \( Z \) itself. Therefore, Theorem 2.3 is valid for \( Z \).

The image of a parabolic subgroup \( P \) of \( G \) by the projection \( G \to G' \) is a parabolic subgroup of \( G' \). Moreover, all parabolic subgroups of \( G' \) arise this way. The image, in \( G' \), of a Levi factor \( H \subset P \) is a Levi factor of the corresponding parabolic subgroup of \( G' \).

Consequently, to establish Theorem 2.3 for \( G \), it suffices to prove it for the semisimple group \( G' \).

Any parabolic subgroup of \( G_1 \times G_2 \), where \( G_1 \) and \( G_2 \) are semisimple, is of the form \( P_1 \times P_2 \), where \( P_i \) is a parabolic subgroup of \( G_i \). Furthermore, from the Proof of Proposition 2.2 it follows immediately that if we have \( G_i \)-bundle \( E_{G_i} \), \( i = 1, 2 \), over \( X \), then both \( E_{G_1} \) and \( E_{G_2} \) admit a connection if and only if the principal \((G_1 \times G_2)\)-bundle \( E_{G_1} \times X E_{G_2} \) admits a connection. Therefore, it suffices to prove Theorem 2.3 under the assumption that \( G \) is simple.

If \( H \subset G \) is a Levi factor of a parabolic subgroup of \( G \), and if \( H_1 \subset H \) is a Levi factor of a parabolic subgroup of \( H \), then \( H_1 \), as a subgroup of \( G \), is a Levi factor of some parabolic subgroup of \( G \). Since \( H_1 \) is a Levi factor of \( G \), using reverse induction, we may reduce the structure group of \( G \) to such a situation where it does not admit any further reduction to some
Levi factor. A Levi factor $H$ of some parabolic subgroup of $G$ will be called *nontrivial* if $H$ is a proper subgroup of $G$. Theorem 2.3 follows from the following theorem:

**Theorem 3.2.** Let $p > 5$ and $G$ simple. Assume that either of the following two is valid:

1) $G$ is not isomorphic to $\text{SL}(n)/Z$ for some subgroup $Z$ of the center $Z(\text{SL}(n))$ of $\text{SL}(n)$;
2) if $G$ is isomorphic to $\text{SL}(n)/Z$, where $Z \subseteq Z(\text{SL}(n))$, then $p$ does not divide $n$.

Let $E_G$ be a $G$-bundle over $X$ such that $E_G$ does not admit any reduction of structure group to any nontrivial Levi factor. Such a $G$-bundle $E_G$ admits a connection.

This theorem will be proved in the next section.

### 4. Obstruction for connection.

Let $G$ be a simple algebraic group over $k$. As in Theorem 3.2, assume that either of the following two is valid:

1) $G$ is not isomorphic to $\text{SL}(n)/Z$ for some subgroup $Z$ of the center $Z(\text{SL}(n))$ of $\text{SL}(n)$;
2) if $G$ is isomorphic to $\text{SL}(n)/Z$, where $Z \subseteq Z(\text{SL}(n))$, then $p$ does not divide $n$.

This assumption ensures that the Lie algebra $\mathfrak{g}$ of $G$ is isomorphic to $\mathfrak{g}^*$ as a $G$-module. Indeed, from [5, 0.13] we know that $\mathfrak{g}$ is simple. Now, as $\mathfrak{g}$ and $\mathfrak{g}^*$ are simple modules of same highest-weight, they are isomorphic [6, p. 200, Proposition 2.4(a)]. In other words, $\mathfrak{g}$ is self-dual.

Consequently, for $G$-bundle $E_G$ we have $\text{ad}(E_G) = \text{ad}(E_G)^*$. Now the Serre duality gives

$$H^1(X, K_X \otimes \text{ad}(E_G)) = H^0(X, \text{ad}(E_G))^*.$$  \hspace{1cm} (4.1)

Assume that $E_G$ satisfies the conditions in Theorem 3.2. Let

$$\tau \in H^1(X, K_X \otimes \text{ad}(E_G))$$

be the extension class for the Atiyah exact sequence for $E_G$. Let

$$\theta \in H^0(X, \text{ad}(E_G))^*$$

be the functional that corresponds to $\tau$ by the isomorphism (4.1). Theorem 3.2 will be proved by showing that the functional $\theta$ vanishes identically. For this we need to study the section of $\text{ad}(E_G)$.

For any $x \in X$ the fiber of $\text{ad}(E_G)$ over $x$ will be denoted by $\text{ad}(E_G)_x$. Note that $\text{ad}(E_G)_x$ is isomorphic, as a Lie algebra, with $\mathfrak{g}$. 


Lemma 4.1. Let \( \phi \) be a section of \( \text{ad}(E_G) \) such that for some point \( x \in X \), the evaluation \( \phi(x) \) is a nilpotent element of the Lie algebra \( \text{ad}(E_G)_x \). Then we have \( \theta(\phi) = 0 \).

Proof. We noted earlier that the assumption on \( G \) (stated at the beginning of this section) ensures that \( \mathfrak{g} \) is simple. Therefore, an element \( v \) of the simple Lie algebra \( \mathfrak{g} \) is nilpotent if \( \text{ad}(v) \) is nilpotent. If \( f \) is a \( G \)-invariant function on \( \mathfrak{g} \), then evaluating \( f \) on \( \phi \) we get a function on \( X \). Note that an element \( v \) of \( \mathfrak{g} \) is nilpotent if and only if all \( G \)-invariant functions on \( \mathfrak{g} \) vanishing at \( 0 \in \mathfrak{g} \) also vanishes at \( v \). Since \( X \) is connected and complete there are no nonconstant functions on \( X \). Consequently, if \( \phi \) is nilpotent over some point, this observation implies that \( \phi(y) \) is a nilpotent element of \( \text{ad}(E_G)_y \) for every point \( y \in X \).

Using \( \phi \) we will construct a reduction of structure group of \( E_G \) to a parabolic subgroup of \( G \). For that we will first construct a parabolic subalgebra bundle of \( \text{ad}(E_G) \).

Take any point \( y \in X \) such that \( \phi(y) \neq 0 \). Let \( V_y \) be the line in \( \text{ad}(E_G)_y \) generated by \( \phi(y) \). Let \( n^1_y \subset \text{ad}(E_G)_y \) be the normalizer of \( V_y \) and \( \tau^1_y \subset n^1_y \) be the nilpotent radical.

Now inductively define \( n^{i+1}_y \) to be the normalizer of \( \tau^i_y \) in \( \text{ad}(E_G)_y \) and \( \tau^{i+1}_y \) to be the nilpotent radical of \( n^{i+1}_y \).

Let \( n_y := \lim n^i_y \) and \( \tau_y := \lim \tau^i_y \) be the limits of these two increasing sequences. From the construction of the two sequences it is obvious that \( n_y \) is the normalizer, in \( \text{ad}(E_G)_y \), of \( \tau_y \). Also, \( \tau_y \) is the nilpotent radical of \( n_y \). Therefore, \( n_y \) is a parabolic subalgebra of \( \text{ad}(E_G)_y \). See [4, 30.3] for the details of this construction.

Consider the action of \( G \) on itself by inner conjugation. Let \( \text{Ad}(E_G) := (E_G \times G)/G \) be the gauge bundle (adjoint bundle) constructed using this action. Let \( P_y \subset \text{Ad}(E_G)_y \) be the parabolic subgroup of the fiber \( \text{Ad}(E_G)_y \) whose Lie algebra coincides with the parabolic subalgebra \( n_y \) constructed above.

Since there are only finitely many conjugacy classes of parabolic subalgebras of \( G \), there is a nonempty Zariski open subset \( U \) of \( X \) such that the conjugacy class of \( n_z \) is independent of \( z \in U \). Fix a parabolic subgroup \( P \) of \( G \) whose Lie algebra is in the same conjugacy class as \( n_z \), where \( z \in U \). To explain this with more details, we observe that the variety of nilpotent elements in \( \mathfrak{g} \) is irreducible. Indeed, it is the image of \( \mathfrak{g} \) by the Jordan decomposition. The variety of nilpotent elements in \( \mathfrak{g} \) is filtered by conjugacy classes. Therefore, on some nonempty Zariski open subset of \( X \), the evaluation of \( \phi \) must lie in some particular stratum of this filtered variety.

Consider the obvious projection

\[
q(y) : (E_G)_y \times G \longrightarrow \text{Ad}(E_G)_y.
\]
Let \((E_P)_y \subset (E_G)_y\) be the subvariety consisting all elements \(z\) such that \(q(z,g) \in P_z\) for every \(g \in P\). It is easy to check that \(E_P \subset E_G\) is a reduction of structure group over \(U\) of \(E_G\) to the parabolic subgroup \(P\) of \(G\). Indeed, this is an immediate consequence of the fact that the normalizer of \(P\) in \(G\) is \(P\) itself.

Since \(G/P\) is a complete variety and \(\dim X = 1\), the reduction over \(U\) extends to a reduction of structure group over \(X\) of \(E_G\) to \(P\). Let \(E_P \subset E_G\) denote this reduction of structure group.

Let \(\text{ad}(E_P) \subset \text{ad}(E_G)\) be the adjoint bundle. The commutativity of the diagram

\[
0 \rightarrow \text{ad}(E_P) \rightarrow \text{At}(E_P) \rightarrow TX \rightarrow 0
\]

ensures that the cohomology class \(\tau \in H^1(X, K_X \otimes \text{ad}(E_G))\) defined in (4.2) lies inside the image of \(H^1(X, K_X \otimes \text{ad}(E_P))\) for the homomorphism defined by the inclusion of \(\text{ad}(E_P)\) in \(\text{ad}(E_G)\).

Now, since \(\phi(y)\) is in the nilpotent radical \(\mathfrak{r}_y\) of the parabolic subalgebra \(\mathfrak{n}_y\), it follows that

\[\theta(\phi) = 0,\]

where \(\theta\) is defined in (4.3). Indeed, the subalgebra \(\mathfrak{r}_y\) is contained in the annihilator of \(\mathfrak{n}_y\) for each \(y \in X\). Now, \(\tau\) defined in (4.2) is in the image of \(H^1(X, K_X \otimes \text{ad}(E_P))\). This immediately implies that for the nondegenerate pairing

\[H^1(X, K_X \otimes \text{ad}(E_P)) \otimes H^0(X, \text{ad}(E_P)) \rightarrow k\]

defining the Serre duality, we have \(\tau \otimes \phi \mapsto 0\). In other words, \(\theta(\phi) = 0\). This completes the proof of the lemma.

In view of Lemma 4.1, to complete the proof of Theorem 3.2 it suffices to show that \(\theta(\phi) = 0\), where \(\phi\) is a everywhere semisimple section of \(\text{ad}(E_G)\). Indeed, by the Jordan decomposition theorem, any section \(\phi\) of \(\text{ad}(E_G)\) decomposes uniquely as \(\phi_s + \phi_n\), where \(\phi_s\) is everywhere semisimple and \(\phi_n\) is everywhere nilpotent. So, to prove that the \(\theta(\phi) = 0\), it is enough to show that \(\theta(\phi_s) = 0\) and \(\theta(\phi_n) = 0\).

We will show that \(\text{ad}(E_G)\) does not admit any nonzero section which is semisimple everywhere. This is the content of the following lemma:

**Lemma 4.2.** Let \(E_G\) be as in Theorem 3.2. Let \(\phi \in H^0(X, \text{ad}(E_G))\) be such that \(\phi(y)\) is a semisimple vector of \(\text{ad}(E_G)_y\) for every \(y \in X\). Then the section \(\phi\) vanishes identically.
Proof. Let $\phi$ be a nonzero section of $\text{ad}(E_G)$ which is semisimple everywhere. Since $X$ is connected and complete, the characteristic polynomial for the adjoint action of $\phi(y)$ on $\text{ad}(E_G)_y$ is independent of $y$. So $\phi$ does not vanish at point of $X$.

We have a decomposition
\begin{equation}
\text{ad}(E_G)_y = \bigoplus_{\lambda \in \Lambda} V_y^\lambda,
\end{equation}
where $V_y^\lambda$ is the eigenspace for the eigenvalue $\lambda$ for the adjoint action of $\phi(y)$ on $\text{ad}(E_G)_y$. So, $V_y^0$ coincides with the subalgebra of $\text{ad}(E_G)_y$ that centralizes $\phi(y)$.

Let $V_y \subset \text{ad}(E_G)_y$ denote the direct sum of all eigenspaces in (4.2) with eigenvalue less than or equal to zero in the lexicographic ordering. Since $[V_y^\lambda_1, V_y^\lambda_2] \subset V_y^{\lambda_1 + \lambda_2}$, unless $[V_y^\lambda_1, V_y^\lambda_2] = 0$, we have $V_y$ as a subalgebra of $\text{ad}(E_G)_y$. Note that the direct sum of all eigenspaces in (4.5) with eigenvalue strictly positive (in the lexicographic ordering) is a nilpotent subalgebra. In other words, $V_y$ has a complement which is nilpotent. Using [2, p. 473, Corollary 4.10], [3, p. 747, Lemma 4], it now follows that $V_y$ is a parabolic subalgebra of $\text{ad}(E_G)_y$ and $V_y^0$ its Levi factor.

Let $P_y \subset \text{Ad}(E_G)_y$ be the parabolic subgroup with $V_y$ as its Lie algebra. Let $H_y$ be the Levi factor of $P_y$ whose Lie algebra is $V_y^0$.

Let $\mathcal{W}_y$ denote the direct sum of all eigenspaces in (4.5) with eigenvalue greater than or equal to zero. Just as before, $\mathcal{W}_y$ is a parabolic subalgebra of $\text{ad}(E_G)_y$ with $V_y^0$ as its Levi factor.

Let $Q_y \subset \text{Ad}(E_G)_y$ be the parabolic subgroup whose Lie algebra is the direct sum of all eigenspaces in (4.5) with nonnegative eigenvalues. For the same reason as for $P_y$, the subgroup $H_y$ is a Levi factor of $Q_y$. Clearly we have $P_y \cap Q_y = H_y$.

It is easy to see that the conjugacy classes of $P_y$ and $H_y$ are independent of $y$. Fix subgroups $P$, $Q$ and $H$ in $G$ such that some identification of $G$ with $\text{Ad}(E_G)_y$ takes them to $P_y$, $Q_y$ and $H_y$ respectively. Note that $P$, $Q$ and $H$ have been fixed independent of $y$. So $P \cap Q = H$ and $H$ is a Levi factor for both $P$ and $Q$.

As in the Proof of Lemma 4.1, using $P_y$ we have a reduction of structure group $E_P \subset E_G$ to $P$. More precisely, let $(E_P)_y \subset (E_G)_y$ be the subvariety consists of all $z$ with such that $q(y)(z) \in P_y$ for every $g \in P$, where $q(y)$ is the projection in (4.4). Similarly, we have a reduction of structure group $E_Q \subset E_G$ to $Q \subset G$. Let
\[ E_H := E_P \cap E_Q \subset E_G \]
be the intersection of the two subvarieties $E_P$ and $E_Q$ of $E_G$. Clearly $E_H$ defines a reduction of structure group of $E_G$ to $P \cap Q = H$.

Recall the assumption in Theorem 3.2 that $E_G$ does not admit any reduction to a nontrivial Levi. Therefore, we have $H = G$. This immediately implies that $\phi = 0$. This completes the proof of the lemma.

Lemma 4.1 and Lemma 4.2 together complete the proof of Theorem 3.2. It was noted in Section 3 that Theorem 3.2 completes the proof of Theorem 2.3. Therefore, the proof of Theorem 2.3 is complete.

If $G$ is a classical group but not isomorphic to $\text{SL}(n)/Z$ for some subgroup $Z$ of the center $Z(\text{SL}(n))$ of $\text{SL}(n)$, then $g = g^*$ as a $G$-module if $p > 2$ [5, 0.13]. Therefore, Theorem 2.3 remains valid in this case.

We note that for $G = E_6$, the $G$-module $g$ fails to be isomorphic to $g^*$ if $p = 3$ [5, 0.13, p. 9]. For $G = E_7$, the $G$-module $g$ fails to be isomorphic to $g^*$ if $p = 2$ [5, 0.13, p. 9]. For classical groups of type $B_r$ and $C_r$, we have $g \neq g^*$ if $p = 2$ [5, 0.13].

Acknowledgments. We thank the referee for helpful comments that helped in improving the exposition.

References


Received April 25, 2001 and revised April 17, 2002.
School of Mathematics
Tata Institute of Fundamental Research
Homi Bhabha Road
Bombay 400005
India
E-mail address: indranil@math.tifr.res.in

School of Mathematics
Tata Institute of Fundamental Research
Homi Bhabha Road
Bombay 400005
India
E-mail address: subramnn@math.tifr.res.in
CUBIC SINGULAR MODULI, RAMANUJAN’S CLASS INVARIANTS $\lambda_n$ AND THE EXPLICIT SHIMURA RECIPROCITY LAW

HENG HUAT CHAN, VICTOR TAN, AND ALICE GEE

In this paper, we use the explicit Shimura Reciprocity Law to compute the cubic singular moduli $\alpha^*_n$, which are used in the constructions of new rapidly convergent series for $1/\pi$. We also complete a table of values for the class invariant $\lambda_n$ initiated by S. Ramanujan on page 212 of his Lost Notebook.

1. Introduction.

In his famous paper [26], S. Ramanujan offers several beautiful series representations for $1/\pi$, one of which is

$$4/\pi = \sum_{k=0}^{\infty} \frac{(6k+1)\left(\frac{1}{2}\right)_k^3}{(k!)^34^k},$$

where $(a)_0 = 1$ and for each positive integer $k$,

$$(a)_k = (a)(a+1)(a+2)\ldots(a+k-1).$$

Motivated by Ramanujan’s series, J.M. Borwein and P.B. Borwein [10] obtained many general representations for $1/\pi$. One generalization of (1.1) takes the form

$$\frac{1}{\pi} = \sum_{k=0}^{\infty} \left\{ \frac{\left(\frac{1}{2}\right)_k}{k!} \right\}^3 (a_n + b_nk)(G_n^{-12})^{2k},$$

where $n$ is a positive integer (usually odd) and $a_n, b_n$ and $G_n$ are certain special values of modular forms. It turns out that these special values can be expressed in terms of the singular modulus $\alpha_n$, which is defined to be the unique positive number between 0 and 1 satisfying

$$\frac{2F_1\left(\frac{1}{2}, \frac{1}{2}; 1; 1 - \alpha_n\right)}{2F_1\left(\frac{1}{2}, \frac{1}{2}; 1; \alpha_n\right)} = \sqrt{n}, \quad n \in \mathbb{Q},$$

where

$$2F_1(a, b; c; z) := \sum_{k=0}^{\infty} \frac{(a)_k(b)_k}{(c)_k} \frac{z^k}{k!}, \quad |z| < 1.$$
In his Notebooks, S. Ramanujan recorded many values of $\alpha_n$, one of which is

$$\alpha_3 = \frac{2 - \sqrt{3}}{4}.$$  

This value, when substituted into the Borweins’ formula (1.2) yields (1.1).\(^1\)

The proofs for all the singular moduli recorded in Ramanujan’s Notebooks can now be found in [9] and [6].

Recently, B.C. Berndt, S. Bhargava and F.G. Garvan [3] succeeded in developing theories of elliptic functions to alternative bases vaguely mentioned by Ramanujan in [26]. As indicated in [3], Ramanujan’s elliptic functions to alternative base 3 turns out to be the most interesting case of his theories. For this particular base, an analogue of the singular modulus, which we shall call “cubic singular modulus”, is defined as the unique positive number $\alpha^*_n$ between 0 and 1 such that

$$\frac{\text{2}_F_1(\frac{1}{3}, \frac{2}{3}; 1; 1 - \alpha^*_n)}{\text{2}_F_1(\frac{1}{3}, \frac{2}{3}; 1; \alpha^*_n)} = \sqrt{n}, \quad n \in \mathbb{Q}.$$ 

Although Ramanujan did not record any cubic singular moduli in his Notebooks or Lost Notebook, he must have computed some of them since these values (see [10]) are essential in his derivations of the series [26]

$$\frac{27}{4\pi} = \sum_{k=0}^{\infty} \left(2 + 15k\right) \frac{\left(\frac{1}{2}\right)_k \left(\frac{1}{3}\right)_k \left(\frac{2}{3}\right)_k}{(k!)^3} \left(\frac{2}{27}\right)^k$$

and

$$\frac{15\sqrt{3}}{2\pi} = \sum_{k=0}^{\infty} \left(4 + 33k\right) \frac{\left(\frac{1}{2}\right)_k \left(\frac{1}{3}\right)_k \left(\frac{2}{3}\right)_k}{(k!)^3} \left(\frac{4}{125}\right)^k.$$ 

The first discussion of the computations of the cubic singular moduli was given by the Borweins [10]. They determined $\alpha^*_n$ for $n = 2, 3, 4, 5$ and 6 from known values of Ramanujan-Weber class invariants $G_{3n}$ and $g_{6n}$, and deduced three new series for $1/\pi$ corresponding to $n = 2, 3$ and 6. Recently, Chan and Liaw [16] succeeded in evaluating $\alpha^*_n$ for $n = 2, 5, 7, 11$, and 23 using cubic Russell-type modular equations. From the values of $\alpha^*_7$ and $\alpha^*_{11}$, they discovered that when $3n$ is an Euler convenient number, $\alpha^*_n$ can be determined using Kronecker’s Limit Formula. Using these new $\alpha^*_n$’s, they derived many new series for $1/\pi$. Their method, however, cannot be extended to include the computations of $\alpha^*_n$ when $3n$ is not convenient.

In Sections 2 and 3, we use an explicit version of the Shimura Reciprocity Law to extend the list of $\alpha^*_n$. We show that when the class group of $\mathbb{Q}(\sqrt{-3n})$ is

\(^1\)The determination of $a_n$ from $\alpha_n$ is very challenging. It involves modular equations of degrees dividing $n$. 
takes the form $\mathbb{Z}_2^t \oplus \mathbb{Z}_k$, with $t \in \mathbb{N}$ and $k = 4, 6$ and $8$, $\alpha_n^*$ can be determined explicitly.

On page 212 of his Lost Notebook, Ramanujan defined a certain function $\lambda_n$ (see (3.1)) and recorded its values for $n = 1, 9, 17, 25, 33, 41, 49, 73, 97, 121$. He also indicated that he could compute $\lambda_n$ when $n = 57, 65, 81, 89, 169, 193, 217, 241, 265, 289, 361$ but did not supply any values for these $n$’s. Using cubic Russell-type modular equations, Kronecker’s Limit Formulas and other techniques, Berndt, Chan, S.-Y. Kang and L.-C. Zhang [7] provided proofs of all these values except for $n = 73, 97, 193, 217, 241$. In Section 4, we modify our method in Sections 2 and 3 and determine rigorously these remaining values of $\lambda_n$.

2. Some properties of $\alpha_n^*$.

Let
\begin{equation}
\eta(\tau) := q^{1/24} \prod_{k=1}^{\infty} (1 - q^k), \quad \text{where } q = e^{2\pi i \tau} \text{ with } \text{Im}\, \tau > 0,
\end{equation}
and
\begin{equation}
\mu_n = \frac{1}{3\sqrt{3}} \left\{ \frac{\eta\left(\sqrt{-n/3}\right)}{\eta\left(\sqrt{-3n}\right)} \right\}^6, \quad n \in \mathbb{Q}.
\end{equation}
The relation between $\mu_n$ and the cubic singular moduli $\alpha_n^*$ is given by [17]
\begin{equation}
\frac{1}{\alpha_n^*} = 1 + \mu_n^2.
\end{equation}
Identity (2.3) shows that in order to determine $\alpha_n^*$, it suffices to compute $\mu_n^2$. First, we need the following:

**Theorem 2.1.** Suppose that $n$ is squarefree so that $-12n$ is a fundamental imaginary quadratic discriminant. Then $\mu_n^2$ is a real unit contained in $K_1$, the Hilbert class field of $K := \mathbb{Q}(\sqrt{-3n})$.

To prove Theorem 2.1, we need the following lemmas:

**Lemma 2.2** ([23, p. 159, Corollary]). Let $K$ be as defined in Theorem 2.1, and let $\mathcal{O}_K$ be the ring of integers of $K$. Let $a = [\tau_1, \tau_2]$ be an $\mathcal{O}_K$-ideal and define
\begin{equation}
\Delta(a) := \tau_2^{-12} \eta^{24}(\tau),
\end{equation}
where $\tau = \tau_1/\tau_2$ with $\text{Im}\, \tau > 0$. Then the value $\Delta(a)/\Delta(\mathcal{O}_K)$ lies in $K_1$, where $K_1$ is the Hilbert class field of $K$. 

Lemma 2.3 ([23, p. 166, Corollary]). Let $N(a)$ denote the index $(\mathcal{O}_K : a)$
where $a$ is an $\mathcal{O}_K$-ideal. Then the number
\[
N(a)\frac{|\Delta(a)|^2}{|\Delta(\mathcal{O}_K)|^2}
\]
is a unit.

Lemma 2.4. Recall that for $\tau \in \mathbb{C}$, with $\text{Im} \tau > 0$, the $j$-function is defined
by
\[
j(\tau) = 1728 \frac{g_3^3(\tau)}{g_2^3(\tau) - 27g_3^2(\tau)},
\]
with $g_2$ and $g_3$ given by
\[
g_2(\tau) = 60 \sum_{m,n=-\infty}^{\infty} (m + n\tau)^{-4}, \quad \text{and}
\]
\[
g_3(\tau) = 140 \sum_{m,n=-\infty}^{\infty} (m + n\tau)^{-6}.
\]
If
\[
g(\tau) := \frac{1}{3\sqrt{3}} \left( \frac{\eta(\tau)}{\eta(3\tau)} \right)^6,
\]
then
\[
j(\tau) = 27 \frac{(1 + g^2(\tau))(9 + g^2(\tau))^3}{g^6(\tau)}.
\] (2.5)

Lemma 2.4 follows from the fact that $g^2(\tau)$ generates the function field
associated with the group $\Gamma_0(3)$, which implies the $j(\tau)$ is a rational function
of $g^2(\tau)$. For a more elementary proof of this lemma using Ramanujan’s
identities, see [14] and [5].

Proof of Theorem 2.1. Let $a = [3, \sqrt{-3n}]$ with $n \equiv 3 \pmod{4}$. By (2.4),
\[
\mu_n^4 = 3^{-12} \frac{\eta^{24}(\sqrt{-n/3})}{\eta^{24}(\sqrt{-3n})} = \frac{\Delta(a)}{\Delta(\mathcal{O}_K)} = N(a)^6 = N(a)^6 \frac{\Delta(a)}{|\Delta(\mathcal{O}_K)|}.
\] (2.6)

From the second equality of (2.6) and Lemma 2.2, we find that $\mu_n^4$ belongs
to $K_1$ and from the last equality of (2.6) and Lemma 2.3, we conclude that
$\mu_n^2$ is a real unit. To complete the proof of Theorem 2.1, it remains to show
that $\mu_n^2$ is in $K_1$.

Now, when $\tau = \sqrt{-n/3}$, $g(\tau) = \mu_n$ and
\[
\mu_n^8 + 270\mu_n^4 + 3^6 = ((j(\sqrt{-n/3})/27 - 28)\mu_n^4 - 972)\mu_n^2,
\]
by Lemma 2.4. Since both $\mu_n^2$ and $j(\sqrt{-n/3})$ are in $K_1$, we can conclude that $\mu_n^2 \in K_1$ unless $(j(\sqrt{-n/3})/27 - 28)\mu_n^4 - 972 = 0$. If this is the case, we can deduce that $j(\sqrt{-n/3})$ satisfies the quadratic equation

$$X^2 + 8208X - 5832000 = 0.$$ 

But the two roots of this equation has numerical values 657.8 and -8865.8. This contradicts the fact that $j(\sqrt{-m}) \geq 1728$ for any $m \geq 1$. This completes the Proof of Theorem 2.1. □

A class invariant $\gamma$ of a field $K$ is defined to be a generator for the Hilbert class field of $K$, i.e., $K_1 = K(\gamma)$. Theorem 2.1, Lemma 2.4 and the fact that $j(\sqrt{-3n})$ is a class invariant [20] imply that $\mu_n^2$ is a class invariant of $\mathbb{Q}(\sqrt{-3n})$ when $n \equiv 3 \pmod{4}$. Hence, we conclude that $\alpha_n^*$ is also a class invariant of $\mathbb{Q}(\sqrt{-3n})$ by Theorem 2.1 and (2.3).

We remark here that our result given in this section is not “optimal”. We have shown that $\mu_n^2$ is a class invariant whenever $3 \nmid n$ and $n$ squarefree. It is possible to show further that smaller powers of the $\eta$-quotients given in the definition of $\mu_n^2$, namely, $\eta(\sqrt{-n/3})^s \eta(\sqrt{-3n})$, with $s|12$ and $s < 12$, is a class invariant if we impose further congruence conditions on $n$. This can be established using Gee’s results [21, Section 5].

3. The explicit Shimura reciprocity law and new values of $\alpha_n^*$.

We have seen in Section 2 that $\mu_n^2$ is a class invariant whenever $n$ satisfies the hypothesis of Theorem 2.1. In this section, we identify $\mu_n^2$ as a value of a modular function, construct the explicit action of $\text{Gal}(K_1|K)$ on $\mu_n^2$ and as a result, evaluate $\mu_n^2$.

Let $\mathbb{M}_2^+(\mathbb{Z})$ denote the set of $2 \times 2$ matrices with integer coefficients and positive determinant. For each $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{M}_2^+(\mathbb{Z})$, define the function

$$\eta \circ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : \tau \mapsto \frac{\eta(a\tau + b)}{\eta(c\tau + d)} \eta(\tau).$$

It is easy to see that $\mu_n$ is the value of $g_0(\tau)^6/(3\sqrt{3})$ at $\tau = \sqrt{-3n}$ where

$$g_0(\tau) := \frac{\eta \circ \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}}{\eta}(\tau).$$

The function $g_0(\tau)$ is an element of $F_{72}$, the modular function field of level 72 defined over $\mathbb{Q}(\zeta_{72})$. This means that it is meromorphic on the completed upper half plane $\mathbb{H} \cup \mathbb{Q} \cup \{\infty\}$, admits a Laurent series expansion in the
variable $q^{1/72} = e^{2\pi i q/72}$ centered at $q = 0$ having coefficients in $\mathbb{Q}(\zeta_{72})$ and invariant with respect to the matrix group

$$\Gamma(72) := \ker[SL_2(\mathbb{Z}) \to SL_2(\mathbb{Z}/72\mathbb{Z})].$$

From (2.5), we find that the minimal polynomial for $g_{12}^{0}$ over the modular function field $\mathbb{Q}(j)$ is

$$X^4 + 36 X^3 + 270 X^2 + (756 - j)X + 36.$$

Over $\mathbb{Q}(j)$, the conjugates of $g_{12}^{0}$ are $g_{12}^{1}$, $g_{12}^{2}$ and $g_{12}^{3}$ defined by

$$g_1 := \zeta_{24}^{-1} \frac{\eta \circ \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix}}{\eta}, \quad g_2 := \frac{\eta \circ \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}}{\eta}, \quad \text{and} \quad g_3 := \sqrt{3} \frac{\eta \circ \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}}{\eta}.$$

If $K$ is an imaginary quadratic field of discriminant $D$, class field theory gives an isomorphism

$$\text{Gal}(K_1/K) \simeq C(D)$$

between the Galois group for $K \subset K_1$ and the form class group of discriminant $D$. Among the primitive forms $[a,b,c]$ having discriminant $D = b^2 - 4ac$, one obtains a complete set of representatives in $C(D)$ by choosing the reduced forms

$$|b| \leq a \leq c \quad \text{and} \quad b \geq 0 \text{ if either } |b| = a \text{ or } a = c.$$ 

The class of $[a,-b,c]$ is the inverse of $[a,b,c]$ in $C(D)$, and the elements having order 2 in $C(D)$ correspond to ambiguous forms. These are the reduced forms $[a,b,c]$ for which $a = b$, $a = c$ or $b = 0$ occurs.

Given $h \in F_m$, if $h(\theta) \in K_1$ where $\theta$ is the generator of $\mathcal{O}_K$ over $\mathbb{Z}$ (we assume here the algebraic closure of $K$ is embedded in the complex plane such that $\theta$ lies in the upper half plane $\mathbb{H}$), there is an explicit formula for computing the action of $C(D)$ on $h(\theta)$ which is a consequence of the Shimura Reciprocity Law. This is given as follows:

**Lemma 3.1.** Let $K$ be an imaginary quadratic number field of discriminant $D$ and let $h \in F_m$ be such that $h(\frac{\sqrt{D}}{2}) \in K_1$. Given a primitive quadratic form $[a,b,c]$ of discriminant $D$, let $M = M_{[a,b,c]} \in GL_2(\mathbb{Z}/m\mathbb{Z})$ be the matrix that satisfies the congruences

$$M \equiv \begin{cases} \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \pmod{p^r} & \text{if } p \nmid a, \\
\begin{pmatrix} -b & -c \\ 1 & 0 \end{pmatrix} \pmod{p^r} & \text{if } p \mid a \text{ and } p \nmid c, \\
\begin{pmatrix} -b - a & -b - c \\ 1 & -1 \end{pmatrix} \pmod{p^r} & \text{if } p \mid a \text{ and } p \mid c. \end{cases}$$
at all prime power factors \( p^r \mid m \). The Galois action of the class of \([a, -b, c]\) in \( C(D)\) with respect to the Artin map is given by

\[
\begin{pmatrix} h \left( \frac{\sqrt{D}}{2} \right) \end{pmatrix}^{[a, -b, c]} = h^M \left( \frac{-b + \sqrt{D}}{2a} \right),
\]

where \( h^M \) denote the image of \( h \) under the action of \( M \).

For a proof of Lemma 3.1 and the description of the action of \( M \) on \( h \), see [21].

In view of Lemma 3.1, we first need to discuss the action of \( M \in GL_2(\mathbb{Z}/m\mathbb{Z}) \) on functions \( h \in F_m \). The action of such an \( M \) depends only on \( M_{p^r} \) for all prime factors \( p \mid m \) where \( M_N \in GL_2(\mathbb{Z}/N\mathbb{Z}) \) is the reduction modulo \( N \) of \( M \) and \( r_p \) is the largest power of \( p \) such that \( p^{r_p} \) divides \( m \).

Now every \( M_N \) with determinant \( x \) decomposes as

\[
M_N = \begin{pmatrix} 1 & 0 \\ 0 & x \end{pmatrix}_N \left( \begin{array}{cc} a & b \\ c & d \end{array} \right)_N
\]

for some \( \left( \begin{array}{cc} a & b \\ c & d \end{array} \right)_N \in SL_2(\mathbb{Z}/N\mathbb{Z}) \). Since \( SL_2(\mathbb{Z}/N\mathbb{Z}) \) is generated by \( S_N \) and \( T_N \) where \( S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \) and \( T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \), it suffices to find the action of \( \left( \begin{array}{cc} 1 & 0 \\ 0 & x \end{array} \right)_{p^{r_p}} \), \( S_{p^{r_p}} \) and \( T_{p^{r_p}} \) on \( h \) for all \( p \mid m \).

For \( \left( \begin{array}{cc} 1 & 0 \\ 0 & x \end{array} \right)_{p^{r_p}} \), the action on \( F_m \) is given by lifting the automorphism of \( \mathbb{Q}(\zeta_m) \) determined by

\[
\zeta_{p^{r_p}} \mapsto \zeta_{p^{r_p}}^x \quad \text{and} \quad \zeta_{q^{r_q}} \mapsto \zeta_{q^{r_q}}
\]

for all prime factors \( q \mid m \) such that \( q \neq p \).

In order that the actions of the matrices at different primes commute with each other, we have to lift \( S_{p^{r_p}} \) and \( T_{p^{r_p}} \) to matrices in \( SL_2(\mathbb{Z}/m\mathbb{Z}) \) such that they reduce to the identity matrix in \( SL_2(\mathbb{Z}/q^{r_q}\mathbb{Z}) \) for all \( q \neq p \). In our case for \( m = 72 \), the prime powers are 8 and 9 and we have

\[
S_8 \mapsto \begin{pmatrix} -8 & 9 \\ -9 & -8 \end{pmatrix}_{72}, \quad T_8 \mapsto \begin{pmatrix} 1 & 9 \\ 0 & 1 \end{pmatrix}_{72},
\]

\[
S_9 \mapsto \begin{pmatrix} 9 & -8 \\ 8 & 9 \end{pmatrix}_{72}, \quad T_9 \mapsto \begin{pmatrix} 1 & -8 \\ 0 & 1 \end{pmatrix}_{72}.
\]

When \( h \in F_m \) is an \( \eta \)-quotient, we can use the transformation rule

\[
\eta \circ S_m(\tau) = \sqrt{-i\tau}\eta(\tau) \quad \text{and} \quad \eta \circ T_m(\tau) = \zeta_{24}\eta(\tau)
\]

to determine the action of any \( M_m \in SL_2(\mathbb{Z}/m\mathbb{Z}) \). In particular, we have

\[
(\zeta_{10}^0, \zeta_{10}^1, \zeta_{10}^2, \zeta_{10}^3) \circ S_{72} = (\zeta_{14}^0, \zeta_{14}^1, \zeta_{14}^2, \zeta_{14}^3, \zeta_{14}^0, \zeta_{14}^1, \zeta_{14}^2, \zeta_{14}^3).
\]
and
\[(g_0, g_1, g_2, g_3) \circ T_{72} = (g_1, \zeta_{24}^2 g_2, g_0, \zeta_{24}^2 g_3).\]
Consequently, we derive the following actions:

<table>
<thead>
<tr>
<th></th>
<th>$g_0^{12}$</th>
<th>$g_1^{12}$</th>
<th>$g_2^{12}$</th>
<th>$g_3^{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(1\ 0)$</td>
<td>$g_0^{12}$</td>
<td>$g_1^{12}$</td>
<td>$g_2^{12}$</td>
<td>$g_3^{12}$</td>
</tr>
<tr>
<td>$S_8$</td>
<td>$g_0^{12}$</td>
<td>$g_1^{12}$</td>
<td>$g_2^{12}$</td>
<td>$g_3^{12}$</td>
</tr>
<tr>
<td>$T_8$</td>
<td>$g_0^{12}$</td>
<td>$g_1^{12}$</td>
<td>$g_2^{12}$</td>
<td>$g_3^{12}$</td>
</tr>
<tr>
<td>$(1\ 0)$, $3</td>
<td>(x - 1)$</td>
<td>$g_0^{12}$</td>
<td>$g_1^{12}$</td>
<td>$g_2^{12}$</td>
</tr>
<tr>
<td>$S_9$</td>
<td>$g_3^{12}$</td>
<td>$g_2^{12}$</td>
<td>$g_1^{12}$</td>
<td>$g_0^{12}$</td>
</tr>
<tr>
<td>$T_9$</td>
<td>$g_1^{12}$</td>
<td>$g_2^{12}$</td>
<td>$g_0^{12}$</td>
<td>$g_3^{12}$</td>
</tr>
</tbody>
</table>

Using this, together with Lemma 3.1, we have:

**Theorem 3.2.** The action of a reduced primitive quadratic form $[a, b, c]$ with discriminant $D$ in $C(D)$ on $g_0(\sqrt{D/2})^{12}$ is given by

\[
\begin{cases}
  g_0(\sqrt{D/2})^{12} & [a, b, c] \\
  g_0(-b+\sqrt{D}/2a)^{12} & \text{if } b \equiv 0, a \not\equiv 0 \pmod{3}, \\
  g_1(-b+\sqrt{D}/2a)^{12} & \text{if } ab \equiv -1 \pmod{3}, \\
  g_2(-b+\sqrt{D}/2a)^{12} & \text{if } ab \equiv 1 \pmod{3}, \\
  g_3(-b+\sqrt{D}/2a)^{12} & \text{if } a \equiv 0 \pmod{3}.
\end{cases}
\]

**Proof.** The above result follows from the observation that the action of $M_8$ on $g_0^{12}$ is trivial. Hence, it suffices to consider the action of $M_9$ on $g_0^{12}$. When $3 \nmid a$,

\[
M_9 = \begin{pmatrix} a & b-1 \\ 0 & 2a \end{pmatrix} \equiv \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & b-1 \\ 0 & 2a \end{pmatrix} \equiv S_9 \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix} S_9 T_9^{b-1/2a}.
\]

When $3 \mid a$, then $3 \nmid c$, so

\[
M_9 = \begin{pmatrix} -b-1/2 & -c \\ 1 & 0 \end{pmatrix} \equiv \begin{pmatrix} c & -b-1/2 \\ 0 & 1 \end{pmatrix} S_9 \equiv \begin{pmatrix} c & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -b-1/2 \\ 0 & 2c \end{pmatrix} S_9 T_9^{b-1/2c} S_9.
\]

\[\square\]
Theorem 3.2 should be viewed as a cubic analogue of the results of N. Yui and D. Zagier [28, Proposition, Section 2] and it indicates that all the conjugates of \( \mu_n^2 \) can be computed numerically once we determine the class group of \( \mathbb{Q}(\sqrt{-3n}), n \equiv 3 \mod 4 \). Using these numerical values, we could then determine the minimal polynomial satisfied by \( \mu_n^2 \). If the degree of the minimal polynomial is at most 4, we could solve the minimal polynomial and determine \( \mu_n^2 \) explicitly. In order to calculate \( \mu_n^2 \) for which the class number of \( \mathbb{Q}(\sqrt{-3n}) \) is greater than 4, we need the following lemma, which essentially tells us the action of the ambiguous forms (see the remarks before Lemma 3.1 for the definition of ambiguous forms) on \( \mu_n^2 \).

**Lemma 3.3.** Let \( n \equiv 3 \mod 4 \) and \( K = \mathbb{Q}(\sqrt{-3n}) \), where \( n \) is squarefree. Then

\[
(\mu_n^2)^{[2,2,3n+1]} = -\lambda_n^2,
\]

where

\[
\lambda_n = \frac{1}{3\sqrt{3}} \begin{pmatrix}
\eta \left( \frac{1 + \sqrt{-n/3}}{2} \right) \\
\eta \left( \frac{1 + \sqrt{-3n}}{2} \right)
\end{pmatrix}^6.
\]

If \( n = p_1 p_2 \ldots p_k \) then

\[
(\mu_n^2)^{[p_1 p_2 \ldots p_j,0,\frac{3n}{p_1 p_2 \ldots p_j}]} = \mu_n^2/(p_1 p_2 \ldots p_j)^2,
\]

where \( j \leq k \).

**Proof.** We apply Theorem 3.2 with \( ab \equiv 1 \mod 3 \) and \( b \equiv 0 \mod 3 \), respectively and note that

\[
\lambda_n^2 = -\frac{1}{27} g_1^2 \left( \frac{1 + \sqrt{-3n}}{2} \right)
\]

and

\[
\mu_n^2/(p_1 p_2 \ldots p_j)^2 = \frac{1}{27} g_0^2 \left( \sqrt{\frac{3n}{(p_1 p_2 \ldots p_j)^2}} \right).
\]

We can now explicitly determined \( \mu_n^2 \) by first collecting in a symmetric way the products of the real conjugates of \( \mu_n^2 \).

For example, when \( n = 23 \), \( C(-276) \simeq \mathbb{Z}_2 \oplus \mathbb{Z}_4 \), generated by \( a = [2,2,35] \) and \( b = [5,2,14] \). Now define

\[
P_{23} := (\mu_{23} \lambda_{23})^2 + (\mu_{23} \lambda_{23})^{-2} \quad \text{and} \quad Q_{23} := (\mu_{23}/\lambda_{23})^2 + (\mu_{23}/\lambda_{23})^{-2}.
\]

These numbers are fixed by the Galois action of \( a^2 \) and \( b \) and since \( P_{23} \) and \( Q_{23} \) are algebraic integers, one concludes that \( P_{23} + P_{23}^a, P_{23} P_{23}^a, Q_{23} + Q_{23}^a, \) and \( Q_{23}Q_{23}^a \) are integers. These integers can be found by approximating...
the numerical values of the $g_{12}^{12}$ at the corresponding arguments given by Theorem 3.2. Hence, we obtain
\[ P_{23} + P_{23}^a = 296143772, \]
\[ P_{23}P_{23}^a = -389054012, \]
\[ Q_{23} + Q_{23}^a = 5980, \]
and
\[ Q_{23}Q_{23}^a = -17852. \]
Solving the quadratic polynomials satisfied by $P_{23}$ and $Q_{23}$ and simplifying, we deduce that
\[ \mu_{23}^2 = \left(5\sqrt{3} + 24\right)^{1/2} \left(13\sqrt{23} + 36\sqrt{3}\right)^{1/2} \left(\sqrt{84 + 48\sqrt{3}} + \sqrt{83 + 48\sqrt{3}}\right)^3. \]
Substituting the value $\mu_{23}^2$ into (2.3), we easily determine $\alpha_{23}^*$, which is crucial in the derivation of the following series.
\[ \frac{1}{\pi} = \sum_{m=0}^{\infty} \left( a_{23} + b_{23}m \right) \frac{\left(\frac{1}{2}\right)_m \left(\frac{2}{3}\right)_m (m!)^3}{H_{23}^m}. \]
\[ z_{23} = \frac{1}{23} \left( \sqrt{-83 + 48\sqrt{3}(444 + 252\sqrt{3}) - 56 + 54\sqrt{3}} \right), \]
\[ a_{23} = -\frac{1}{6\sqrt{3}} \left( z_{23} + (8\alpha_{23}^* - 4)\sqrt{23} \right), \]
\[ b_{23} = \frac{2\sqrt{23} \mu_{23}^2 - 1}{\sqrt{3} \mu_{23}^2 + 1}, \quad \text{and} \]
\[ H_{23} = \frac{1}{24^{23^3}} \left( 6\sqrt{-83 + 48\sqrt{3}} + 9\sqrt{3}\sqrt{-83 + 48\sqrt{3}} - 2 - 3\sqrt{3} \right)^3. \]
For methods of deriving series of the above type, and the relation between $\mu_n^2$ and series for $1/\pi$, see [17] and [18].

**Remarks.**

(a) The method illustrated above for the case $n = 23$ works for any $n$ such that $C(-12n)$ is of the type $\mathbb{Z}_2 \oplus \mathbb{Z}_{2s}$, where $s = 1, 2, 3$, or 4.

(b) If $C(-12n)$ is of the type $\mathbb{Z}_2^t \oplus \mathbb{Z}_{2s}$ with $s = 1, 2, 3$ or 4 and $t \in \mathbb{N}$, we need to construct more numbers analogous to $P_n$ and $Q_n$. Examples of such constructions can be found in [15] and Section 4.

(c) One can modify the method in [15] to evaluate the corresponding $\mu_n^2$ whenever the class group is of the form $\mathbb{Z}_2^t \oplus \mathbb{Z}_{2s}$, where $s = 2, 3$, or 4. The method used there avoids the use of the explicit Shimura Reciprocity Law but it cannot be extended to compute $\mu_n^2$ when the associated class groups are different from those mentioned above.
(d) Gee and M. Honsbeek [22] have recently devised a method of computing class invariants without solving their minimal polynomials. Their method involves determining the Lagrange resolvents of these minimal polynomials by determining the conjugates of the corresponding class invariants explicitly.

4. The class invariant $\lambda_n^2$ and the missing entries in the Lost Notebook.

We first note that

$$\lambda_n^2 = \frac{1}{27} g_2^{12} \left( \frac{-1 + \sqrt{-3n}}{2} \right).$$

To compute $\lambda_n$, it suffices to determine the action of the elements in the corresponding class groups. This is given by the following analogue of Theorem 3.2:

**Theorem 4.1.** The action of a reduced primitive quadratic form $[a,b,c]$ with discriminant $D$ in $C(D)$ on $g_2 \left( \frac{-1+\sqrt{D}}{2} \right)^{12}$ is given by

$$g_2 \left( \frac{-1+\sqrt{D}}{2} \right)^{12} [a,-b,c] = \begin{cases} g_0 \left( \frac{-b+\sqrt{D}}{2a} \right)^{12} & \text{if } b \equiv 0, a \not\equiv 0 \pmod{3}, \\ g_1 \left( \frac{-b+\sqrt{D}}{2a} \right)^{12} & \text{if } ab \equiv -1 \pmod{3}, \\ g_2 \left( \frac{-b+\sqrt{D}}{2a} \right)^{12} & \text{if } ab \equiv 1 \pmod{3}, \\ g_3 \left( \frac{-b+\sqrt{D}}{2a} \right)^{12} & \text{if } a \equiv 0 \pmod{3}. \end{cases}$$

To facilitate the computations of $\lambda_n$ we need the analogue of Lemma 3.3.

**Lemma 4.2.** Let $n \equiv 1 \pmod{4}$ and $K = \mathbb{Q}(\sqrt{-3n})$, where $n$ is squarefree. If $n = p_1 p_2 \ldots p_k$, then

$$(\lambda_n^2)^{p_1 p_2 \ldots p_j} = \lambda_n^2 (p_1 p_2 \ldots p_j)^{3n + (p_1 p_2 \ldots p_j)^2},$$

where $j \leq k$.

Lemma 4.2 indicates that instead of calculating the conjugates of $\lambda_n^2$ directly, it suffices to calculate the conjugates of symmetric combinations of all the real conjugates of $\lambda_n^2$.

We may now proceed to complete the table of $\lambda_n$ initiated by Ramanujan on page 212 of his Lost Notebook. For $p = 73, 97$, and 241, all of which are primes, set

$$(4.1) \quad P_p = \lambda_p^2 + \frac{1}{\lambda_p^2}. $$

Since the class groups corresponding to these $p$’s are of the form $\mathbb{Z}_4$, we conclude that $P_p$ each satisfies a quadratic polynomial. We now derive the polynomial satisfied by $P_{73}$. 
Now the class group of $\mathbb{Q}(\sqrt{-219})$ is generated by the form $[5, 1, 11]$. By Theorem 4.1, we easily deduce that

$$P_{73} + P_{73}^{[5,1,11]} = 199044,$$

and

$$P_{73}P_{73}^{[5,1,11]} = 287491,$$

where $P_{73}^{[5,1,11]}$ denotes the image of $P_{73}$ under the action of $[5, 1, 11]$. Hence, $P_{73}$ satisfies the quadratic polynomial

$$x^2 - 199044x + 287491 = 0.$$

Solving and simplifying, we deduce that

$$\lambda_{73} = \left(\sqrt{\frac{11 + \sqrt{73}}{8}} + \sqrt{\frac{3 + \sqrt{73}}{8}}\right)^6.$$

The cases for $n = 97$ and 241 are similar.

We now turn to the case $n = 217$. Here 217 is divisible by two primes, namely, 7 and 31. In this case we consider two numbers $Q_{217}$ and $R_{217}$ defined by

$$Q_{217} = \lambda_{217}^2\lambda_{31/7}^2 + \frac{1}{\lambda_{217}^2\lambda_{31/7}^2}$$

and

$$R_{217} = \frac{\lambda_{217}^2}{\lambda_{31/7}^2} + \frac{\lambda_{31/7}^2}{\lambda_{217}^2}.$$

Note that the class group of $\mathbb{Q}(\sqrt{-651})$ is generated by $a := [5, 3, 33]$ and $b := [3, 3, 55]$. The order of $a$ is 4 and the group generated by $a^2$ and $b$ fixes $Q_{217}$ and $R_{217}$. Hence it suffices to determine the action of $a$ on $Q_{217}$ and $R_{217}$, which can be easily done by Theorem 4.1. The value of $\lambda_{217}$ which results from this consideration is a product of two units, given by

$$\lambda_{217} = \left(\sqrt{\frac{1901 + 129\sqrt{217}}{8}} + \sqrt{\frac{1893 + 129\sqrt{217}}{8}}\right)^{3/2} \cdot \left(\sqrt{\frac{1597 + 108\sqrt{217}}{4}} + \sqrt{\frac{1593 + 108\sqrt{217}}{4}}\right)^{3/2}.$$

Finally, consider the case $n = 193$. This is the case which we cannot evaluate using the previous method given in [15]. Here the class group of $\mathbb{Q}(\sqrt{-579})$ is generated by $a := [5, 1, 29]$ and it is of order 8. We consider the
expression $P_{193}$ where $P_p$ is given by (4.1). To determine $P_{193}$ we compute the image of $P_{193}$ under $a, a^2$, and $a^3$. Our computations show that if $\alpha := P_{193}$,
\[ \beta := P_{193}^a = -\frac{1}{27}g_2^{12} \left( \frac{1 + \sqrt{-579}}{10} \right) - 27g_2^{-12} \left( \frac{1 + \sqrt{-579}}{10} \right) \]
\[ \gamma := P_{193}^{a^2} = -\frac{1}{27}g_0^{12} \left( \frac{3 + \sqrt{-579}}{14} \right) - 27g_0^{-12} \left( \frac{3 + \sqrt{-579}}{14} \right) \]
and
\[ \delta := P_{193}^{a^3} = -\frac{1}{27}g_0^{12} \left( \frac{-9 + \sqrt{-579}}{22} \right) - 27g_0^{-12} \left( \frac{-9 + \sqrt{-579}}{22} \right) \]
then
\[ \alpha + \beta + \gamma + \delta = 3251132424, \]
\[ \alpha\beta + \alpha\gamma + \alpha\delta + \beta\gamma + \beta\delta + \gamma\delta = 82707128352, \]
\[ \alpha\gamma + \alpha\beta + \beta\gamma + \alpha\beta + \beta\gamma + \alpha\gamma = 9465475096, \]
and
\[ \alpha\beta\gamma\delta = 176664526832. \]

Solving the quartic polynomial satisfied by $P_{193}$ and simplifying, we deduce that

\[ \lambda_{193}^{1/3} + \frac{1}{\lambda_{193}^{1/3}} = \frac{1}{4} \left( 39 + 3\sqrt{193} + \sqrt{2690 + 194\sqrt{193}} \right). \]

It was not clear to us what motivated Ramanujan to construct the table of values for $\lambda_n$. Perhaps he intended to set up a table for $\lambda_n$ similar to that for the Ramanujan-Weber class invariants $G_n$ and $g_{2n}$ (see [26]). Recently, Chan, Liaw and Tan offered another reason for the existence of Ramanujan’s table. They succeeded in deriving a new class of series for $1/\pi$ associated with $\lambda_n$. Two of such series are

\[ \frac{4}{\pi\sqrt{3}} = \sum_{k=0}^{\infty} (5k + 1) \left( \frac{1}{3} \right)_k \left( \frac{2}{3} \right)_k \left( \frac{1}{2} \right)_k \left( -\frac{9}{16} \right)^k, \]

and

\[ \frac{12\sqrt{3}}{\pi} = \sum_{k=0}^{\infty} (51k + 7) \left( \frac{1}{3} \right)_k \left( \frac{2}{3} \right)_k \left( \frac{1}{2} \right)_k \left( -\frac{1}{16} \right)^k. \]

These simple series came as a surprise as it was thought that all the possible simple series should have been exhausted after the work of Ramanujan, the Chudnovskys [19] and the Borweins.
References


Received April 13, 2001 and revised November 7, 2001. This research is funded by Academic Research Fund from National University of Singapore, project number: R-146000007112. The first author wishes to thank W. Galway for bringing A. Gee’s work to his attention.

Department of Mathematics  
National University of Singapore  
Kent Ridge, Singapore 119260  
Republic of Singapore  
E-mail address: matchh@nus.edu.sg

Department of Mathematics  
National University of Singapore  
Kent Ridge, Singapore 119260  
Republic of Singapore  
E-mail address: mattanv@nus.edu.sg

Department of Mathematics  
University of Amsterdam  
The Netherlands  
E-mail address: gee@wins.uva.nl
SUMS OF PRODUCTS OF GENERALIZED BERNOULLI POLYNOMIALS

Kwang-Wu Chen

In this paper, we investigate the zeta function
\[
Z(P, \chi, a, s) = \sum_{n_1=1}^{\infty} \cdots \sum_{n_r=1}^{\infty} \chi_1(n_1) \cdots \chi_r(n_r) \\
\cdot P(n_1 + a_1, \ldots, n_r + a_r)^{-s},
\]
where \(a_i \geq 0\), \(\chi_i\) is a Dirichlet character with conductor \(N_i\), and \(P\) is a polynomial satisfying certain conditions. Its special values at nonpositive integers are closely related to generalized Bernoulli polynomials. Using this fact we can easily get sums of products of Euler polynomials and generalized Bernoulli polynomials.

1. Introduction.

Let \(\chi\) be a Dirichlet character with conductor \(N\). Generalized Bernoulli numbers and polynomials are defined by Leopoldt [10] by
\[
\sum_{k=1}^{N} \frac{\chi(k)t^k}{e^{Nt} - 1} = \sum_{n=0}^{\infty} \frac{B_n^{\chi} t^n}{n!},\quad |t| < \frac{2\pi}{N},
\]
\[
\sum_{k=1}^{N} \frac{\chi(k)t^{(k+x)t}}{e^{Nt} - 1} = \sum_{n=0}^{\infty} \frac{B_n^{\chi}(x)t^n}{n!},\quad |t| < \frac{2\pi}{N}.
\]
In particular, if \(\chi_0\) is the trivial character, then
\[
(1)\quad B_n^{\chi_0} = (-1)^n B_n, \quad \text{for } n \geq 0,
\]
\[
(2)\quad B_n^{\chi_0}(x) = B_n(1 + x), \quad \text{for } n \geq 0.
\]
If \(\chi\) is the primitive character with conductor 4, then
\[
(3)\quad B_0^{\chi} = 0 \quad \text{and} \quad B_n^{\chi} = -\frac{n}{2} E_{n-1}, \text{ for } n \geq 1;
\]
\[
(4)\quad B_0^{\chi}(x) = 0 \quad \text{and} \quad B_n^{\chi}(x) = -2^{n-2} n E_{n-1} \left(\frac{x+1}{2}\right), \quad \text{for } n \geq 1.
\]
Let $a = (a_1, \ldots, a_r)$, $a_i \geq 0$, $P(X) = P(X_1, \ldots, X_r)$ be a polynomial of $r$ variables with nonnegative real coefficients such that $P(n + a) > 0$ for all $n \in \mathbb{N}^r$ and the series

$$\sum_{n \in \mathbb{N}^r} P(n)^{-s} = \sum_{n_1=1}^{\infty} \cdots \sum_{n_r=1}^{\infty} P(n_1, \ldots, n_r)^{-s}$$

is absolutely convergent for $\text{Re } s > \sigma > 0$. $\chi_1, \ldots, \chi_r$ are nontrivial Dirichlet characters with conductors $N_1, \ldots, N_r$, respectively. Consider the zeta function

$$Z(P, \chi, a, s) = \sum_{n_1=1}^{\infty} \cdots \sum_{n_r=1}^{\infty} \chi(n_1) \cdots \chi(n_r) \cdot P(n_1 + a_1, \ldots, n_r + a_r)^{-s}, \quad \text{Re } s > \sigma.$$

In [2] the author and Eie considered the zeta function $Z(P, \chi, 0, s)$, and we found the special value at nonpositive integers closely related to generalized Bernoulli numbers. Using the same method as in the proof of the Main Theorem in [2], we have the following similar result for $Z(P, \chi, a, s)$ with generalized Bernoulli polynomials:

**Theorem 1.** $Z(P, \chi, a, s)$ defined above has a meromorphic analytic continuation to the whole complex $s$-plane. For any integer $m \geq 0$, if

$$P^m(X) = \sum_{|\alpha|=0}^{mp} C_{\alpha} X_1^{\alpha_1} \cdots X_r^{\alpha_r}, \quad p = \text{deg } P,$$

then

$$Z(P, \chi, a, -m) = (-1)^r \sum_{|\alpha|=0}^{mp} C_{\alpha} \prod_{j=1}^{r} B_{\chi_j}^{\alpha_j+1}(a_j) / \alpha_j + 1.$$

Since Euler polynomials are special cases of generalized Bernoulli polynomials with the primitive Dirichlet character $\chi$ of conductor 4, we can easily get the following theorem:

**Theorem 2.** Let $P$ and $a$ be defined as in Theorem 1. The zeta function

$$Z(P, s) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_r=0}^{\infty} (-1)^{n_1 + \cdots + n_r} P(n_1 + a_1, \ldots, n_r + a_r)^{-s}$$

has a meromorphic analytic continuation to the whole complex $s$-plane. For any integer $m \geq 0$,\n
$$Z(P, -m) = \sum_{|\alpha|=0}^{mp} C_{\alpha} \prod_{j=1}^{r} E_{\alpha_j}(a_j) / 2.$$
We use the notation \( s(n, k) \) for the Stirling number of the first kind, the number of ways to permute a list of \( n \) items into \( k \) cycles (cf. \([6]\)). Using some relations between different zeta functions and their special values at nonpositive integers, we can get sums of products of Euler polynomials, Bernoulli polynomials, and generalized Bernoulli polynomials.

**Theorem 3.** Let \( y = x_1 + \cdots + x_N \). Then

\[
(7) \quad \sum_{j_1 + \cdots + j_N = m} \binom{m}{j_1, \ldots, j_N} E_{j_1}(x_1) \cdots E_{j_N}(x_N)
\]

\[
= \frac{2^{N-1}}{(N-1)!} \sum_{k=0}^{N-1} s(N, k+1) \sum_{j=0}^{k} \binom{k}{j} (-y)^{k-j} m^j + j(y).
\]

**Theorem 4.** Let \( y = x_1 + \cdots + x_N \). Then

\[
(8) \quad \sum_{j_1 + \cdots + j_N = m} \binom{m}{j_1, \ldots, j_N} B_{j_1}(x_1) \cdots B_{j_N}(x_N)
\]

\[
= \frac{(-1)^{N-1}m!}{(N-1)!}(m-N)! \sum_{k=0}^{N-1} s(N, k+1) \sum_{j=0}^{k} \binom{k}{j} (-y)^{k-j} \frac{B_{m+j-N+1}(y)}{m+j-N+1}.
\]

**Theorem 5.** Let \( r \) be a positive integer and \( \chi_i \) be a nontrivial Dirichlet character with conductor \( N_i \), for \( i = 1, 2, \ldots, r \). Then for any positive integer \( m \),

\[
(9) \quad \sum_{j_1 + \cdots + j_r = m} \binom{m}{j_1, \ldots, j_r} \frac{B_{\chi_1}^{j_1+1}(x_1)}{N_1^{j_1+1}(j_1+1)} \cdots \frac{B_{\chi_r}^{j_r+1}(x_r)}{N_r^{j_r+1}(j_r+1)}
\]

\[
= \frac{-1}{(r-1)!} \sum_{a_1=1}^{N_1} \cdots \sum_{a_r=1}^{N_r} \chi_1(a_1) \cdots \chi_r(a_r)
\]

\[
\cdot \sum_{k=0}^{r-1} s(r, k+1) \sum_{j=0}^{k} \binom{k}{j} (-y)^{k-j} \frac{B_{m+j+1}(y)}{m+j+1},
\]

where \( y = \frac{a_1+x_1}{N_1} + \cdots + \frac{a_r+x_r}{N_r} \).

In the last section, we reproduce some classical identities among Euler polynomials using our method, and also some new identities.

### 2. Sketch of proof of Theorem 1.

Since the proof is exactly the same as \([2]\), we just sketch the outline.
Finding the special value at $s = -m$ of the zeta function

$$Z(P, \chi, a, s) = \sum_{n_1=1}^{\infty} \cdots \sum_{n_r=1}^{\infty} \chi_1(n_1) \cdots \chi_r(n_r) P(n + a)^{-s},$$

is equivalent to finding the coefficient of $t^m$ in the asymptotic expansion at $t = 0$ of the function

$$\sum_{n_1=1}^{\infty} \cdots \sum_{n_r=1}^{\infty} \chi_1(n_1) \cdots \chi_r(n_r) \exp\{-P(n + a)t\}.$$

It is also equivalent to finding the constant term in the asymptotic expansion at $t = 0$ of the function

$$g(t) = \sum_{n_1=1}^{\infty} \cdots \sum_{n_r=1}^{\infty} \chi_1(n_1) \cdots \chi_r(n_r) P^m(n + a) \exp\{-P(n + a)t\}.$$

For the given polynomial

$$P(X) = \sum_{|\alpha| = 0}^{p} A_\alpha X^\alpha, \quad p = \deg P,$$

we let

$$Q(X, Y) = \sum_{|\alpha| = 0}^{p} A_\alpha X^\alpha Y^{p-|\alpha|}$$

be the corresponding homogeneous polynomial in $r+1$ variables. Obviously, $Q((n + a)t, t) = P(n + a)t^p$ and so

$$g(t^p) = \sum_{n \in \mathbb{N}^r} \chi_1(n_1) \cdots \chi_r(n_r) P^m(n + a) \exp\{-P(n + a)t^p\}$$

$$= \sum_{n \in \mathbb{N}^r} \chi_1(n_1) \cdots \chi_r(n_r) P^m(n + a) \exp\{-Q((n + a)t, t)\}$$

$$= \sum_{|\alpha| = 0}^{mp} C_\alpha \sum_{n \in \mathbb{N}^r} \chi_1(n_1) \cdots \chi_r(n_r)(n + a)^\alpha \exp\{-Q((n + a)t, t)\}$$

where

$$P^m(X) = \sum_{|\alpha| = 0}^{mp} C_\alpha X^\alpha \quad \text{and} \quad n^\alpha = n_1^{\alpha_1} \cdots n_r^{\alpha_r}.$$
has the form $\sum_{n=0}^{\infty} d_n t^n$ with the constant term $d_0$ given by

$$d_0 = (-1)^r \prod_{j=1}^{r} \frac{B_{\chi_j}^{a_j+1}(a_j)}{\beta_j + 1}.$$

Therefore, we get our assertion for generalized Bernoulli polynomials.

### 3. Proof of Theorem 2.

Let $\chi$ be the primitive Dirichlet character with conductor 4. Then the zeta function can be rewritten as

$$Z(P,s) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_r=0}^{\infty} (-1)^{n_1 + \cdots + n_r} P(n_1 + a_1, \ldots, n_r + a_r)^{-s}$$

$$= \sum_{n_1=0}^{\infty} \cdots \sum_{n_r=0}^{\infty} \chi(2n_1 + 1) \cdots \chi(2n_r + 1) P(n_1 + a_1, \ldots, n_r + a_r)^{-s}$$

$$= \sum_{k_1=1}^{\infty} \cdots \sum_{k_r=1}^{\infty} \chi(k_1) \cdots \chi(k_r) P\left(\frac{k_1-1}{2} + a_1, \ldots, \frac{k_r-1}{2} + a_r\right)^{-s}.$$

Now we assume that $P^m(X) = \sum_{|\alpha|=0}^{mp} C_{\alpha} X^\alpha$, where $p = \deg P$. Thus

$$P^m\left(\frac{k-1}{2} + a\right) = \sum_{|\alpha|=0}^{mp} C_{\alpha} \left(\frac{k_1-1}{2} + a_1\right)^{\alpha_1} \cdots \left(\frac{k_r-1}{2} + a_r\right)^{\alpha_r}$$

$$= \sum_{|\alpha|=0}^{mp} \frac{C_{\alpha}}{2^{2|a|}} (k_1 + 2a_1 - 1)^{\alpha_1} \cdots (k_r + 2a_r - 1)^{\alpha_r}.$$

Now we apply Theorem 1 and Equation (4) to this zeta function

$$Z(P, -m) = (-1)^r \sum_{|\alpha|=0}^{mp} \frac{C_{\alpha}}{2^{2|a|}} \prod_{j=1}^{r} \frac{B_{\chi_j}^{a_j+1}(2a_j - 1)}{\alpha_j + 1}$$

$$= (-1)^r \sum_{|\alpha|=0}^{mp} \frac{C_{\alpha}}{2^{2|a|}} \prod_{j=1}^{r} \frac{-2^{a_j-1}(\alpha_j + 1) E_{\alpha_j}(a_j)}{\alpha_j + 1}$$

$$= \sum_{|\alpha|=0}^{mp} \frac{C_{\alpha}}{2} \prod_{j=1}^{r} \frac{E_{\alpha_j}(a_j)}{2}.$$

This completes our proof.
4. Sums of products of Euler polynomials.

We use a result stated in [5]. For $m_i$ positive integers and $\deg P < m_1 + \cdots + m_r$, we consider the rational function

$$F(T) = \frac{P(T)}{(1 - T^{m_1}) \cdots (1 - T^{m_r})} = \sum_{k=0}^{\infty} a(k)T^k,$$

where $|T| < 1$, and

$$a(k) = \frac{1}{2\pi i} \int_C \frac{F(z)}{z^{k+1}} dz$$

is determined by $F$ via Cauchy’s integral formula, with $C$ a sufficiently small circle centered at the origin and going counterclockwise. The zeta function (cf. Chapter XVII of [8])

$$Z_F(s) = \sum_{k=1}^{\infty} a(k)k^{-s},$$

is related to $F(T)$ via a Mellin transform

$$Z_F(s)\Gamma(s) = \int_0^\infty t^{s-1}[F(e^{-t}) - F(0)] dt,$$

for $\Re s$ sufficiently large. The main tool that we use to prove the following theorems and propositions is as follows:

**Lemma (Lemma 3 of [5]).** Given

$$P(T) = \sum_{j=0}^{m} b_j T^j \quad \text{and} \quad F(T) = \frac{P(T)}{(1 - T^{m_1}) \cdots (1 - T^{m_r})}$$

with $m_1 + \cdots + m_r > m$, then, for $|T| < 1$ we have

$$F(T) = \sum_{j=0}^{m} b_j \sum_{n_1=0}^{\infty} \cdots \sum_{n_r=0}^{\infty} T^{n_1 m_1 + \cdots + n_r m_r + j}$$

and hence the associated zeta function

$$Z_F(s) = b_0 \sum_{n_1, \ldots, n_r \geq 0, |n| > 0} (n_1 m_1 + \cdots + n_r m_r)^{-s}$$

$$+ \sum_{j=1}^{m} b_j \sum_{n_1, \ldots, n_r \geq 0} (n_1 m_1 + \cdots + n_r m_r + j)^{-s}.$$
Using the above statements we can prove Theorem 3. Consider the rational function
\[
F(T) = \frac{T^{x_1}}{1 + T} \cdots \frac{T^{x_N}}{1 + T} = \left[ \frac{T^{x_1}}{1 + T} \sum_{n_1 = 0}^{\infty} (-T)^{n_1} \right] \cdots \left[ \frac{T^{x_N}}{1 + T} \sum_{n_N = 0}^{\infty} (-T)^{n_N} \right] = \sum_{n_1 = 0}^{\infty} \cdots \sum_{n_N = 0}^{\infty} (-1)^{n_1 + \cdots + n_N} T^{(n_1 + x_1) + \cdots + (n_N + x_N)}.
\]
Its associated zeta function is
\[
Z_F(s) = \sum_{n_1 = 0}^{\infty} \cdots \sum_{n_N = 0}^{\infty} (-1)^{n_1 + \cdots + n_N} [(n_1 + x_1) + \cdots + (n_N + x_N)]^{-s}.
\]
Using the result of Theorem 2 we know that for \( m \geq 0 \),
\[
(10) \quad Z_F(-m) = 2^{-N} \sum_{j_1 + \cdots + j_N = m} \binom{m}{j_1, \ldots, j_N} E_{j_1}(x_1) \cdots E_{j_N}(x_N).
\]
On the other hand, let \( y = x_1 + \cdots + x_N \); we can rewrite the rational function \( F(T) \) as
\[
F(T) = \frac{T^y}{(1 + T)^N} = \sum_{n = 0}^{\infty} (-1)^n \frac{(n + N - 1)(n + N - 2) \cdots (n + 1)}{(N - 1)!} T^{n + y}.
\]
The associated zeta function can also be rewritten as
\[
Z_F(s) = \sum_{n = 0}^{\infty} \frac{(-1)^n}{(N - 1)!} (n + N - 1)(n + N - 2) \cdots (n + 1)(n + y)^{-s} = \sum_{n = 0}^{\infty} \frac{(-1)^n}{(N - 1)!} \sum_{k = 0}^{N - 1} s(N, k + 1)n^k \cdot (n + y)^{-s},
\]
since (cf. Eq. (7.48) of [6])
\[
(n + 1)(n + 2) \cdots (n + N - 1) = \sum_{k = 0}^{N - 1} s(N, k + 1)n^k.
\]
Thus
\[
Z_F(s) = \sum_{n = 0}^{\infty} \frac{(-1)^n}{(N - 1)!} \sum_{k = 0}^{N - 1} s(N, k + 1)(n + y - y)^k \cdot (n + y)^{-s} = \sum_{n = 0}^{\infty} \frac{(-1)^n}{(N - 1)!} \sum_{k = 0}^{N - 1} s(N, k + 1) \sum_{j = 0}^{k} \binom{k}{j} (-y)^{k-j}(n + y)^{j-s}.
\]
Again using the result of Theorem 2 (this time for \(r = 1\) and \(P(x) = x\)) we know that for \(m \geq 0\)

\[
Z_F(-m) = \sum_{k=0}^{N-1} s(N, k + 1) \frac{k^m}{(N - 1)!} \sum_{j=0}^{k} \frac{(-y)^{k-j}E_{m+j}(y)}{2}.
\]

Now combine Equation (10) and Equation (11) to obtain our assertion.

5. Sums of products of generalized Bernoulli polynomials.

We first prove Theorem 4, then apply it to prove Theorem 5. The proof of Theorem 4 is similar to the proof of Theorem 3. We just consider the different rational function

\[
F(T) = \frac{T^{x_1}}{1-T} \cdots \frac{T^{x_N}}{1-T} = \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} T^{(n_1+x_1)+\cdots+(n_N+x_N)}.
\]

Its associated zeta function is

\[
Z_F(s) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} [(n_1 + x_1) + \cdots + (n_N + x_N)]^{-s}.
\]

Using the result of Proposition 2 in [5], we know that for \(m \geq 0\)

\[
Z_F(-m) = (-1)^N \sum_{j_1+\cdots+j_N=m+N \atop j_i \geq 0} \binom{m}{j_1, \ldots, j_N} B_{j_1}(x_1) \cdots B_{j_N}(x_N).
\]

On the other hand, let \(y = x_1 + \cdots + x_N\); we can rewrite the rational function \(F(T)\) as

\[
F(T) = \frac{T^y}{(1-T)^N} = \sum_{n=0}^{\infty} \frac{T^{n+y}}{(N-1)!} \sum_{k=0}^{N-1} s(N, k + 1)n^k.
\]

The associated zeta function can also rewrite as

\[
Z_F(s) = \sum_{n=0}^{\infty} \frac{1}{(N-1)!} \sum_{k=0}^{N-1} s(N, k + 1) \sum_{j=0}^{k} \binom{k}{j} (-y)^{k-j}(n+y)^{j-s}.
\]

Again using the same result of Proposition 2 in [5], we have for \(m \geq 0\)

\[
Z_F(-m) = \sum_{k=0}^{N-1} s(N, k + 1) \frac{k^m}{(N - 1)!} \sum_{j=0}^{k} \frac{(-y)^{k-j}E_{m+j}(y)}{2}.
\]

Now combine Equation (12), Equation (13), and change \(m + N\) to \(m\), to conclude the proof of Theorem 4.
To prove Theorem 5, we consider the zeta function

\[ Z(s) = \sum_{n_1=1}^{\infty} \cdots \sum_{n_r=1}^{\infty} \chi_1(n_1) \cdots \chi_r(n_r) \left( \sum_{j=1}^{r} \prod_{i=1 \atop i \neq j}^{r} N_i \right) (n_j + x_j)^{-s} \]

Substitute \( n_i = a_i + N_i m_i \) where \( a_i = 1, \ldots, N_i \) and \( m_i \geq 0 \) for \( i = 1, \ldots, r \). Thus \( Z(s) \) becomes

\[ \sum_{a_1=1}^{N_1} \cdots \sum_{a_r=1}^{N_r} \sum_{m_1=0}^{\infty} \sum_{m_r=0}^{\infty} \left( \prod_{i=1}^{r} \chi_i(a_i + m_i N_i N_i^{-s}) \left( \sum_{j=1}^{r} m_j + \frac{a_j + x_j}{N_j} \right) \right)^{-s} \]

Now we let

\[ Z_B(s) = \sum_{m_1=0}^{\infty} \cdots \sum_{m_r=0}^{\infty} \left( \prod_{i=1}^{r} N_i \right)^{-s} \left[ \sum_{j=1}^{r} m_j + \frac{a_j + x_j}{N_j} \right]^{-s} \]

Then we can represent the zeta function \( Z(s) \) as

\[ Z(s) = \sum_{a_1=1}^{N_1} \cdots \sum_{a_r=1}^{N_r} \left( \prod_{i=1}^{r} \chi_i(a_i) \right) Z_B(s). \]

From [4] we know that this zeta function \( Z_B(s) \) has an analytic continuation to the whole complex plane, and the special values at nonpositive integers \( s = -m \) are given by

\[ Z_B(-m) = \left( \prod_{i=1}^{r} N_i^{m_i} \right) \sum_{p_1^{m_1} \cdots p_r^{m_r} \geq 0} \frac{m!}{p_1! \cdots p_r!} \prod_{j=1}^{r} B_{p_j} \left( -\frac{a_j + x_j}{N_j} \right). \]

Using the result of Theorem 4 we can rewrite \( Z_B(-m) \) as

\[ \frac{(\prod_{i=1}^{r} N_i^{m_i})}{(r-1)!} \sum_{k=0}^{r-1} s(r, k + 1) \sum_{j=0}^{k} \binom{k}{j} (-y)^{k-j} B_{m+j+1}(y) \frac{B_{m+j+1}(x_j)}{m+j+1}, \]

where \( y = \frac{a_1 + x_1}{N_1} + \cdots + \frac{a_r + x_r}{N_r} \). Now applying Theorem 1, the special values at nonpositive integers \( s = -m \) of the zeta function \( Z(s) \) are

\[ Z(-m) = \sum_{p_1^{m_1} \cdots p_r^{m_r} \geq 0} \binom{m}{p_1, \ldots, p_r} (-1)^r \left( \prod_{i=1}^{r} N_i^{m_i-p_i} B_{p_i+1}(x_i) \right). \]

On the other hand, using the equality

\[ Z(-m) = \sum_{a_1=1}^{N_1} \cdots \sum_{a_r=1}^{N_r} \prod_{i=1}^{r} \chi_i(a_i) Z_B(-m) \]

and the above values of \( Z(-m) \) and \( Z_B(-m) \), we get our assertion.
Remark.
(1) Dilcher in [3] produced Equations (7) and (8) in a different way. These formulae are the same, except for the definition of the Stirling numbers of the first kind.
(2) The author and Eie in [2] produced a formula with sums of products of generalized Bernoulli numbers; here we have used the same ideas to prove a similar formula with generalized Bernoulli polynomials.
(3) Huang and Huang in [9] gave some generalized formulas for sums of products of Bernoulli numbers and polynomials via a different method called algebraic residues.

6. Some further identities.
Applying the method of proof of Theorems 3 and 4 to different rational functions, we can get different identities between generalized Bernoulli polynomials, Euler polynomials, and Bernoulli polynomials. Here we list some classical identities among Euler polynomials (cf. [1]).

**Proposition 1** (see 23.1.7 of [1]).

\[ E_m(x+h) = \sum_{k=0}^{m} \binom{m}{k} E_k(x)h^{m-k} = \sum_{k=0}^{m} \binom{m}{k} E_k(h)x^{m-k}, \]

for any nonnegative integer \( m \).

**Proof.** Consider the zeta function

\[ Z(P,s) = \sum_{n=0}^{\infty} (-1)^n (n + x + h)^{-s}, \quad \text{where} \quad P^m(z) = (z + x + h)^m. \]

We can express \( P^m(z) \) in a different way, as

\[ P^m(z) = \sum_{k=0}^{m} \binom{m}{k} (z + x)^k h^{m-k}, \quad \text{or} \quad \sum_{k=0}^{m} \binom{m}{k} (z + h)^k x^{m-k}. \]

Then we apply Theorem 2 to this zeta function \( Z(P,s) \) to obtain the assertion. \( \square \)

**Proposition 2** (see 23.1.10 of [1]).

\[ E_m(kx) = \begin{cases} k^m \sum_{i=0}^{k-1} (-1)^i E_m(x + \frac{i}{k}), & \text{if } k \text{ is odd,} \\ -2^{m+1}k^m \sum_{i=0}^{k-1} (-1)^i B_{m+1}(x + \frac{i}{k}), & \text{if } k \text{ is even,} \end{cases} \]

for any nonnegative integer \( m \).

**Proof.** We consider the zeta function

\[ Z(P,s) = \sum_{n=0}^{\infty} (-1)^n (n + kx)^{-s} = \sum_{n=0}^{\infty} \sum_{i=0}^{k-1} (-1)^{nk+i} (nk + i + kx)^{-s}. \]
Now we separate $k$ into two cases, odd and even.

\[
Z(P, s) = \begin{cases} 
\sum_{i=0}^{k-1} (-1)^i k^{-s} \sum_{n=0}^{\infty} (-1)^n (n + x + \frac{i}{k})^{-s}, & \text{if } k \text{ is odd}, \\
\sum_{i=0}^{k-1} (-1)^i k^{-s} \sum_{n=0}^{\infty} (n + x + \frac{i}{k})^{-s}, & \text{if } k \text{ is even}.
\end{cases}
\]

Again we apply Theorem 2 to $Z(P, s)$ and complete the proof. \(\square\)

**Proposition 3** (see Eq. (51.6.5) of [7]). For any nonnegative integer $m$,

\[
\sum_{k=0}^{m} \binom{m}{k} 2^k E_{m-k}(x) E_k(y) = E_m(x + 2y) + 2^m E_m\left(\frac{x + 2y}{2}\right) - 2^m E_m\left(\frac{x + 2y + 1}{2}\right).
\]

**Proof.** Follow a similar argument as in the proof of the previous proposition, but consider the different fraction

\[
F(T) = 4 \cdot \frac{T^{2x}}{1 + T^2} \cdot \frac{T^{4y}}{1 + T^4} = 2 \cdot \frac{T^{2x+4y}}{1 + T^2} + 2 \cdot \frac{T^{2x+4y}}{1 + T^4} - 2 \cdot \frac{T^{2x+4y+2}}{1 + T^4}.
\]

From the associated zeta functions, we get the identity

\[
4 \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} (-1)^{n_1+n_2} [(2n_1 + 2x) + (4n_2 + 4y)]^{-s}
\]

\[
= 2 \sum_{n=0}^{\infty} (-1)^n (2n + 2x + 4y)^{-s} + 2 \sum_{n=0}^{\infty} (-1)^n (4n + 2x + 4y)^{-s} - 2 \sum_{n=0}^{\infty} (-1)^n (4n + 2x + 4y + 2)^{-s}.
\]

Then we calculate the special value of $s = -m$ with $m \geq 0$ and obtain Proposition 3. \(\square\)

The following results give new identities for Euler polynomials:

**Proposition 4.** Let $a$ be any positive odd integer. Then for any nonnegative integer $m$, we have:
(1) If $a = 4k + 1$, then

$$
\sum_{l=0}^{m} \binom{m}{l} 2^l E_l(x) a^{m-l} E_{m-l}(y)
= \frac{1}{a} E_m(2x + ay) + 2^m E_m \left( \frac{2x + ay}{2} \right) - 2^m E_m \left( \frac{2x + ay + 1}{2} \right) 
+ \sum_{n=1}^{k} \left[ a^{m-1}(a-1) E_m \left( \frac{2x + ay + 4n}{a} \right) 
- a^{m-1}(a-1) E_m \left( \frac{2x + ay + 4n - 1}{a} \right) 
- a^{m-1}(a+1) E_m \left( \frac{2x + ay + 4n - 2}{a} \right) 
+ a^{m-1}(a+1) E_m \left( \frac{2x + ay + 4n - 3}{a} \right) \right] 
+ a^{m-1}(a-1) E_m \left( \frac{2x + ay}{a} \right) .
$$

(2) If $a = 4k + 3$, then

$$
\sum_{l=0}^{m} \binom{m}{l} 2^l E_l(x) a^{m-l} E_{m-l}(y)
= \frac{1}{a} E_m(2x + ay) + 2^m E_m \left( \frac{2x + ay}{2} \right) + 2^m E_m \left( \frac{2x + ay + 1}{2} \right) 
- \left\{ \sum_{n=1}^{k} \left[ a^{m-1}(a+1) E_m \left( \frac{2x + ay + 4n + 2}{a} \right) 
+ a^{m-1}(a-1) E_m \left( \frac{2x + ay + 4n + 1}{a} \right) 
- a^{m-1}(a-1) E_m \left( \frac{2x + ay + 4n}{a} \right) 
- a^{m-1}(a+1) E_m \left( \frac{2x + ay + 4n - 1}{a} \right) \right] 
+ a^{m-1}(a+1) E_m \left( \frac{2x + ay + 2}{a} \right) + a^{m-1}(a-1) E_m \left( \frac{2x + ay + 1}{a} \right) 
- a^{m-1}(a-1) E_m \left( \frac{2x + ay}{a} \right) \right\} .
$$
Proof. The proof is similar to that of Proposition 3, but with

\[
F(T) = 4 \cdot \frac{T^{2x}}{1 + T^2} \cdot \frac{T^{ay}}{1 + T^a} \\
= \frac{2}{a} \frac{T^{2x+ay}}{1 + T} + 2 \frac{T^{2x+ay}}{1 + T^2} - 2 \frac{T^{2x+ay+1}}{1 + T^2} \\
+ \frac{2}{1 + T^a} \left\{ \sum_{n=1}^{k} \left[ \frac{a - 1}{a} T^{2x+ay+4n} - \frac{a - 1}{a} T^{2x+ay+4n-1} \\
- \frac{a + 1}{a} T^{2x+ay+4n+2} + \frac{a + 1}{a} T^{2x+ay+4n+3} \right] + \frac{a - 1}{a} T^{2x+ay} \right\},
\]

for \( a = 4k + 1 \) and

\[
F(T) = 4 \cdot \frac{T^{2x}}{1 + T^2} \cdot \frac{T^{ay}}{1 + T^a} \\
= \frac{2}{a} \frac{T^{2x+ay}}{1 + T} + 2 \frac{T^{2x+ay}}{1 + T^2} + 2 \frac{T^{2x+ay+1}}{1 + T^2} \\
- \frac{2}{1 + T^a} \left\{ \sum_{n=1}^{k} \left[ \frac{a + 1}{a} T^{2x+ay+4n+2} \\
+ \frac{a - 1}{a} T^{2x+ay+4n+1} - \frac{a - 1}{a} T^{2x+ay+4n} - \frac{a + 1}{a} T^{2x+ay+4n-1} \right] \\
+ \frac{a + 1}{a} T^{2x+ay+2} + \frac{a - 1}{a} T^{2x+ay+1} - \frac{a - 1}{a} T^{2x+ay} \right\},
\]

for \( a = 4k + 3 \), respectively. \( \Box \)

Remark. We can generalize the previous propositions to formulas involving \( a^k \) and \( b^{m-k} \) for arbitrary integers \( a \) and \( b \), depending on a suitable partial fraction decomposition of the function \( F(T) \).

Acknowledgements. The author would like to thank the referee for some useful comments and suggestions. And this paper was supported by the National Science Council of R. O. C. grant under NSC 89-2115-M-122-006 and the Yuanpei University of Science and Technology.

References


Received May 9, 2001 and revised November 13, 2001.
THE MODULI SPACE OF REAL ALGEBRAIC CURVES OF GENUS 2

FRANCISCO-JAVIER CIRRE

We give an explicit description of each of the five connected components of the moduli space of real algebraic curves of genus 2. This is done in terms of the coefficients of equations defining such curves. We also find the subsets consisting of curves with prescribed automorphism group.

Introduction.

Riemann’s moduli problem could be formulated as that of describing the set of isomorphism classes of complex algebraic curves of a given topological type. Its real counterpart, i.e., the description of the set of real isomorphism classes of real algebraic curves, was initiated by Klein [8] and [9], who mainly worked with polynomial equations that define the real curve in question. The problem has acquired importance in the last three decades due to both the development of proper techniques of real algebraic geometry and the right definition of the “analytic” surface associated to a real algebraic curve. Such definition is given in [1] and the associated surface is known as Klein surface. This establishes the well-known coequivalence between real algebraic curves (algebraic objects) and Klein surfaces (geometric ones). A large number of works have appeared since; a good reference for many of them is [11].

Usual methods to deal with moduli spaces are based on Teichmüller ones. Different points of view can be considered, such as quasiconformal mappings, see for example [12, 14, 7], abelian varieties [13, 15] or non-euclidean crystallographic groups [10]. They rely upon the geometry of the surface associated to a curve. The approach here is more classical in the sense that we directly work with polynomial equations of curves.

By considering antianalytic involutions on Riemann surfaces of genus 0 and 1, Alling and Greenleaf classified all isomorphism classes of rational and elliptic real curves ([1], Thms. 1.9.4 and 1.9.8). So it is interesting to understand the situation in higher genera. In this paper we study the case of hyperelliptic real curves, which in particular covers the genus 2 case. Using coefficients of polynomial equations we describe the connected components of the space $\mathcal{M}_2^\mathbb{R}$ of isomorphism classes of real algebraic curves of genus 2.
We also find their subsets consisting of curves with prescribed automorphism group.

The paper is organized as follows. In Section 1 we recall some well-known facts about real curves. The reader may find them, e.g., in Chapters 3 and 5 in [14]. The second section deals with hyperelliptic curves. In it we develop a simple method to approach the problem of classifying hyperelliptic real algebraic curves of arbitrary genus up to birational isomorphism. The method is based on a detailed analysis of the action of real Möbius transformations on the sphere. It is worth pointing out how much information can be obtained from this simple method. It is applied in Sections 3 to 6 to the genus 2 case; namely, for each topological type \((2, k, \varepsilon)\) we find a bijection \(\Psi_{(2, k, \varepsilon)}\) between a basic semialgebraic subset \(\Delta_{(2, k, \varepsilon)}\) of \(\mathbb{R}^3\) and the real moduli space \(\mathcal{M}_{(2, k, \varepsilon)}\) of real algebraic curves of topological type \((2, k, \varepsilon)\). The bijection \(\Psi_{(2, k, \varepsilon)}\) is given in terms of the branch points of the curve. The moduli space \(\mathcal{M}_{(2, k, \varepsilon)}\) admits a semianalytic structure (see [13, 6]) which is natural in the sense that the positions of the branch points could be chosen as analytic parameters for the curves. (In the complex case, the branch-point positions were the first (heuristic) method used by Riemann to predict the dimension \(3g - 3\) of the moduli space of complex algebraic curves of genus \(g\).) We exhibit this naturality in Section 7 by showing that each map \(\Psi_{(2, k, \varepsilon)}\) is indeed real analytic.

1. Preliminaries.

A real algebraic curve of genus \(g\) is a pair \((X, \sigma)\) where \(X\) is a projective, smooth, irreducible complex algebraic curve of genus \(g\) and \(\sigma : X \to X\) is an antianalytic involution on \(X\). Two real curves \((X, \sigma)\) and \((Y, \tau)\) are isomorphic if there exists an isomorphism \(f : X \to Y\) such that \(f \circ \sigma = \tau \circ f\). We denote by \(\mathcal{M}_g^\mathbb{R}\) the moduli space of real algebraic curves of genus \(g\):

\[
\mathcal{M}_g^\mathbb{R} = \{\text{isomorphism classes of real algebraic curves of genus } g\}.
\]

Two isomorphic real curves are homeomorphic but the converse is not true. \(((X, \sigma)\) and \((Y, \tau)\) are homeomorphic if \(f \circ \sigma = \tau \circ f\) for some homeomorphism \(f : X \to Y\).) Weichold [16] showed that the homeomorphism class of \((X, \sigma)\) is determined by its topological type \((g, k, \varepsilon)\), where \(g\) is the genus of \(X\), \(k\) is the number of connected components of the real part \(X_\sigma\) of \((X, \sigma)\), which is the fixed point set of \(\sigma\), and \(\varepsilon = 1\) if \(X - X_\sigma\) is connected and 0 otherwise. For each value of \(g\), there are exactly \(\lfloor (3g + 4)/2 \rfloor\) topologically different real algebraic curves of genus \(g\), where \(\lfloor r \rfloor\) stands for the integer part of \(r\). For each admissible triple \((g, k, \varepsilon)\) we write

\[
\mathcal{M}_{(g, k, \varepsilon)} = \{(X, \sigma) \in \mathcal{M}_g^\mathbb{R} : (X, \sigma) \text{ has topological type } (g, k, \varepsilon)\},
\]
where we have used \((X, \sigma)\) to denote both a real algebraic curve and its isomorphism class. This convention will be assumed throughout this paper.

For abbreviation, a real algebraic curve of topological type \((g, k, \varepsilon)\) will be called a \((g, k, \varepsilon)\)-curve. The usual topology on \(\mathcal{M}_g^R\) makes the sets \(\mathcal{M}_{(g,k,\varepsilon)}\) be the connected components of \(\mathcal{M}_g^R\). For example, \(\mathcal{M}_2^R\) has the following 5 connected components:

\[
\mathcal{M}_2^R = \mathcal{M}_{(2,3,0)} \cup \mathcal{M}_{(2,2,1)} \cup \mathcal{M}_{(2,1,1)} \cup \mathcal{M}_{(2,1,0)} \cup \mathcal{M}_{(2,0,1)}.
\]

In terms of Klein surfaces, \((2,3,0)\)-curves correspond to spheres with 3 holes, \((2,2,1)\)-curves to projective planes with 2 holes, \((2,1,1)\)-curves to connected sums of two projective planes with a hole, \((2,1,0)\)-curves to tori with a hole and \((2,0,1)\) to connected sums of three projective planes.

A natural partition of the moduli space \(\mathcal{M}_{(g,k,\varepsilon)}\) arises when considering its subsets \(\mathcal{M}_{(g,k,\varepsilon)}(H)\) formed by curves with the same automorphism group \(H\):

\[
\mathcal{M}_{(g,k,\varepsilon)}(H) = \{(X, \sigma) \in \mathcal{M}_{(g,k,\varepsilon)} : \text{Aut}(X, \sigma) = H\},
\]

where \(\text{Aut}(X, \sigma) = \{f : X \to X : f \text{ analytic selfhomeomorphism such that } f \circ \sigma = \sigma \circ f\}\) is the automorphism group of \((X, \sigma)\). By means of combinatorial methods, the list of groups which are realized as the automorphism groups of real algebraic curves of genus 2 was calculated in [2] for \((2,0,1)\)-curves and in [3] for the rest of cases. As a by-product we obtain here the same results although in a very different way. We will denote the cyclic group of order \(N\) by \(C_N\) and the dihedral group of order \(2N\) by \(D_N\). Thus \(D_2\) stands for the non-cyclic group of order 4.

2. Hyperelliptic real algebraic curves.

A real algebraic curve \((X, \sigma)\) of genus \(g \geq 2\) is hyperelliptic if so is the complex curve \(X\), that is, if \(X\) admits a meromorphic function of degree 2. Throughout this paper we represent \(X\) by its affine plane model

\[
X = \{(x, y) \in \mathbb{C}^2 : y^2 = P_X(x) := (x - e_1) \cdots (x - e_{2g+1+\delta})\}
\]

with \(e_i \neq e_j\) and \(\delta = 0\) or 1. In this model we identify the characteristic elements of a hyperelliptic curve. First, a meromorphic function of degree 2 is given by the projection \(\pi_X : (x, y) \mapsto x\) onto the Riemann sphere \(\hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}\). Its branch points are thus the roots of \(P_X\) and possibly \(\infty\). They constitute what we call (by abuse of language) the branch point set of \(X\),

\[
B_X = \begin{cases} 
\{e_1, \ldots, e_{2g+2}\} & \text{if } \delta = 1, \\
\{e_1, \ldots, e_{2g+1}, \infty\} & \text{if } \delta = 0.
\end{cases}
\]

The automorphism interchanging the two sheets of \(\pi_X\) is known as the hyperelliptic involution \(h_X : (x, y) \mapsto (x, -y)\). It is central in \(\text{Aut}(X, \sigma)\).
Let \( Y = \{ w^2 = P_Y(z) \} \) be another hyperelliptic curve and \( B_Y \) its branch point set. Every birational isomorphism \( f : X \to Y \) induces a Möbius transformation \( \hat{f} : \hat{\mathbb{C}} \to \hat{\mathbb{C}} \) which maps \( B_X \) onto \( B_Y \). In fact, \( \hat{f} \) is defined by \( \hat{f} : \pi_X(p) \mapsto \pi_Y(f(p)) \) for any \( p \in X \).

\[
\begin{array}{ccc}
X & \xrightarrow{f} & Y \\
\pi_X & \downarrow & \pi_Y \\
\hat{\mathbb{C}} & \xrightarrow{\hat{f}} & \hat{\mathbb{C}}
\end{array}
\]

Conversely, every Möbius transformation \( m : \hat{\mathbb{C}} \to \hat{\mathbb{C}} \) which maps \( B_X \) onto \( B_Y \) induces exactly two birational isomorphisms \( f_1, f_2 : X \to Y \) such that \( \hat{f}_i = m, i = 1, 2 \) (in fact, \( f_2 = f_1 \circ h_X = h_Y \circ f_1 \)). We call these isomorphisms liftings of \( m \). Their formulae can be calculated explicitly, [4].

**Liftings of Möbius transformations.** Writing \( m(x) = \frac{ax + b}{cx + d} \) with \( \det m := ad - bc \neq 0 \), we have:

- If \( \infty \in B_X \) and \( m(\infty) = \infty \) then
  \[
  f_1(x, y) = \left( \frac{ax + b}{d}, y \cdot \left( \frac{a}{d} \right)^g \sqrt{\det m} \right).
  \]

- If \( \infty \in B_X \) and \( m(\infty) \neq \infty \) then
  \[
  f_1(x, y) = \left( \frac{ax + b}{cx + d}, \frac{y \cdot c^g}{(cx + d)^{g+1}} \sqrt{-\det m \cdot P'_Y(m(\infty))} \right),
  \]
  where \( P'_Y \) denotes the derivative of \( P_Y \).

- If \( \infty \notin B_X \) and \( m(\infty) = \infty \) then
  \[
  f_1(x, y) = \left( \frac{ax + b}{d}, y \cdot \left( \frac{a}{d} \right)^{g+1} \right).
  \]

- If \( \infty \notin B_X \) and \( m(\infty) \neq \infty \) then
  \[
  f_1(x, y) = \left( \frac{ax + b}{cx + d}, \frac{y \cdot c^{g+1}}{(cx + d)^{g+1}} \sqrt{P_Y(m(\infty))} \right).
  \]

We now return to real algebraic curves. If the hyperelliptic complex curve \( Y \) admits an antianalytic involution \( \tau \) then its branch point set \( B_Y \) is preserved by the antianalytic involution \( \hat{\tau} : \hat{\mathbb{C}} \to \hat{\mathbb{C}} \) induced by \( \tau \). Further, \( \hat{\tau} \) is conjugate to either the complex conjugation \( x \mapsto \overline{x} \) or the antipodal map \( x \mapsto -1/\overline{x} \) ([1], Thm 1.9.4). In the latter case \( Y \) has odd genus (see 2.2.(a) below). Since we are interested here in genus 2 curves, we only have to
consider those \((Y, \tau)\) such that \(\hat{\tau}\) is conjugate to the complex conjugation. The next lemma provides suitable plane models for such curves, [4].

**Lemma 2.1.** With the above notations, each hyperelliptic real curve \((Y, \tau)\) with \(\hat{\tau}(x) = \overline{x}\) is isomorphic to one and only one of the following:

(i) \((X = \{y^2 = P_X(x)\}, \sigma_X\)) where \(P_X\) is a monic real polynomial without multiple roots, or

(ii) \((X = \{y^2 = -P_X(x)\}, \sigma_X\)) where \(P_X\) is a monic real polynomial without multiple roots and without real roots,

and in both cases, \(\sigma_X : (x, y) \mapsto (\overline{x}, \overline{y})\).

Note that curves of type (ii) have empty real part \(X_{\sigma}\) since \(P_X\) is always positive on \(\mathbb{R}\). On the contrary, curves of type (i) have nonempty real part.

The complete topological description of any hyperelliptic real curve is given in the next lemma. It may be seen as a reformulation of results in [5], Section 6 adapted to our situation.

**Topological Classification Lemma 2.2.**

(a) Let \((X = \{y^2 = P_X\}, \tau)\) be a hyperelliptic real curve such that the antianalytic involution that \(\tau\) induces in \(\hat{\mathbb{C}}\) is conjugate to \(\hat{\tau} : x \mapsto -1/\overline{x}\). Then its topological type is \((g, 0, 1)\) with \(g = [(\deg P_X - 1)/2]\) odd.

(b) Let \((X = \{y^2 = P_X(x)\}, \sigma_X)\) be a hyperelliptic real curve of type (i) in the above lemma and let \(z_R\) be the number of real roots of \(P_X\). Then its topological type is \((g, k, \varepsilon)\) with \(g = [(\deg P_X - 1)/2]\) and

(b.1) \(k = g + 1, \varepsilon = 0\) if \(z_R = \deg P_X\),

(b.2) \(k = [(z_R + 1)/2]\) and \(\varepsilon = 1\) if \(0 < z_R < \deg P_X\),

(b.3) \(k = 1, \varepsilon = 0\) if \(z_R = 0\) and \(g\) is even,

(b.4) \(k = 2, \varepsilon = 0\) if \(z_R = 0\) and \(g\) is odd.

(c) Let \((X = \{y^2 = -P_X(x)\}, \sigma_X)\) be a hyperelliptic real curve of type (ii) in the above lemma. Then its topological type is \((g, 0, 1)\) with \(g = (\deg P_X - 2)/2\).

Since for curves of type (i) or (ii) (the only ones to be considered throughout this paper), the antianalytic involution is the complex conjugation, we will omit it and in the sequel we will simply say that \(X\) is a real curve.

Let us return to the problem of classifying hyperelliptic curves up to isomorphism. Recall that a necessary and sufficient condition for two complex hyperelliptic curves to be isomorphic is the existence of a (complex) Möbius transformation mapping the branch point set of one of them onto that of the other. However, the situation in the real case is more complicated: The existence of a real Möbius transformation mapping \(B_X\) onto \(B_Y\) does not assure \(X\) and \(Y\) to be isomorphic. This obstruction in the real case is detected by means of the formulae of liftings (see above). A case by case examination
of such formulae proves the following lemma. (A similar result for curves appearing in (a) of the Topological Classification Lemma is given in [4].)

\textbf{\(\mathbb{R}\)-lifting Lemma 2.3.} \textit{Let} \(X\) \textit{and} \(Y = \{y^2 = P_Y(x)\}\) \textit{be two hyperelliptic real curves of type (i) in Lemma 2.1 and let} \(m\) \textit{be a real Möbius transformation mapping the branch point set of} \(X\) \textit{onto that of} \(Y\). \textit{The liftings of} \(m\) \textit{are isomorphisms between} \(X\) \textit{and} \(Y\) \textit{if and only if:}

\begin{itemize}
  \item[-] \(\det m > 0\) if \(\infty \in B_X\) and \(m(\infty) = \infty\);
  \item[-] \(\det m \cdot P_Y(m(\infty)) < 0\) if \(\infty \in B_X\) and \(m(\infty) \neq \infty\);
  \item[-] \(\text{always if} \ \infty \not\in B_X\) \text{and} \(m(\infty) = \infty\);
  \item[-] \(P_Y(m(\infty)) > 0\) if \(\infty \not\in B_X\) \text{and} \(m(\infty) \neq \infty\).
\end{itemize}

\textit{The same holds true if both} \(X\) \textit{and} \(Y\) \textit{are of type (ii).}

In either case we will say that \(m\) has \textit{real liftings}. Isomorphisms between \(X\) and \(Y\) become automorphisms of \(X\) when \(Y = X\). The automorphism group \(\text{Aut}X\) of \(X\) consists of the liftings of those real Möbius transformations preserving \(B_X\) which, in addition, have real liftings. We denote by \(\text{Aut}_{\mathbb{R}}X\) the group of such Möbius transformations:

\[\text{Aut}_{\mathbb{R}}X := \{m : m \text{ real, } m(B_X) = B_X \text{ and } m \text{ has real liftings}\}.\]

Results of this section are applied to the particular case of genus 2 curves. In each of the following sections we give an explicit description of each \(\mathcal{M}(2,k,\varepsilon)\). We develop in detail only the case of \((2,3,0)\)-curves. In the rest of the cases, proofs are outlined.

\section{3. Moduli of \((2,3,0)\)-curves.}

The Topological Classification Lemma shows that all the branch points of a \((2,3,0)\)-curve lie on the real line \(\mathbb{R} \cup \{\infty\}\). An easy consequence of the \(\mathbb{R}\)-lifting lemma is that \(\infty\) may be fixed as one of them. Moreover, we have the following:

\textbf{Proposition 3.1.} \textit{Each} \((2,3,0)\)-\textit{curve is isomorphic to another of the form}

\[X(a,b,c) = \{y^2 = P_X(a,b,c)(x) := x(x-1)(x-a)(x-b)(x-c)\}\]

\textit{with} \(0 < a < b < c < 1\).

\textit{Proof.} Let \(B_Y = \{e_1, \ldots, e_5, \infty\}\) with \(e_1 < \cdots < e_5\) be the branch point set of a \((2,3,0)\)-curve \(Y\). The Möbius transformation \(m : x \mapsto (x - e_1)/(e_5 - e_1)\) maps \(B_Y\) onto \(B_X = \{0, a := m(e_2), b := m(e_3), c := m(e_4), 1, \infty\}\) with \(0 < a < b < c < 1\). So the curve \(X(a,b,c)\) \textit{we are looking for is the curve which ramifies over} \(B_X\) \textit{and given by the above equation. Indeed, it follows from the} \(\mathbb{R}\)-\textit{lifting lemma that a lifting of} \(m\) \textit{makes} \(X\) \textit{and} \(Y\) \textit{isomorphic.} \(\square\)
Let $T$ be the following open subset of $\mathbb{R}^3$:

$$T = \{(a, b, c) \in \mathbb{R}^3 : 0 < a < b < c < 1\}.$$  

The above proposition may be restated by saying that the mapping from $T$ to $M_{(2,3,0)}$ given by

$$T \rightarrow M_{(2,3,0)}(a,b,c) \mapsto \ X(a,b,c)$$

is surjective. Thus, in order to describe $M_{(2,3,0)}$ we have to find the fibres of this mapping. It turns out that they are orbits of points under the action of a finite group acting on $T$.

**Proposition 3.2.** Let $G$ be the dihedral group of order 6 generated by the involutions

$$\alpha : (a,b,c) \mapsto (a, a^{-1}c, a^{-1}b) \quad \text{and} \quad \beta : (a,b,c) \mapsto \left(\frac{b-c}{b-1}, \frac{a-c}{a-1}, c\right).$$

Then, with the above notations, $X(a',b',c')$ is isomorphic to $X(a,b,c)$ if and only if $(a',b',c') = \gamma(a,b,c)$ for some $\gamma \in G$.

**Proof.** We first have to determine all the real Möbius transformations which map $B_{X(a,b,c)} = \{\infty, 0, a, b, c, 1\}$ onto $B_{X(a',b',c')} = \{\infty, 0, a', b', c', 1\}$. If $m$ is such a transformation then its restriction to $\mathbb{R} - \{m^{-1}(\infty)\}$ is either strictly increasing or strictly decreasing according to the sign of its determinant. Therefore, the images by $m$ of the branch points of $X(a,b,c)$ are completely determined by the image of one of them and by the increasing or decreasing nature of $m$. This gives 12 Möbius transformations, one for each choice of, e.g., $m(\infty)$ and the sign of det $m$. We then have to apply the R-lifting lemma in order to find those having real liftings. For that we compute the sign of $P_{X(a',b',c')}$ at $m(\infty)$ (if $m(\infty) \neq \infty$). It appears in the fourth column of Table 1. In the fifth one we write “yes” or “no” according to whether $m$ has real liftings or not.

We next calculate the formula of each Möbius transformation having real liftings. In order to express it in terms of $a, b$ and $c$ we use the data $m_j^{-1}(0), m_j^{-1}(\infty)$ and $m_j^{-1}(1)$. We obtain the following six formulae:

\[
\begin{align*}
m_1(x) &= x; & m_4(x) &= \frac{a}{x}; & m_5(x) &= \frac{b-c}{b-1} \cdot \frac{x-1}{x-c}; \\
m_8(x) &= \frac{c-b}{c-a} \cdot \frac{x-a}{x-b}; & m_9(x) &= \frac{a}{b} \cdot \frac{x-b}{x-a}; & m_{12}(x) &= \frac{x-c}{x-1}.
\end{align*}
\]

Calculating finally $m_j(B_{X(a,b,c)})$ for each $m_j$ and arranging in increasing order the images different to 0, 1 and $\infty$, we get the points $(a',b',c')$ with $0 < a' < b' < c' < 1$ we are looking for. They are precisely the images of $(a,b,c)$ under the elements of the group $G$ in the statement of the proposition. □
### Table 1

<table>
<thead>
<tr>
<th>$m_j(\infty)$</th>
<th>$\det m_j$</th>
<th>$P'_{X(a',b',c')}(m_j(\infty))$</th>
<th>Real liftings</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_1$</td>
<td>$\infty$</td>
<td>+</td>
<td>Yes</td>
</tr>
<tr>
<td>$m_2$</td>
<td>$\infty$</td>
<td>$-$</td>
<td>No</td>
</tr>
<tr>
<td>$m_3$</td>
<td>0</td>
<td>+</td>
<td>No</td>
</tr>
<tr>
<td>$m_4$</td>
<td>0</td>
<td>$-$</td>
<td>Yes</td>
</tr>
<tr>
<td>$m_5$</td>
<td>$a'$</td>
<td>+</td>
<td>Yes</td>
</tr>
<tr>
<td>$m_6$</td>
<td>$a'$</td>
<td>$-$</td>
<td>No</td>
</tr>
<tr>
<td>$m_7$</td>
<td>$b'$</td>
<td>+</td>
<td>No</td>
</tr>
<tr>
<td>$m_8$</td>
<td>$b'$</td>
<td>$-$</td>
<td>Yes</td>
</tr>
<tr>
<td>$m_9$</td>
<td>$c'$</td>
<td>+</td>
<td>Yes</td>
</tr>
<tr>
<td>$m_{10}$</td>
<td>$c'$</td>
<td>$-$</td>
<td>No</td>
</tr>
<tr>
<td>$m_{11}$</td>
<td>1</td>
<td>+</td>
<td>No</td>
</tr>
<tr>
<td>$m_{12}$</td>
<td>1</td>
<td>$-$</td>
<td>Yes</td>
</tr>
</tbody>
</table>

As a consequence, each isomorphism class of a $(2,3,0)$-curve is represented by one and only one point of the quotient space $T/G$. The next proposition describes a fundamental set for the action of $G$ on $T$.

**Proposition 3.3.** The mapping $(a, b, c) \mapsto \text{orbit } (a, b, c)$ is a bijection between the subset $\Delta$ of $T$ given by $\Delta = \{ bc \leq a \leq b - c \}$ and the quotient $T/G$.

**Proof.** The generating involutions $\alpha$ and $\beta$ act as two reflections with respect to two surfaces of $T$ intersecting along a curve of $T$. More precisely, the fixed point set of $\alpha$ is the surface of $T \text{ fix}(\alpha) = \{ a = bc \}$, and the two semispaces in which it divides $T$ are interchanged by $\alpha$, i.e., $(a, b, c) \in \{ a > bc \}$ if and only if $\alpha(a, b, c) \in \{ a < bc \}$. Similarly, $(a, b, c) \in \{ a > (b - c)/(b - 1) \}$ if and only if $\beta(a, b, c) \in \{ a < (b - c)/(b - 1) \}$, where $\{ a = (b - c)/(b - 1) \}$ is the fixed point set of $\beta$. Straightforward calculations show that the intersection $\{ a \geq bc \} \cap \{ a \leq (b - c)/(b - 1) \}$ is a fundamental set $\Delta$ for the action of $G$ on $T$. \[ \Box \]

Summarizing, we have the following:

**Theorem 3.4.** The mapping

$$\Psi : (a, b, c) \mapsto X(a, b, c) = \{ y^2 = x(x - 1)(x - a)(x - b)(x - c) \}$$

is a bijection between

$$\Delta = \left\{ (a, b, c) \in \mathbb{R}^3 : 0 < a < b < c < 1, \; bc \leq a \leq \frac{b - c}{b - 1} \right\} \text{ and } \mathcal{M}_{(2,3,0)}.$$
The next step for a better description of $\mathcal{M}_{(2,3,0)}$ is to identify in $\Delta$ its strata, i.e., its subsets $\mathcal{M}_{(2,3,0)}(H)$ consisting of curves whose automorphism group is $H$. It is clear that points in $T$ with nontrivial $G$-stabilizer represent curves with "nontrivial" automorphism group, i.e., other than $\langle h_X \rangle$. Hence, the strata of $\mathcal{M}_{(2,3,0)}$ are described in terms of fixed point sets of elements of $G$. In addition, we can calculate explicitly the automorphism group of each $(2,3,0)$-curve since we know all the Möbius transformations which permute its branch points.

**Proposition 3.5.** The subsets $D = \{ bc < a < \frac{b-c}{b+1} \}$, $S = \{ bc = a < \frac{b-c}{b+1} \}$ \cup $\{ bc < a = \frac{b-c}{b-1} \}$ and $C = \{ bc = a = \frac{b-c}{b+1} \}$ correspond, via the bijection $\Psi$ described above, to $\mathcal{M}_{(2,3,0)}(C_2)$, $\mathcal{M}_{(2,3,0)}(D_2)$ and $\mathcal{M}_{(2,3,0)}(D_6)$ respectively. These are the only strata of $\mathcal{M}_{(2,3,0)}$. Furthermore, the explicit formulae of the automorphisms of the curves belonging to each stratum are given.

**Proof.** Let $m$ be any of the 6 Möbius transformations with real liftings we find in Proposition 3.2. We just have to calculate which conditions $a$, $b$ and $c$ must fulfill so that $m(B_{X(a,b,c)})$ coincides with $B_{X(a,b,c)}$. With the same notations:

- $m_4(x) = \frac{a}{x}$ preserves $B_{X(a,b,c)}$ if and only if $a = bc$.
- $m_5(x) = \frac{b-c}{b+1} \frac{x-1}{x-c}$ preserves $B_{X(a,b,c)}$ if and only if $bc = a = \frac{b-c}{b+1}$.
- $m_8(x) = \frac{c-b}{c-a} \frac{x-c}{x-b}$ preserves $B_{X(a,b,c)}$ if and only if $a = 1 + c - \frac{c}{b}$.
- These are the only strata of $\mathcal{M}_{(2,3,0)}$. It is clear that points in $H$ have real liftings that preserve $B_{X(a,b,c)}$ if and only if $a = \frac{b-c}{b+1}$; therefore $m_8$ preserves $B_{X(a,b,c)}$ if and only if $bc = a = \frac{b-c}{b+1}$.
- $m_9(x) = \frac{a}{b} \frac{x-b}{x-a}$ preserves $B_{X(a,b,c)}$ if and only if $bc = a = \frac{b-c}{b+1}$.
- $m_{12}(x) = \frac{x-c}{x-a}$ preserves $B_{X(a,b,c)}$ if and only if $a = \frac{b-c}{b+1}$.

Recall that the group of Möbius transformations with real liftings that preserve $B_{X(a,b,c)}$ is denoted by $\text{Aut} X(p) \overline{\mathbb{C}}$, where $p = (a, b, c)$ for short. Let

\[ f : (x, y) \mapsto \left( \frac{a}{x}, \frac{-ya^{3/2}}{x^3} \right) \quad \text{and} \quad g : (x, y) \mapsto \left( \frac{x-c}{x-1}, \frac{y(1-c)^{3/2}}{(x-1)^3} \right) \]

be liftings of $m_4$ and $m_{12}$, respectively. As a consequence of the above computations:

(i) If $p \in D = \{ bc < a < \frac{b-c}{b+1} \}$ then none of the above Möbius transformations preserves $B_{X(p)}$, i.e., $\text{Aut} X(p) \overline{\mathbb{C}} = \{ \text{id} \overline{\mathbb{C}} \}$ and so $\text{Aut} X(p) = \langle h_X \rangle \simeq C_2$.

(ii) If $p \in \{ bc = a < \frac{b-c}{b+1} \}$ then $\text{Aut} X(p) \overline{\mathbb{C}} = \{ \text{id} \overline{\mathbb{C}}, m_4 \}$ and so $\text{Aut} X(p) = \langle h_X, f \rangle \simeq D_2$. If $p \in \{ bc < a = \frac{b-c}{b+1} \}$ then $\text{Aut} X(p) \overline{\mathbb{C}} = \{ \text{id} \overline{\mathbb{C}}, m_{12} \}$ and so $\text{Aut} X(p) = \langle h_X, g \rangle \simeq D_2$.

(iii) If $p \in C = \{ bc = a = \frac{b-c}{b+1} \}$ then $\text{Aut} X(p) \overline{\mathbb{C}} = \{ \text{id} \overline{\mathbb{C}}, m_4, m_5, m_8, m_9, m_{12} \}$. It turns out that $\text{Aut} X(p) = \langle f, g \rangle \simeq D_6$. \qed
Figure 1 illustrates what $\Delta$ looks like (dashed lines do not belong to $\Delta$). In it we identify its subsets corresponding to the strata of $\mathcal{M}_{(2,3,0)}$. The stratum $\mathcal{M}_{(2,3,0)}(C_2)$ corresponds to the interior of $\Delta$ in $T$, which is the domain bounded by the two shadowed surfaces.

Figure 1. The moduli space of $(2,3,0)$-curves.

4. Moduli of $(2,2,1)$-curves.

By the Topological Classification Lemma, exactly 2 of the 6 branch points of a $(2,2,1)$-curve are complex non-real. It is easy to see that we may fix $i := \sqrt{-1}$, $-i$ and $\infty$ as branch points.

**Proposition 4.1.** Write $T = \{(a, b, c) \in \mathbb{R}^3 : a < b < c\}$. The mapping $T \to \mathcal{M}_{(2,2,1)}$ given by $(a, b, c) \mapsto X(a, b, c) = \{y^2 = (x^2 + 1)(x - a)(x - b)(x - c)\}$ is surjective.

Its fibres are orbits under the action of a finite group acting on $T$.

**Proposition 4.2.** Let $G$ be the dihedral group of order 4 generated by the involutions
\[
\alpha : (a, b, c) \mapsto \left( a, \frac{ac + 1}{c - a}, \frac{ab + 1}{b - a} \right) \quad \text{and} \quad \beta : (a, b, c) \mapsto \left( \frac{bc + 1}{b - c}, \frac{ac + 1}{a - c}, c \right).
\]

Then, with the above notations, $X(a', b', c')$ is isomorphic to $X(a, b, c)$ if and only if $(a', b', c') = \gamma(a, b, c)$ for some $\gamma \in G$. 

Outline of Proof. A real Möbius transformation mapping $B_{X(a,b,c)} = \{\infty, a, b, c, i, -i\}$ onto $B_{X(a',b',c')} = \{\infty, a', b', c', i, -i\}$ preserves the set $\{i, -i\}$. There exist 8 such Möbius transformations, half of them with real liftings. Namely, (writing their formulae in terms of $a, b$ and $c$):

$$m_1(x) = x, \quad m_2(x) = \frac{ax + 1}{x - a}, \quad m_3(x) = \frac{bx + 1}{-x + b} \quad \text{and} \quad m_4(x) = \frac{cx + 1}{x - c}.$$  

Calculating then $m_j(\{\infty, a, b, c, i, -i\})$ for each $m_j$ and arranging in increasing order the finite real values, we get the points $(a', b', c') \in T$ we are looking for. They are precisely the images of $(a, b, c)$ under the elements of $G$. \(\square\)

The generating involutions $\alpha$ and $\beta$ act as two reflections with respect to the surfaces $\text{fix}(\alpha) = \{b = (ac + 1)/(c - a)\}$ and $\text{fix}(\beta) = \{b = (ac + 1)/(a - c)\}$, which intersect along a curve in $T$. It follows that $\Delta = \{(ac + 1)(a - c) \leq b \leq (ac + 1)(c - a)\}$ is a fundamental set for the action of $G$ in $T$. This proves half of the following theorem.

**Theorem 4.3.** The mapping  

$$\Psi : (a, b, c) \mapsto X(a, b, c) = \{y^2 = (a^2 + 1)(x - a)(x - b)(x - c)\}$$  

is a bijection between  

$$\Delta = \left\{ (a, b, c) \in \mathbb{R}^3 : a < b < c, \quad \frac{ac + 1}{a - c} \leq b \leq \frac{ac + 1}{c - a} \right\} \quad \text{and} \quad \mathcal{M}_{(2,2,1)}.$$  

The subsets $D = \left\{ \frac{ac + 1}{a - c} < b < \frac{ac + 1}{c - a} \right\}$, $S = \left\{ \frac{ac + 1}{a - c} = b < \frac{ac + 1}{c - a} \right\} \cup \left\{ \frac{ac + 1}{a - c} < b = \frac{ac + 1}{c - a} \right\}$ and $C = \{ac + 1 = b = 0\}$ correspond, via $\Psi$, to $\mathcal{M}_{(2,2,1)}(C_2)$, $\mathcal{M}_{(2,2,1)}(D_2)$ and $\mathcal{M}_{(2,2,1)}(D_4)$ respectively. These are the only strata of $\mathcal{M}_{(2,2,1)}$. Furthermore, the explicit formulae of the automorphisms of the curves belonging to each stratum are given.

**Proof.** The only Möbius transformations with real liftings which may preserve $B_{X(a,b,c)}$ are those appearing in the proof of Proposition 4.2. With the same notations:

- $m_2(x) = \frac{ax + 1}{x-a}$ preserves $B_{X(a,b,c)}$ if and only if $b = \frac{ac+1}{c-a}$.
- $m_3(x) = \frac{bx+1}{x-b}$ preserves $B_{X(a,b,c)}$ if and only if $ac + 1 = b = 0$.
- $m_4(x) = \frac{cx+1}{x-c}$ preserves $B_{X(a,b,c)}$ if and only if $b = \frac{ac+1}{a-c}$.

Let

$$f : (x, y) \mapsto \left( \frac{cx + 1}{x - c}, \frac{y(c^2 + 1)^{3/2}}{(x - c)^3} \right),$$  

and

$$g : (x, y) \mapsto \left( \frac{ax + 1}{x - a}, \frac{y(a^2 + 1)^{3/2}}{(x - a)^3} \right).$$
be liftings of \( m_4 \) and \( m_2 \) respectively. Denoting \( p = (a,b,c) \) for short, we conclude:

(i) If \( p \in D = \{ \frac{ac+1}{a-c} < b < \frac{ac+1}{c-a} \} \) then \( \text{Aut}_X(p)_{\bar{\mathbb{C}}} = \{ \text{id}_{\bar{\mathbb{C}}} \} \) and so \( \text{Aut}_X(p) = \langle h_X \rangle \simeq C_2 \).

(ii) If \( p \in \{ \frac{ac+1}{a-c} = b < \frac{ac+1}{c-a} \} \) then \( \text{Aut}_X(p)_{\bar{\mathbb{C}}} = \{ \text{id}_{\bar{\mathbb{C}}}, m_4 \} \) and so \( \text{Aut}_X(p) = \langle h_X, f \rangle \simeq D_2 \).

(iii) If \( p \in C = \{ ac+1 = b = 0 \} \) then \( \text{Aut}_X(p)_{\bar{\mathbb{C}}} = \{ \text{id}_{\bar{\mathbb{C}}}, m_2, m_3, m_4 \} \). In this case \( \text{Aut}_X(p) = \langle f, g \rangle \simeq D_4 \).

\[ \square \]

Figure 2 illustrates what \( \Delta \) looks like (dashed lines do not belong to \( \Delta \)). The stratum \( M_{(2,2,1)}(C_2) \) corresponds to the interior of \( \Delta \) in \( T \), which is the domain bounded by the two shadowed surfaces.

\[ \begin{array}{c}
\text{Figure 2. The moduli space of (2,2,1)-curves.}
\end{array} \]

5. Moduli of (2,1,1)-curves.

These curves have two real branching points and four complex non-real ones. We may fix 0 and \( \infty \) as the real ones and the imaginary part of two of the others.
Proposition 5.1. Write \( T = \{ b > 0, (a, b) \neq (c, 1) \} \subset \mathbb{R}^3 \). The mapping \( T \to \mathcal{M}_{(2,1,1)} \) given by \((a, b, c) \mapsto X(a, b, c) = \{ y^2 = x((x - a)^2 + b^2)((x - c)^2 + 1) \} \) is surjective.

Its fibres are calculated in the next proposition. Its proof is similar to that of Proposition 4.2 and so we just indicate the formulae of the unique 4 Möbius transformations with real liftings mapping \( B_{X(a, b, c)} \) onto \( B_{X(a', b', c')}. \)

Apart from the identity, they are the following:

\[
m_2(x) = \frac{x}{b}, \quad m_3(x) = \frac{c^2 + 1}{x}, \quad \text{and} \quad m_4(x) = \frac{a^2 + b^2}{bx}.
\]

Proposition 5.2. Let \( G \) be the dihedral group of order 4 generated by the involutions

\[
\alpha : (a, b, c) \mapsto \left( a \cdot \frac{c^2 + 1}{a^2 + b^2}, b \cdot \frac{c^2 + 1}{a^2 + b^2}, c \right) \quad \text{and}
\]

\[
\beta : (a, b, c) \mapsto \left( \frac{c}{b}, \frac{a^2 + b^2 - 1}{c^2 + 1}, \frac{a^2 + b^2 - a}{c^2 + 1}, \frac{1}{b} \right).
\]

Then \( X(a', b', c') \) is isomorphic to \( X(a, b, c) \) if and only if \((a', b', c') = \gamma(a, b, c)\) for some \( \gamma \in G \).

A fundamental set for the action of \( G \) on \( T \) is the set \( \Delta \) appearing in the next theorem.

Theorem 5.3. The mapping \( \Psi : (a, b, c) \mapsto X(a, b, c) = \{ y^2 = x((x - a)^2 + b^2)((x - c)^2 + 1) \} \) is a bijection between

\[
\Delta = \{ (a, b, c) \in \mathbb{R}^3 : b > 0, (a, b) \neq (c, 1), a \geq bc, a^2 + b^2 \geq c^2 + 1 \}
\]

and \( \mathcal{M}_{(2,1,1)} \). The subsets \( D = \{ a > bc, a^2 + b^2 > c^2 + 1 \} \) and \( S = \{ a = bc \} \cup \{ a^2 + b^2 = c^2 + 1 \} \) correspond, via \( \Psi \), to \( \mathcal{M}_{(2,1,1)}(C_2) \) and \( \mathcal{M}_{(2,1,1)}(D_2) \) respectively. These are the only strata of \( \mathcal{M}_{(2,1,1)} \). Furthermore, the explicit formulae of the automorphisms of the curves belonging to each stratum are given.

Proof. We prove the claims concerning the strata. The only Möbius transformations with real liftings which may preserve \( B_{X(a, b, c)} \) are given above. With the same notations:

- \( m_2(x) = x/b \) preserves \( B_{X(a, b, c)} \) if and only if \( a = c \) and \( b = 1 \). However, \( a, b \) and \( c \) cannot take these values.
- \( m_3(x) = (c^2 + 1)/x \) preserves \( B_{X(a, b, c)} \) if and only if \( a^2 + b^2 = c^2 + 1 \).
- \( m_4(x) = (a^2 + b^2)/(bx) \) preserves \( B_{X(a, b, c)} \) if and only if \( a = bc \).
Let
\[ f : (x, y) \mapsto \left( \frac{b(c^2 + 1)}{x}, \frac{y(b(c^2 + 1))^{3/2}}{x^3} \right) \]
and
\[ g : (x, y) \mapsto \left( \frac{c^2 + 1}{x}, \frac{y(c^2 + 1)^{3/2}}{x^3} \right) \]
be liftings of \( m_4 \) and \( m_3 \) respectively. Taking into account that no point \( p = (a, b, c) \in \Delta \) fulfills simultaneously the equations \( a = bc \) and \( a^2 + b^2 = c^2 + 1 \), we get:

(i) If \( p \in D = \{ a > bc, a^2 + b^2 > c^2 + 1 \} \) then \( \text{Aut}_X(p)\mathfrak{C} = \{ \text{id}\mathfrak{C} \} \), and so \( \text{Aut}_X(p) = \langle h_X \rangle \simeq C_2 \).

(ii) If \( p \in \{ a = bc \} \) then \( \text{Aut}_X(p)\mathfrak{C} = \{ \text{id}\mathfrak{C}, m_4 \} \) and so \( \text{Aut}_X(p) = \langle h_X, f \rangle \simeq D_2 \). If \( p \in \{ a^2 + b^2 = c^2 + 1 \} \) then \( \text{Aut}_X(p)\mathfrak{C} = \{ \text{id}\mathfrak{C}, m_3 \} \)
and so \( \text{Aut}_X(p) = \langle h_X, g \rangle \simeq D_2 \).

Figure 3 illustrates what \( \Delta \) looks like (dashed lines do not belong to \( \Delta \)). Note that, unlike the preceding cases, the closures in \( \Delta \) of the two surfaces corresponding to \( \mathcal{M}_{(2,2,1)}(D_2) \) do not intersect.

**Figure 3.** The moduli space of \((2,1,1)\)-curves.
6. Moduli of (2,1,0) and (2,0,1)-curves.

We may study simultaneously the moduli sets of (2,1,0) and (2,0,1)-curves. Indeed, as a consequence of the Topological Classification Lemma, if \( P_X \) is a monic real polynomial of degree 6 with no real root then \( X = \{ y^2 = P_X \} \) is a (2,1,0)-curve (and any such curve is like this) while \( X' = \{ y^2 = -P_X \} \) is a (2,0,1)-curve (and any such curve is like this). Therefore, roots of a polynomial as the above serve as parameters to describe the moduli sets of curves of both types. Furthermore, concerning the classification up to isomorphism, an easy computation shows the following.

**Lemma 6.1.** A mapping \( f : X \to Y \) is an isomorphism between \( X = \{ y^2 = P_X \} \) and \( Y = \{ y^2 = P_Y \} \) if and only if \( f \) is an isomorphism between \( X' = \{ y^2 = -P_X \} \) and \( Y' = \{ y^2 = -P_Y \} \). In particular, the automorphism group of \( X \) coincides with that of \( X' \).

Consequently, along this section we only need to deal with (2,1,0)-curves, for instance.

**Remark 6.2.** Curves of these topological types are described by polynomials which are positive on \( \mathbb{R} \). In particular \( \infty \) is not a branch point and so the \( \mathbb{R} \)-lifting lemma is superfluous along this section: Every real Möbius transformation mapping the branch point set of \( X \) onto that of \( Y \) has real liftings.

The description of \( \mathcal{M}_{(2,1,0)} \) and \( \mathcal{M}_{(2,0,1)} \) by means of the real and imaginary parts of roots of \( P_X \) is more involved than in the preceding sections. However, the fact that no branch point of \( X \) lies in the real axis \( \mathbb{R} \cup \{ \infty \} \) allows us to consider *hyperbolic distances* between those with positive imaginary part. The advantage of working with such distances is that they are preserved by real Möbius transformations. Indeed, restrictions to the upper half plane \( \mathcal{H} \) of those with positive determinant are the direct isometries of \( \mathcal{H} \), whilst the composite of the complex conjugation with those having negative determinant are its inverse ones. Now, the 3 branch points of a (2,1,0)-curve lying on \( \mathcal{H} \) define a hyperbolic triangle or a hyperbolic segment. Therefore, since here the \( \mathbb{R} \)-lifting lemma is superfluous, *two (2,1,0)-curves are isomorphic if and only if their corresponding hyperbolic triangles (or segments) are \( \mathcal{H} \)-isometric*.

It is clear that any hyperbolic triangle or hyperbolic segment is isometric to another with vertices \( i, ai \) and \( b + ci \) for some \( a \in (0,1) \) and \( b \geq 0 \). We can now introduce the following parameters:

\[
\delta_1 = \rho(i, ai), \quad \delta_2 = \rho(ai, b + ci), \quad \delta_3 = \rho(i, b + ci),
\]

where \( \rho \) is the hyperbolic distance. Note that \( 0 < \delta_i \leq \delta_j + \delta_k \), for \( \{i, j, k\} = \{1, 2, 3\} \) because they are distances between different points. This way, each isomorphism class of a (2,1,0)-curve may be represented by a
triple \((\delta_1, \delta_2, \delta_3)\). In order to get unicity in this representation we impose \(\delta_1 \leq \delta_2 \leq \delta_3\). This shows that

\[
\Delta_{\rho} = \{0 < \delta_1 \leq \delta_2 \leq \delta_3 \leq \delta_1 + \delta_2\}
\]

is in bijective correspondence with \(\mathcal{M}_{(2,1,0)}\).

We now recover the real and imaginary parts of roots of polynomials defining \((2,1,0)\)-curves as the parameters to describe \(\mathcal{M}_{(2,1,0)}\). For that we translate the restrictions defining \(\Delta_{\rho}\) to restrictions on \(a, b, c\). Using the explicit formula for the hyperbolic distance

\[
\rho(z, w) = \ln \left\{ \frac{|z - w| + |z - w|}{|z - w| - |z - w|} \right\}
\]

we get the expressions of \(\delta_1, \delta_2, \delta_3\) in terms of \(a, b, c\). Then, easy computations give:

- \(\delta_1 \leq \delta_2\) if and only if \((a^2 - c)(c - 1) \leq b^2\),
- \(\delta_2 \leq \delta_3\) if and only if \(b^2 \leq a - c^2\),
- \(\delta_3 \leq \delta_1 + \delta_2\) always, and \(\delta_3 = \delta_1 + \delta_2\) if and only if \(b = 0\).

Note that if \(\delta_3 = \delta_1 + \delta_2\) then \(ai\) lies in the interior of the segment joining \(i\) with \(ci\); that is, if \(b = 0\) then \(0 < c < a < 1\), and so in this case condition \(\delta_1 \leq \delta_2\) is equivalent to \(c \leq a^2\).

We can now formulate the main result of this section:

**Theorem 6.3.** The mapping

\[
\Psi : (a, b, c) \mapsto X(a, b, c) = \{y^2 = (x^2 + 1)(x^2 + a^2)((x - b)^2 + c^2)\},
\]

(respectively

\[
\Psi : (a, b, c) \mapsto X'(a, b, c) = \{y^2 = -(x^2 + 1)(x^2 + a^2)((x - b)^2 + c^2)\},
\]

is a bijection between

\[
\Delta = \left\{0 < a < 1, \ b \geq 0, \ c > 0, \ (0, a) \neq (b, c) \neq (0, 1), \right. \\
\left. (a^2 - c)(c - 1) \leq b^2 \leq a - c^2\right\}
\]

and \(\mathcal{M}_{(2,1,0)}\) (respectively \(\mathcal{M}_{(2,0,1)}\)). Furthermore, the subsets of \(\Delta\)

\[
D = \{b > 0, \ (a^2 - c)(c - 1) < b^2 < a - c^2\},
\]

\[
S = \{b > 0, \ (a^2 - c)(c - 1) = b^2 < a - c^2\} \cup \{b > 0, \ (a^2 - c)(c - 1) < b^2 = a - c^2\} \cup \{b = 0, \ c < a^2\},
\]

\[
C = \{b = 0, \ c = a^2\} \quad \text{and}
\]

\[
L = \{b > 0, \ (a^2 - c)(c - 1) = b^2 = a - c^2\}
\]

correspond, via \(\Psi\), respectively to \(\mathcal{M}_{(2,1,0)}(C_2), \mathcal{M}_{(2,1,0)}(D_2), \mathcal{M}_{(2,1,0)}(D_4)\)

and \(\mathcal{M}_{(2,1,0)}(D_6)\) (respectively to \(\mathcal{M}_{(2,0,1)}(C_2), \mathcal{M}_{(2,0,1)}(D_2), \mathcal{M}_{(2,0,1)}(D_4)\)

and \(\mathcal{M}_{(2,0,1)}(D_6)\)). These are the only strata of \(\mathcal{M}_{(2,1,0)}\) (respectively of
$\mathcal{M}_{(2,0,1)}$. Furthermore, the explicit formulae of the automorphisms of the curves belonging to each stratum are given.

Proof. In order to calculate the automorphism group of $X(a,b,c)$ we have to find the isometry group of the hyperbolic triangle (or segment) determined by $i$, $ai$ and $b + ci$.

Case 1. If $\delta_3 = \delta_1 + \delta_2$ then $i$, $ai$ and $ci$ determine a segment with $ai$ in its interior. So any isometry $s$ of this segment fixes $ai$.

1.1: If $s$ also fixes $i$ and $ci$ then $s$ is either the identity or the reflection $x \leftrightarrow -\overline{x}$ in the imaginary axis.

1.2: If $s$ interchanges $i$ and $ci$ then $s$ is the reflection in the $\mathcal{H}$-line orthogonal to the imaginary axis at $ai$, i.e., $x \mapsto a^2/\overline{x}$. Note that this case happens only if $\delta_1 = \delta_2$.

Case 2. If $\delta_3 \neq \delta_1 + \delta_2$ then $i$, $ai$ and $b + ci$ determine a hyperbolic triangle.

2.1: If $\delta_1 < \delta_2 < \delta_3$ then the triangle is scalene, and so only the identity preserves it.

2.2: If $\delta_1 = \delta_2 < \delta_3$ then the triangle is isosceles non-equilateral; the only nontrivial isometry that preserves it is the reflection in the angle bisector at $ai$, i.e., $x \mapsto (b\overline{x} + a^2(1 - c))/((1 - c)x - b)$.

2.3: If $\delta_1 < \delta_2 = \delta_3$ then the triangle is also isosceles non-equilateral; the only nontrivial isometry that preserves it is the reflection in the angle bisector at $b + ci$, i.e., $x \mapsto a/\overline{x}$.

2.4: If $\delta_1 = \delta_2 = \delta_3$ then the triangle is equilateral and so its isometry group is generated by the two reflections described above.

Viewing isometries of $\mathcal{H}$ as real Möbius transformations we obtain the corresponding liftings:

$$f_1 : (x,y) \mapsto (-x,y), \quad f_2 : (x,y) \mapsto \left(\frac{a^2}{x}, \frac{ya^3}{x^3}\right),$$

$$f_3 : (x,y) \mapsto \left(\frac{a}{x}, \frac{ya^{3/2}}{x^3}\right) \quad \text{and}$$

$$f_4 : (x,y) \mapsto \left(\frac{bx + a^2(1 - c)}{(1 - c)x - b}, \frac{y \cdot [(1 - a^2)(c - c^2)]^{3/2}}{[(1 - c)x - b]^3}\right).$$

Translating conditions on $\delta_1$, $\delta_2$ and $\delta_3$ into conditions of $a$, $b$ and $c$ we get the following, where $p = (a,b,c)$ for short:

(i) If $p \in \{b > 0, (a^2 - c)(c - 1) < b^2 < a - c^2\}$ then $\text{Aut}X(p) = \langle h_X \rangle \simeq C_2$.
(ii) If $p \in \{b = 0, c < a^2\}$ then $\text{Aut}X(p) = \langle h_X, f_1 \rangle \simeq D_2$. If $p \in \{b > 0, (a^2 - c)(c - 1) = b^2 < a - c^2\}$ then $\text{Aut}X(p) = \langle h_X, f_4 \rangle \simeq D_2$. If $p \in \{b > 0, (a^2 - c)(c - 1) < b^2 = a - c^2\}$ then $\text{Aut}X(p) = \langle h_X, f_3 \rangle \simeq D_2$. 


(iii) If \( p \in \{ b = 0, \ a^2 = c \} \) then \( \text{Aut} \, X(p) = \langle f_1, f_2 \rangle \simeq D_4 \).

(iv) If \( p \in \{ b > 0, \ (a^2 - c)(c - 1) = b^2 = a - c^2 \} \) then \( \text{Aut} \, X(p) = \langle f_3, f_4 \rangle \simeq D_6 \).

\[ \square \]

In order to see what \( \Delta \) looks like, it is easier to think of its description in terms of \( \delta_1, \delta_2 \) and \( \delta_3 \), i.e., to think of \( \Delta_\rho \). Figure 4 illustrates the relative position of the strata of \( \Delta_\rho \). There are three boundary surfaces, which correspond to \( \mathcal{M}_{(2,1,0)}(D_2) \). Their closures intersect pairwise in three lines; two of them correspond to \( \mathcal{M}_{(2,1,0)}(D_4) \) and \( \mathcal{M}_{(2,1,0)}(D_6) \), whilst the third (the dashed one) does not belong to \( \Delta \). The domain bounded by the three shadowed surfaces corresponds to \( \mathcal{M}_{(2,1,0)}(C_2) \).

\[ \text{Figure 4. The moduli spaces of } (2,1,0) \text{ and } (2,0,1)-\text{curves.} \]

### 7. Real analyticity.

In the preceding sections we gave, for each of the five different topological types of real algebraic curves of genus 2, a bijection \( \Psi_{(2,k,\varepsilon)} : \Delta_{(2,k,\varepsilon)} \to \mathcal{M}_{(2,k,\varepsilon)} \) between a semialgebraic subset \( \Delta_{(2,k,\varepsilon)} \) of \( \mathbb{R}^3 \) and the moduli space \( \mathcal{M}_{(2,k,\varepsilon)} \). This latter is known to have a natural semianalytic structure. A natural question arises: Whether the above mapping \( \Psi_{(2,k,\varepsilon)} \) is real analytic or not. In this section we answer the question in the affirmative.

**Theorem 7.1.** Each bijection \( \Psi_{(2,k,\varepsilon)} : \Delta_{(2,k,\varepsilon)} \to \mathcal{M}_{(2,k,\varepsilon)} \) described in the preceding sections is real analytic.
Let us begin with the case of \((2,3,0)\)-curves. With the notations of Section 3, consider the real analytic manifold \(T_{(2,3,0)} = T = \{(a, b, c) \in \mathbb{R}^3 : 0 < 0 < b < c < 1\}\). It contains \(\Delta_{(2,3,0)}\) and the mapping \(\Psi_{(2,3,0)}\) extends to \(T\) in the obvious way. Of course, it suffices to show that this extension is real analytic. Let \(\mathcal{R}\) be the sheaf of real analytic functions on \(T\) and consider the polynomial \(f \in \mathcal{R}(T)[X,Y]\) given by

\[
f = Y^2 - X(X - 1)(X - u_1)(X - u_2)(X - u_3),
\]

where each \(u_i \in \mathcal{R}(T)\) is the \(i\)th projection. Consider, with the notations in [7, Section 22], the triple \((\mathcal{C}, \mathcal{U}, \varphi)\) where \(\mathcal{U}\) is the trivial open covering of \(T\), \(\mathcal{C}\) is given by \(\mathcal{C}(T) = \mathcal{R}(T)[X,Y] / (f)\), and the gluing data \(\varphi\) is superfluous here. Notice that for each \(p = (a, b, c) \in T\) the fiber \(\mathcal{C}_p\) coincides with \(\mathcal{C}(T) \otimes_{\mathcal{R}(T)} \mathbb{R}\), where \(\mathbb{R}\) is considered as an \(\mathcal{R}(T)\)-algebra via the evaluation map at \(p\). Hence, \(\mathcal{C}_p\) coincides with \(\mathbb{R}[X,Y] / (f_p)\), where \(f_p = Y^2 - X(X - 1)(X - a)(X - b)(X - c)\). Since \(p \in T\), the geometric fiber

\[
\text{Spec} \mathcal{C}_p(\mathbb{R}) = \{(x, y) \in \mathbb{R}^2 : f_p(x, y) = 0\}
\]

is a real algebraic curve of topological type \((2,3,0)\).

Since \(f\) is monic it follows that the triple \((\mathcal{C}, \mathcal{U}, \varphi)\) is an analytic family of real algebraic curves of topological type \((2,3,0)\) as defined in [7]. Therefore a direct application of Theorem 22.2 in [7] gives that the map \(\Psi_{(2,3,0)} : T \to \mathcal{M}_{(2,3,0)}\) is real analytic.

For the rest of topological types the proof is exactly the same, changing \(f\) by the appropriate polynomial, because the mapping \(\Psi_{(2,k,\varepsilon)}\) extends to an open subset \(T\) of \(\mathbb{R}^3\). For \((2,2,1)\) and \((2,1,1)\)-curves, we may take \(T\) as in Propositions 4.1 and 5.1, respectively. For \((2,1,0)\) and \((2,0,1)\)-curves, it is enough to choose as \(T\) a sufficiently small open neighbourhood of the subset \(\Delta\) defined in Theorem 6.3 to which \(\Psi_{(2,k,\varepsilon)}\) extends.

**Remark.** The image under \(\Psi_{(2,k,\varepsilon)}\) of the interior of \(\Delta_{(2,k,\varepsilon)}\) is the set of (isomorphism classes of) \((2,k,\varepsilon)\)-curves whose full automorphism group has order 2. By Theorem 5.1 in [6] this is the complement in \(\mathcal{M}_{(2,k,\varepsilon)}\) of its boundary and so it is a real analytic manifold of pure dimension 3. Hence by the invariance domain theorem, the restriction of \(\Psi_{(2,k,\varepsilon)}\) to the interior of \(\Delta_{(2,k,\varepsilon)}\) is a homeomorphism onto its image.

**Acknowledgments.** This is a part of the author’s Ph.D., written at the Complutense University of Madrid under the supervision of Professors J. M. Gamboa and E. Bujalance. I wish to express my gratitude to them for their encouragement and patience. I also wish to thank J. Huisman for calling my attention to the naturality of the semianalytic structure of the real moduli space, and for many helpful explanations concerning it.
References


Received April 11, 2001. This work was partially supported by DGICYT PB98-0756.

**Dep. Matemáticas Fundamentales**

**Universidad Nacional de Educación a Distancia**

**Madrid**

**Spain**

**E-mail address**: jcirre@mat.uned.es
EQUIVARIANT FRAMINGS, LENS SPACES AND CONTACT STRUCTURES

Siddhartha Gadgil

We construct a simple topological invariant of certain 3-manifolds, including quotients of $S^3$ by finite groups, based on the fact that the tangent bundle of an orientable 3-manifold is trivialisable. This invariant is strong enough to yield the classification of lens spaces of odd, prime order. We also use properties of this invariant to show that there is an oriented 3-manifold with no universally tight contact structure. We generalise and sharpen this invariant to an invariant of a finite covering of a 3-manifold.

It is well-known that the tangent bundle of an orientable 3-manifold is trivialisable. This is in particular true for manifolds of the form $M = S^3/G$, where $G$ is a finite group acting without fixed points on $S^3$. These are the so-called topological spherical space forms.

Using the fact that $TM$ is trivialisable, we can define an invariant $\mathfrak{F}(M)$ of $M$ with a fixed orientation, which we call the *equivariant framing* of $M$. Namely, the homotopy classes of trivialisations of the tangent bundle of $S^3$ are a torseur of $\mathbb{Z}$ (i.e., a set on which $\mathbb{Z}$ acts freely and transitively), which can moreover be canonically identified with $\mathbb{Z}$ by using the Lie group structure of $S^3$ as the unit quaternions and identifying a left-invariant framing with $0 \in \mathbb{Z}$. Now, find a trivialisation $\tau$ of $TM$ and pull it back to one of $TS^3$. Under the above identification, this gives an element $\mathfrak{F}(M, \tau) \in \mathbb{Z}$. This certainly depends on $\tau$, but we shall see that its reduction modulo $|G|$, when $H_1(M, \mathbb{Z}_2) = 0$ (in particular when $|G|$ is odd), and modulo $|G|/2$ otherwise, is well-defined. Observe that this is the same as the collection of homotopy classes of *equivariant* trivialisations with respect to the action of $G$ on $S^3$.

The above definition does not depend on the identification of $S^3$ with the universal cover of $M$, since two such identifications differ by an orientation preserving self-homeomorphism of $S^3$, which must be isotopic to the identity.

Notice that this definition makes essential use of the fact that we have a quotient of $S^3$, rather than a homology (or even homotopy) sphere. In the more general situation, where we have a quotient of a homology sphere by a finite cyclic group, we can use canonical 2-framings, as introduced by Atiyah.
This in fact gives an invariant corresponding to any finite cover of any manifold.

More interestingly, we can obtain an integer-valued invariant. To do this we compare the pullback of the canonical 2-framing of $M$ with that of a finite cover of $M$. We shall define this in Section 4.

The invariant $\mathcal{F}(L(p, q))$ can readily be computed for odd $p$. As it turns out, it suffices to classify lens spaces with $p$ prime.

Thus, $\mathcal{F}(\cdot)$ is an invariant sensitive enough to distinguish between homotopy equivalent manifolds. It is arguably one of the simplest such invariants.

Furthermore, there is a transparent relation between $\mathcal{F}(\cdot)$ and the exceptional isomorphism $SO(4) = (SU(2) \times SU(2))/\pm 1$. This means $\mathcal{F}(\cdot)$ is likely to be useful in studying free finite group actions on $S^3$ - the most elegant classification of orthogonal actions can be obtained using the exceptional isomorphism, so it is useful to have a related topological invariant. At the time of writing this work is in progress.

Another immediate consequence of the existence of this invariant is that corresponding to one of the orientations of the Poincaré homology sphere, we do not have a positive universally tight contact structure. Further applications in a similar vein are given in [8].

For this application, one can use tangent plane fields that are trivialisable rather than framings. The framing invariant in this context is equivalent to an invariant of tangent plane fields defined by Gompf. The relation between these is explained in Section 5. Under this equivalence, our proof translates to Gompf’s proof.

Both framings of and tangent plane fields in 3-manifolds have been studied classically. There has also been recent work on these, motivated by relations to Topological Quantum field theories, Seiberg-Witten invariants and contact geometry. Motivated by the work of Witten [17], framings have been studied by Atiyah [1], Freed and Gompf [7], Reshetikhin and Turaev [15] and Kirby and Melvin [10]. Tangent plane fields on a 3-manifold were first classified in terms of a framing by Pontrjagin [13]. More recently, intrinsic invariants of these have been studied and used by Lisca and Matic [12], G. Kuperberg [9] and finally by Gompf [9]. We clarify the relation with Gompf’s work in Section 5.

The principle novelty of this paper is the use of framings to define a useful invariant of 3-manifolds. While Gompf studies the behaviour under pullbacks of his invariants of tangent plane fields, which is equivalent for some applications, he does not define or use an invariant of 3-manifolds.

In Section 1 we show that the invariant $\mathcal{F}(M)$ can indeed be defined as above. In Section 2 it is computed for odd-order lens spaces. In Section 3 we relate this to contact structures and prove the result regarding contact structures. We generalise and sharpen the framing invariant in Section 4.
1. The definition of the framing invariant.

To define the framing invariant, we need some (straightforward) results.

**Proposition 1.1.** The set of homotopy classes of trivialisations of $TS^3$ is a torseur of $\mathbb{Z}$.

**Proof.** Given two trivialisations, expressing one in terms of the other gives a map from $S^3$ to $SO(3)$. The homotopy classes of trivialisations are the homotopy classes of such maps. But as $S^3$ is simply connected such maps lift to maps $S^3 \to SU(2) \cong S^3$, as do homotopies between them. The homotopy class of a map from $S^3$ to itself is determined by its degree (as orientations have been fixed), making the homotopy classes of trivialisations a torseur of $\mathbb{Z}$. □

**Definition 1.1.** Consider $S^3 \cong SU(2)$ as the Lie group of unit quaternions. The canonical framing of $TS^3$ is the framing of $S^3$ which is invariant under left multiplication and is $(i,j,k)$ at the identity.

**Proposition 1.2.** The homotopy class of the canonical framing is determined by an orientation of $S^3$, and does not depend on the identification with $SU(2)$.

**Proof.** Suppose $f : S^3 \to SU(2)$ is an isomorphism giving a second Lie group structure to $S^3$. Then we have an induced orientation-preserving diffeomorphism $\phi : S^3 \to S^3$, and we need to show that the pullback of the canonical trivialisation under $\phi$ is homotopic to the canonical trivialisation. But it is well-known that any orientation preserving homeomorphism from $S^3$ to itself is isotopic to the identity. Thus, the pullback fixes the homotopy class of any trivialisation, in particular the identity trivialisation. □

For the remainder of the section, let $M = S^3/G$, where $G$ is a finite group that acts without fixed points on $S^3$. Let $\pi : S^3 \to M$ be the projection map.

**Proposition 1.3.** The set of homotopy classes of trivialisations of $TM$ is a torseur of $\mathbb{Z}$ when $H_1(M, \mathbb{Z}/2\mathbb{Z}) = 0$ (in particular when $|G|$ is odd). When $\mathbb{H}_1(M, \mathbb{Z}/2\mathbb{Z}) \neq 0$, the trivialisation is determined by a map $M \to SO(3)$.

**Proof.** As with $S^3$, the difference between trivialisations is determined by a map $f : M \to SO(3)$. When $H_1(M, \mathbb{Z}/2\mathbb{Z}) = 0$, this lifts to a map $\phi : \pi_1(M) \to S^3$. It is well-known that the homotopy class of such a map is determined by its degree. □

Thus, when $H_1(M, \mathbb{Z}/2\mathbb{Z}) = 0$, the difference between two trivialisations of $M$ can be regarded as an integer. In particular this is true for $S^3$.

**Proposition 1.4.** Suppose $\tau_i, \ i = 1,2$ are trivialisations of $TM$ and $\pi^*(\tau_i)$ are their pullbacks. Then $\pi^*(\tau_1) - \pi^*(\tau_2)$ is divisible by $|G|$ when $H_1(M, \mathbb{Z}/2\mathbb{Z}) = 0$ and by $|G|/2$ when $H_1(M, \mathbb{Z}/2\mathbb{Z}) \neq 0$. 

Proof. As above, when $H_1(M, \mathbb{Z}/2\mathbb{Z}) = 0$, we have a map $\phi: M \to S^3$ representing the difference between the $\tau_i$. It is easy to see that the map $\phi \circ \pi$ represents the difference between the pullbacks. As $\pi$ has degree $|G|$ and the degree is multiplicative, it follows that $\pi^*(\tau_1) - \pi^*(\tau_2) = \deg(\phi \circ \pi)$ is divisible by $|G|$.

In the case when $H_1(M, \mathbb{Z}/2\mathbb{Z}) \neq 0$ (as also when $H_1(M, \mathbb{Z}/2\mathbb{Z}) = 0$), there is a map $\phi: M \to SO(3)$ representing the difference between the trivialisations. On composing with the covering map, this gives the a map representing $\pi^*(\tau_1) - \pi^*(\tau_2)$ which lifts to $\tilde{\phi}: S^3 \to S^3$. Thus, if $\alpha: S^3 \to SO(3)$ is the covering map, we get a commutative diagram
\[
\begin{array}{ccc}
S^3 & \xrightarrow{\tilde{\phi}} & S^3 \\
\downarrow{\pi} & & \downarrow{\alpha} \\
M & \xrightarrow{\phi} & SO(3).
\end{array}
\]

As the degree of maps is multiplicative, and $\deg(\pi) = |G|$ and $\deg(\alpha) = 2$, we get $2 \cdot \deg(\tilde{\phi}) = |G| \cdot \deg(\phi)$, or $\deg(\tilde{\phi}) = \frac{|G|}{2} \cdot \deg(\phi)$. The result follows. \qed

We are now in a position to define the invariant $\mathfrak{F}(M)$:

**Definition 1.2.** Let $M = S^3/G$, where $G$ is a finite group acting without fixed points on $S^3$. The framing invariant $\mathfrak{F}(M) \in \mathbb{Z}/\langle G \rangle \mathbb{Z}$, where $\langle G \rangle = |G|$ when $H_1(M, \mathbb{Z}/2\mathbb{Z}) = 0$ and $\langle G \rangle = |G|/2$ when $H_1(M, \mathbb{Z}/2\mathbb{Z}) \neq 0$, is the equivalence class of the trivialisation of $TS^3$ obtained by pulling back a trivialisation of $M$.

2. Computation for lens spaces.

It is easy to see that $\mathfrak{F}(\cdot)$ is a nontrivial invariant, and is in fact sensitive enough to distinguish between homotopy equivalent manifolds. For, the lens spaces $L(p, 1)$ are quotients of $S^3$ by a subgroup of the unit quaternions acting on themselves by left multiplication, and hence the canonical trivialisation is equivariant. On the other hand, the spaces $L(p, -1)$ are quotients by a subgroup of the unit quaternions acting on themselves by right multiplication, making the right-invariant trivialisation equivariant. These two trivialisations differ by the adjoint action of $SU(2)$ on its Lie algebra. This lifts to a degree 1 map from $SU(2)$ to itself. Thus the framing invariant suffices to show that there is no orientation preserving homeomorphism between $L(p, 1)$ and $L(p, -1)$ when $p \neq 2$.

But for $p \equiv 1 \pmod{4}$, with $p$ a prime, there is an orientation preserving homotopy equivalence between $L(p, 1)$ and $L(p, -1)$. By uniqueness of prime
decompositions of 3-manifolds, it follows, for example, that $L(5,1)\#L(5,1)$ is homotopy equivalent but not homeomorphic to $L(5,1)\#L(5,-1)$.

Thus, $\mathfrak{F}(\cdot)$ depends essentially on the homeomorphism type, and not just the homotopy type, of a lens space. Our goal here is to compute this explicitly for odd-order lens spaces, and show that if $p$ is an odd prime, $\mathfrak{F}(\cdot)$ suffices to classify lens spaces.

**Theorem 2.1.** Suppose $p$ is odd. Then $\mathfrak{F}(L(p,q)) = (q^{-1} - 1)^{-1}$, where $q^{-1}$ is a multiplicative inverse of $q$ modulo $p$, and $q$ and $q^{-1}$ have been chosen to be odd representatives of their mod $p$ equivalence class.

**Remark 2.2.** The formula above does not depend on the choice of the odd representatives for $q$ and $q^{-1}$.

**Proof.** It will be useful to regard $S^3$ as the join $S^1 \ast S^1$, which is embedded in the natural way in the quaternions. Then the action corresponding to the lens space $L(p,q)$ is the join of actions on the circle $S^1 \subset \mathbb{C}$ generated respectively by $z_1 \mapsto e^{2\pi i/p} z_1$ and $z_2 \mapsto e^{2\pi i q/p} z_2$. We first find an equivariant trivialisation along the circles $(z_1,0)$ and $(0,z_2)$ and then extend these to $S^3$.

Henceforth the equivariant trivialisation is given in terms of the map $f: S^3 \to SU(2) \cong S^3$ representing the difference with the left-invariant trivialisation. Along the circle $C_1 = \{(z_1,0): z_1 \in S^1\}$, we can take the first vector to point along the circle. An equivariant trivialisation is obtained by choosing the other two vectors so that they rotate $q$ times, for some choice in the mod $p$ class of $q$. This follows as the arc joining $(1,0)$ to $(e^{2\pi i/p},0)$ is a fundamental domain, with the only identification due to the group action being that of the endpoints induced by the element $(z_1, z_2) \mapsto (e^{2\pi i q/p} z_1, e^{2\pi i q/p} z_2)$, and this element rotates the normal plane to the circle by $e^{2\pi i q/p}$.

This trivialisation differs from the left-invariant trivialisation by $q - 1$ rotations, and if $q$ is chosen odd, this gives a comparison map which lifts to $S^3$. The resulting restriction of $f$ is the map $f: (z_1,0) \mapsto \left(\frac{q-1}{2},0\right)$ of degree $2\frac{q-1}{2}$ from the circle to itself.

Likewise, we can trivialise the tangent bundle along the circle $C_2 = \{(0,z_2): z_2 \in S^1\}$. It is convenient to choose the trivialisation so that it differs from the left invariant one at $(0,1)$ by $j = (0,1)$. Then the map $f$ restricted to this circle is the map $f: (0,z_2) \mapsto \left(0, \frac{q-1}{2}\right)$ of degree $\frac{q-1-1}{2}$ from the circle to itself as the identification on the boundary of the fundamental domain is induced in this case by $(z_1, z_2) \mapsto \left(e^{2\pi i q^{-1}/p} z_1, e^{2\pi i q/p} z_2\right)$.  

We now extend this map to a map from the disc \( D_0 = \{(z_1, r) \in S^3 : |z_1| \leq 1, r \in \mathbb{R}\} \subset S^3 \subset \mathbb{C} \) of the form \( re^{i\theta} \mapsto re^{i\frac{q-1}{2} \theta}, r, \theta \in \mathbb{R}\). Since the only points in the disc which are identified by the group action are on the boundary, and the trivialisation on the boundary is equivariant, we get an equivariant trivialisation of the disc corresponding to this map.

We extend this by requiring equivariance to the disc \( D_1 = \{(z_1, re^{2\pi i p}) \in S^3 : |z_1| \leq 1, r \in \mathbb{R}\}\). The map is previously defined on the boundary of this disc, where it is equivariant. Thus, we have a unique equivariant extension.

The midpoint of \( D_1 \) is part of the circle \( C_2 \), as is an arc \( \alpha \) joining the midpoints of \( D_1 \) and \( D_2 \). The map \( f \) as previously defined on \( C_2 \), must agree with the definition on \( D_1 \) at the midpoint as both these have been defined so that the trivialisation is equivariant.

The discs \( D_0 \) and \( D_1 \) bound a lens \( N \), which is a fundamental domain for the group action. The map \( f \) has already been defined on the boundary disc as well as the arc \( \alpha \). This extends to a map on \( N \), and hence a trivialisation of the tangent bundle. Up to homotopy, any other choice of \( f \) can be obtained by replacing \( f \) in a small neighbourhood of an interior point \( x \) of the fundamental domain by a degree \( k \) map from \( S^3 \) to itself for some \( k \). More precisely, \( f \) is homotopic to a map that is constant on the neighbourhood of the point \( x \). Replace this map in this neighbourhood by a map to \( S^3 \) that maps the boundary of the neighbourhood to a single point and so that the inverse image of a generic point has algebraic multiplicity \( k \). We shall call such a transformation ‘blowing a degree-\( k \) bubble’.

We now extend the trivialisation equivariantly from the fundamental domain to all of the manifold, and define \( f \) accordingly. This is an extension of the map on \( C_1 \) and \( C_2 \) that was previously defined.

The resulting map \( f : S^3 \to S^3 \) is homotopic to a map obtained from the join of maps \( C_i \to C_i \) defined by taking powers on the unit circle in \( \mathbb{C} \) by blowing a degree-\( k \) bubble in each of the \( p \) images of the fundamental domain. Thus, it has degree \( \frac{(q-1)(q^{-1}-1)}{4} + kp \). This proves our claim. \( \square \)

**Corollary 2.3.** Suppose \( p \) is a prime, then \( L(p, q) = L(p, q') \) as oriented manifolds if and only if \( q' = q^\pm 1 \).

**Proof.** It is well-known that \( L(p, q) = L(p, q^\pm 1) \) as oriented manifolds (a homeomorphism is induced by \((z_1, z_2) \mapsto (z_2, z_1))\). Conversely, for a fixed \( q \), the condition \( \mathcal{F}(L(p, q')) = \mathcal{F}(L(p, q)) \) is a quadratic equation in \( q' \) over the field \( \mathbb{Z}/p\mathbb{Z} \), with roots \( q \) and \( q^{-1} \).

If \( q^{-1} \neq q \), then these are two distinct root of the quadratic equation, and hence the only solutions for \( q' \). If \( q = q^{-1} \), and \( q' \neq q \) is another root, then we also have \( q' = (q')^{-1} \), otherwise we would have three distinct roots. But this means that \( q' = \pm 1 \) and \( q = \pm 1 \), and we have already seen that
\[ \mathfrak{f}(L(p, 1)) \neq \mathfrak{f}(L(p, -1)) \]. Alternatively, using \( q' = q^{-1} \), the above reduces to a linear equation for \( q' \) that is satisfied by \( q \), and hence \( q' = q \). \qed

The same statement is well-known to be true for all values of \( p \) and there are several proofs of this (see, for instance, [14], [3], [2] and [16]).

**Remark 2.4.** It is more natural to declare the left invariant framing to be \(-\frac{1}{2}\) rather than 0. Then we have the relation \( \mathfrak{f}(L(p, -q)) = -\mathfrak{f}(L(p, q)) \).

**Remark 2.5.** More generally, after re-normalising as above, \( \mathfrak{f}(-M) = -\mathfrak{f}(M) \). This is immediate from Section 4 and can also be proved directly.

### 3. Universally tight contact structures.

Let \( M \) be a closed, orientable 3-manifold. Recall that a contact structure \( \xi \) on \( M \) is a totally non-integrable tangent plane field. We shall assume that the tangent plane field is co-orientable (we shall say that the contact structure is co-orientable). In this situation, we can express \( \xi = \ker(\alpha) \), where \( \alpha \) is a 1-form consistent with the co-orientation.

The hypothesis of \( \xi \) being nowhere integrable is equivalent to \( \alpha \wedge d\alpha \) being a nondegenerate 3-form. Thus, this is everywhere a nonzero multiple of the volume form, and hence induces an orientation on \( M \). We say that \( \xi \) is **positive** if this orientation agrees with the orientation of \( M \).

A fundamental dichotomy among contact structures on 3-manifolds is between **tight** and **overtwisted** contact structures. An **overtwisted** contact structure is one that contains an unknot that is everywhere tangent to the contact structure so that the framing induced by the contact structure is the 0-framing. A contact structure that is not overtwisted is said to be **tight**. A **universally tight** contact structure is one that pulls back to a tight contact structure on every cover of \( M \).

A fundamental result of Eliashberg is that \( S^3 \) with a fixed orientation has a unique positive tight contact structure, namely the contact structure invariant under left multiplication. Our results follow from this and some simple observations.

**Proposition 3.1.** Let \( M \) be an integral homology 3-sphere with a contact structure \( \xi \). Then there is a canonical framing associated to \( \xi \). Further, the pullback of this framing to any homology sphere that covers \( M \) is the framing induced by the pullback of the contact structure.

**Proof.** As \( M \) is a homology sphere, the contact-structure is co-orientable. Choose and fix a co-orientation. This induces an orientation on the plane-bundle given by the contact structure, which we identify with \( \xi \).

As \( H^2(M) = 0 \), the Euler class of \( \xi \) is trivial. Hence there is a trivialisation of \( \xi \). Further, two trivialisations differ by a map onto \( S^1 \). As \( H^1(M) = 0 \), any such map is homotopic to a constant map.
Thus, there is a trivialisation \((X_1, X_2)\) of \(\xi\), canonical up to homotopy. This, together with a vector \(X_3\) normal to \(\xi\), that is consistent with the co-orientation, gives a framing \((X_1, X_2, X_3)\).

The homotopy class of this trivialisation does not depend on the choice of co-orientation since \((X_1, -X_2, -X_3)\) gives a trivialisation corresponding to the opposite co-orientation, and this is clearly homotopic to the trivialisation \((X_1, X_2, X_3)\).

As the trivialisation of \(\xi\) pulls back to give a trivialisation of the pullback to any cover of \(\xi\), the second claim follows. □

Now let \(\mathcal{P}\) be the Poincare homology sphere with a fixed orientation, and let \(-\mathcal{P}\) denote the same manifold with the opposite orientation. These manifolds have finite fundamental group. The Poincaré homology sphere is the quotient of \(S^3\) by a group acting by left multiplication, and \(-\mathcal{P}\) is the quotient of an action by right multiplication. We can now prove the following theorem. Note that any contact structure on a homology sphere is automatically co-orientable.

**Theorem 3.2** (Gompf). The manifold \(-\mathcal{P}\) does not have a universally tight positive contact structure.

**Proof.** As \(-\mathcal{P}\) is the quotient of \(S^3\) by a group acting by right multiplication, it follows that any framing on \(-\mathcal{P}\) pulls back to one homotopic to a framing invariant under right Lie multiplication, or one differing from this by \(|\pi_1(\mathcal{P})|\) units (as \(H_1(\mathcal{P}, \mathbb{Z}_2) = 0\)). However, if \(-\mathcal{P}\) had a universally tight positive contact structure, then the associated framing must pulls back to give the framing associated to left Lie multiplication. This gives the required contradiction. □

**Remark 3.3.** Etnyre and Honda [6] have shown that \(-\mathcal{P}\) does not have a tight contact structure.

**Remark 3.4.** We see in Section 5 that the proof of the above result translates to Gompf’s proof under the correspondence between framings and trivialisable tangent plane fields.

V. Colin [4] shows that tight contact structures on connected sums of manifolds are connected sums of tight contact structures on each summand. Using this, we obtain the following result:

**Corollary 3.5.** The manifold \(\mathcal{P}# -\mathcal{P}\) does not admit a universally tight contact structure.

4. 2-Framings and invariants of covers.

We now generalise and sharpen the framing invariant using so called canonical 2-framings as introduced by Atiyah [1]. Atiyah has shown that any
3-manifold has associated to it a canonical framing $\mathcal{F}$ of the Whitney sum $2TM = TM \oplus TM$ of the tangent bundle with itself, considered as a Spin(6)-bundle with the natural spin structure. The framing $\mathcal{F}$ is characterised by

$$\sigma(W^4) = \frac{1}{6}p_1(2TW,\mathcal{F})$$

for any smooth 4-manifold $W$ with $\partial W = M$. Here $\sigma$ denotes the signature and $p_1$ the relative Pontrjagin class. By the Hirzebruch signature formula this does not depend on the choice of $W$.

Atiyah has shown that such a 2-framing always exists, and the 2-framings form a torseur of $\mathbb{Z}$. Now, we can define the framing invariant $\mathfrak{F}(M, N)$ associated to a cover $M \to N$ - pull back the canonical 2-framing of $N$ and compare this with the canonical 2-framing of $M$. Thus, we get an integer-valued invariant. We state for reference the following lemma, which is an immediate consequence of Atiyah’s result:

**Lemma 4.1.** Suppose $M$ is a 3-manifold bounding a 4-manifold $W$ and let $\mathcal{F}$ be any 2-framing of $M$. The difference between $\mathcal{F}$ and the canonical framing is $\sigma(W^4) - \frac{1}{6}p_1(2TW,\mathcal{F})$.

As a framing gives a 2-framing, we see that we have a sharpening of the framing invariant defined earlier. We show here that this is a nontrivial invariant.

**Remark 4.2.** A hyperbolic 3-manifold has many covers, hence many invariants associated to it. It is not clear whether these are useful.

**Theorem 4.3.** Suppose $N$ and $N'$ are $h$-cobordant 3-manifolds, $M$ is a cover of $N$ and $M'$ the corresponding cover of $N'$. Then $\mathfrak{F}(M, N) = \mathfrak{F}(M', N')$.

**Proof.** Let $X$ be the $h$-cobordism between $N$ and $N'$, so that $\partial W = N' - N$. If $W$ is a 4-manifold with boundary $N$, then $W' = W \coprod_N X$ is a 4-manifold with boundary $N'$ with the same signature as $W$. By considering these manifolds, it is immediate that that if $\mathfrak{F}$ is the canonical 2-framing for $M$, then the canonical framing $\mathfrak{F}'$ of $N'$ is characterised by

$$p_1(2TX, \mathfrak{F}, \mathfrak{F}') = 0.$$

The $h$-cobordism $X$ lifts to an $h$-cobordism $Y$ between $M$ and $M'$, and the Pontrjagin class relative to the framings pulled back is the pullback of the Pontrjagin class and hence is zero. Let $U$ be a 4-manifold with boundary $M$, and let $U' = U \coprod_M Y$. Applying Lemma 4.1 to $U$ and to $U'$, the result follows. $\square$

We can generalise the above theorem to the following:
Theorem 4.4. Let $X$ be a cobordism between $N_1$ and $N_2$ and let $\phi: \pi_1(W) \to H$ be a surjective map onto a finite group that restricts to surjections on $\pi_1(N_1)$ and $\pi_1(N_2)$. Suppose that the cover $\tilde{X}$ with fundamental group $\ker(\phi)$ satisfies

$$\sigma(\tilde{W}) = |H| \sigma(W)$$

and $M_1$ and $M_2$ are the covers of $N_i$ with fundamental group $\ker(\phi)$. Then $\mathfrak{F}(M,N) = \mathfrak{F}(M',N')$.

Proof. We use the additivity of the signature and relative Pontrjagin classes. The above proof generalises immediately.

Example 4.1. Let $N_1$ be a lens space, $M_1$ be $S^3$ and $K$ be a homologically trivial knot in $N_1$ the components of whose lift to $S^3$ are unlinked (for instance, the untwisted Whitehead double of any homologically trivial knot). Add a 2-handle to $N_1$ along $K$ with framing 1 to get a 4-manifold $W$, and let its other boundary component be $N_2$. Let $\phi$ be the map onto $\pi_1(N_1)$ that extends the identity map on $N_1$. Then $\mathfrak{F}(M,N) = \mathfrak{F}(M',N')$.

The manifold $N_2$ is the result of surgery about the knot $K$. Using this construction and taking connected sums, we get hyperbolic manifolds with covers having various framings.

5. Relation to Gompf’s invariants.

In this section we relate equivariant framings to Gompf’s invariant, equivalent to the 3-dimensional obstruction in Pontrjagin’s classification, for tangent plane fields. Assume henceforth that $M$ is an oriented rational homology sphere (for instance, $M = S^3/G$). We first establish a canonical correspondence between homotopy classes of framings $\mathcal{F}$ of $M$ and homotopy classes of orientable tangent plane fields $\xi$ on $M$ with $c_1(\xi) = 0$.

Proposition 5.1. Let $M$ be an rational homology 3-sphere. Then there is a natural bijective correspondence between homotopy classes of framings $\mathcal{F}$ of $M$ and homotopy classes of orientable tangent plane fields $\xi$ on $M$ with $c_1(\xi) = 0$.

Proof. We proceed as in Proposition 3.1. Given an orientable tangent plane fields $\xi$ on $M$ with $c_1(\xi) = 0$, fix an orientation on $\xi$. There is a trivialisation $(X_1, X_2)$ of $\xi$ respecting the orientation. Further, the homotopy classes of such trivialisations are classified by $H^1(M) = 0$. Hence, as $M$ is a rational homology sphere, the trivialisation is canonical up to homotopy.

Find a vector field $X_3$ normal to $\xi$ such that $(X_1, X_2, X_3)$ respects the orientation on $M$. This is possible as $\xi$ and $M$ are orientable. Then $\mathcal{F} = (X_1, X_2, X_3)$ is a framing of $M$. As before this does not depend on the choice of orientation.
Conversely, given a framing $\mathcal{F} = (X_1, X_2, X_3)$ let $\xi$ be the span of $X_1$ and $X_2$. Evidently these constructions are inverses of each other.

We now recall the invariant of Gompf for such tangent plane fields. To do this one finds an almost complex 4-manifold $(X, J)$ with $\partial X = M$ so that the plane field $TM \cap J(TM)$ on $M$ induced by the almost complex structure is $\xi$. Using this, one would wish to define the invariant $c_1^2(X) - 2\chi(X) - 3\sigma(X)$, where $\chi$ and $\sigma$ denote the Euler characteristic and the signature.

One cannot always define such an invariant as $c_1(X) \in H^2(X) \cong H_2(X, \partial X)$ and there is no natural pairing on $H_2(X, \partial X)$. Gompf works instead with a pairing on surfaces representing elements of $H_2(X, \partial X)$ with given framings on their boundaries.

In case of tangent plane fields $\xi$ on $M$ with $c_1(\xi) = 0$, we can define the invariant directly by the following lemma:

**Lemma 5.2.** Let $X$ be an almost complex 4-manifold $X$ with $\partial X = M$ so that the plane field $\xi$ on $M$ induced by the almost complex structure satisfies $c_1(\xi) = 0$. Then the Poincaré dual $PD(c_1(X)) \in H_2(X, \partial X)$ of $c_1(X)$ is contained in the image of $H_2(X)$ under the inclusion map.

**Proof.** If $[F] = PD(c_1(X))$ for a properly embedded surface $F$ with boundary, then it is easy to see that $[\partial F] = PD(c_1(\xi))$ as $TX|_{\partial X}$ splits as the sum of a trivial complex line bundle and $\xi$. It follows that $[\partial F] = 0$. Hence $PD(c_1(X))$ is in the image of $H_2(X)$ by the long exact sequence of homology groups.

Thus, Gompf’s invariant reduces to $c_1^2(X) - 2\chi(X) - 3\sigma(X)$. Using $p_1(X, \mathcal{F}) = c_1^2(X) + 2c_2(X)$ and $c_2(X) = \chi(X)$, we see that this is $-3(\sigma(X) - \frac{1}{6}p_1(2TX, \mathcal{F}))$. We saw that $\sigma(X) - \frac{1}{6}p_1(2TX, \mathcal{F}) = 0$ characterises the canonical framing, which is used to give an integral sharpening of the framing invariant.

To prove Theorem 3.2, Gompf uses the fact that if we pullback two tangent plane fields from $-\mathfrak{F}$ to $S^3$, then the difference between the values of the invariant for the two plane fields is a multiple of 4 times the degree of the cover (the factor of 4 comes about in the process of defining the invariant for general plane fields). This is equivalent to our argument using equivariant framings.

**Acknowledgements.** I would like to Slava Matveyev and Kalyan Mukherjea for several helpful conversations and comments. I would also like to thank the referee for several helpful comments and corrections.
References


Received May 4, 2001 and revised November 28, 2001.

Stat-Math Unit
Indian Statistical Institute
8th Mile, Mysore Road
Bangalore 560059, India
E-mail address: gadgil@ns.isibang.ac.in
CARLESON’S CONVERGENCE THEOREM FOR DIRICHLET SERIES

HÅKAN HEDENMALM AND EERO SAKSMAN

A Hilbert space of Dirichlet series is obtained by considering the Dirichlet series \( f(s) = \sum_{n=1}^{\infty} a_n n^{-s} \) that satisfy \( \sum_{n=0}^{\infty} |a_n|^2 < +\infty \). These series converge in the half plane \( \text{Re } s > \frac{1}{2} \) and define a functions that are locally \( L^2 \) on the boundary \( \text{Res} = \frac{1}{2} \). An analog of Carleson’s celebrated convergence theorem is obtained: Each such Dirichlet series converges almost everywhere on the critical line \( \text{Res} = \frac{1}{2} \). To each Dirichlet series of the above type corresponds a “trigonometric” series \( \sum_{n=1}^{\infty} a_n \chi(n) \), where \( \chi \) is a multiplicative character from the positive integers to the unit circle. The space of characters is naturally identified with the infinite-dimensional torus \( T^\infty \), where each dimension comes from a prime number. The second analog of Carleson’s theorem reads: The above “trigonometric” series converges for almost all characters \( \chi \).

1. Introduction.

The study of Dirichlet series of the form \( \sum_{n=1}^{\infty} a_n n^{-s} \) has a long history beginning in the nineteenth century, and the interest was due mainly to the central role that such series play in analytic number theory. The general theory of Dirichlet series was developed by Hadamard, Landau, Hardy, Riesz, Schnee, and Bohr, to name a few. As regards the modern development, we mention the work of Helson and Kahane. Helson \([9, 10]\) should probably be credited for pioneering modern harmonic analysis methods in the theory of Dirichlet series. As a sample of Kahane’s work, we mention the papers \([15, 14]\). Recently, in \([7]\), Hedenmalm, Lindqvist, and Seip considered a natural Hilbert space \( \mathcal{H} \) of Dirichlet series and began a systematic study thereof. The elements of \( \mathcal{H} \) are analytic functions on the half-plane \( \text{Re } s > \frac{1}{2} \) of the form

\[
(1.1) \quad f(s) = \sum_{n=1}^{\infty} a_n n^{-s}
\]
where the coefficients $a_1, a_2, a_3, \ldots$ are complex numbers subject to the norm boundedness condition

$$\|f\|_H = \left(\sum_{n=1}^{\infty} |a_n|^2\right)^{\frac{1}{2}} < +\infty.$$ 

In [7], the pointwise multipliers of $H$ were characterized, and the result was applied to a problem of Beurling concerning 2-periodic dilation bases in $L^2([0,1])$. The reader is referred to [8] for some historical comments on the topic. In [5], Gordon and Hedenmalm followed up by characterizing the bounded composition operators of $H$.

The convergence and analyticity of $f \in H$ given by the series (1.1) in the half-plane $\text{Re} s > \frac{1}{2}$ is a simple consequence of the Cauchy-Schwarz inequality. A deeper fact is that the boundary values of $f$ on the ‘critical’ line $\text{Re} s = \frac{1}{2}$ are locally $L^2$-functions (see [17, formula (29), p. 140] or [7, Theorem 4.11]).

Here, we establish for Dirichlet series the counterpart of the celebrated Carleson convergence theorem [1] for square summable Fourier series ($\mathbb{R}$ is the set of all real numbers):

**Theorem 1.1.** Let $\sum_{n=1}^{\infty} |a_n|^2 < +\infty$. Then the series

$$\sum_{n=1}^{\infty} a_n n^{-\frac{1}{2} + it}$$

converges for almost every $t \in \mathbb{R}$.

Let us digress on Carleson’s convergence theorem in the context of a square summable Taylor series

$$f(z) = \sum_{n=0}^{\infty} \hat{f}(n) z^n, \quad z \in \mathbb{D},$$

where $\mathbb{D}$ is the open unit disk and the coefficients satisfy

$$\|f\|_{H^2} = \left(\sum_{n=0}^{\infty} |\hat{f}(n)|^2\right)^{\frac{1}{2}} < +\infty.$$ 

The Hilbert space of such functions $f$ is denoted by $H^2(\mathbb{D})$, and in a natural fashion, it is a closed subspace of $L^2(\mathbb{T})$, the space of (equivalence classes of) Lebesgue square summable functions $g$ on the unit circle $\mathbb{T}$, supplied with the norm

$$\|g\|_{L^2} = \left(\int_{\mathbb{T}} |g(z)|^2 \, d\sigma(z)\right)^{\frac{1}{2}},$$

where $d\sigma$ is normalized arc length measure on $\mathbb{T}$: $d\sigma(e^{i\theta}) = d\theta/(2\pi)$. By Carleson’s theorem, we have convergence in (1.2) for almost all $z \in \mathbb{T}$. And
since a general Fourier series decomposes into an analytic and an antianalytic component, the statement that Taylor series for \( H^2(\mathbb{D}) \) functions converge almost everywhere on \( \mathbb{T} \) implies the almost everywhere convergence on \( \mathbb{T} \) for Fourier series of \( L^2(\mathbb{T}) \) functions. We shall need the Taylor series maximal function associated with \( f \in H^2(\mathbb{D}) \),

\[
Mf(z) = \sup_j \left| \sum_{n=j}^{\infty} \hat{f}(n) z^n \right|, \quad z \in \mathbb{T},
\]

where \( j \) runs over \( \{0, 1, 2, \ldots\} \), and the infinite sum is interpreted in the sense of the identity

\[
\sum_{n=j}^{\infty} \hat{f}(n) z^n = f(z) - \sum_{n=1}^{j-1} \hat{f}(n) z^n.
\]

The Taylor series maximal function operator \( M \) is a nonlinear operator from \( H^2(\mathbb{D}) \) to the Lebesgue measurable functions on \( \mathbb{T} \) with values in \([0, +\infty]\). Carleson’s convergence theorem is a consequence of the following estimate of \( M \), due to Hunt [11]. The proof is essentially a modification of Carleson’s original argument, which, as Hunt explains in [12], is equivalent to a weak type estimate for \( M \). It should be mentioned that whereas Carleson’s proof is based on a careful analysis of individual square summable functions, the later proof of Fefferman [4] concentrates on analyzing the linearized maximal operator.

**Theorem 1.2** (Carleson-Hunt [1, 11]). There exists an absolute constant \( A = A_{CH} \) such that for every \( f \in H^2(\mathbb{D}) \),

\[
\int_{\mathbb{T}} |Mf(z)|^2 \, d\sigma(z) \leq A \int_{\mathbb{T}} |f(z)|^2 \, d\sigma(z).
\]

A natural question is whether the absolute constant \( A_{CH} \) is astronomical. We have been assured that it is not: It can be chosen at the order of magnitude of 100. Note that in particular, \( M \) maps \( H^2(\mathbb{D}) \) into \( L^2(\mathbb{T}) \). To obtain the convergence theorem from the above statement, we approximate in the \( H^2(\mathbb{D}) \)-norm the function \( f \in H^2(\mathbb{D}) \) by another function \( g \in H^2(\mathbb{D}) \) which is \( C^\infty \)-smooth up to the boundary. The function \( g \) has a nicely convergent Taylor series on \( \mathbb{T} \). The maximal function estimate is then applied to the difference \( f - g \), and it shows that the partial sums of the Taylor series for \( f \) are uniformly close in the index parameter to those of \( g \). Writing this down carefully, we see that the desired convergence assertion follows.

Let \( K(x, \theta) \) be the kernel function

\[
K(x, \theta) = \frac{e^{-ix\theta}}{\|x\|}, \quad x \in \mathbb{R} \setminus \{0\}, \quad \theta \in \mathbb{R},
\]
which for fixed $\theta$ is interpreted as a distribution on $\mathbb{R}$ in the principal value sense.

The following dual reformulation of Theorem 1.2 appears to be essentially due to Vinogradov [22].

**Theorem 1.3 (Strong Hilbert inequality).** Suppose $\varphi$ is a compactly supported $C^\infty$-smooth function on the set $]-\pi, \pi[ \times ]0, +\infty[$. Then the following estimate holds:

$$
\int_0^{+\infty} \int_0^{+\infty} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi}
K(x - y, \max(\theta_1, \theta_2)) \varphi(x, \theta_1) \overline{\varphi(y, \theta_2)}
dxdy d\theta_1 d\theta_2
\leq B \int_{-\pi}^{+\pi} \left( \int_0^{+\infty} |\varphi(x, \theta)| d\theta \right)^2
dx,
$$

where $B$ is an absolute constant related to the constant $A_{CH}$ of the Carleson-Hunt theorem (Theorem 1.2).

In the above statement, the integration against the kernel in the $(x,y)$ coordinates is to be interpreted in the weaker sense of singular integral operator theory. For instance, we can treat the integration in $y$ as a convolution of a distribution and a smooth function, so that the rest of the integrations are well-defined in the Lebesgue sense. In this guise, we may view the Carleson-Hunt’s theorem as a far-reaching generalization of Hilbert’s inequality [23], which results by letting the functions $\varphi(x, \theta)$ tend to $\phi(x) \delta_0(\theta)$, where $\phi$ is a smooth function and $\delta_0$ is the unit point mass at 0. For the sake of keeping the presentation as self-contained as reasonably possible, the short and simple proof of Theorem 1.3 is reproduced later on in the paper.

We return to the main topic, Dirichlet series. The point with reformulating the Carleson-Hunt theorem as Theorem 1.3 is the following. In the same fashion as the Carleson-Hunt theorem has a dual formulation, the analogous maximal function statement for Dirichlet series in the spirit of Theorem 1.1 has a similarly dual formulation with a different kernel $K^D(x, \theta)$. Surprisingly, the kernels $K$ and $K^D$ prove to be so similar that the inequality of Theorem 1.3 for $K$ immediately entails the corresponding inequality with $K^D$ (and vice versa).

We have yet another convergence result for Dirichlet series, but this time it concerns the typical convergence behavior of a function $f \in \mathcal{H}$. Given a function $f$ of the form (1.1), we form the functions

$$
(1.3) \quad f_\chi(s) = \sum_{n=1}^{\infty} a_n \chi(n) n^{-s},
$$

where $\chi(n)$ is a character, which means that $\chi(1) = 1, \chi(n) \in \mathbb{T}$ for all $n$, and $\chi(mn) = \chi(m) \chi(n)$ for all $m$ and $n$. The functions $f_\chi$ are known as the vertical limit functions for $f$. The terminology is explained by the fact that
\( f_{\chi}(s) \) is obtained from \( f \) as a limit of some of the vertical translates \( f(s - it) \), with \( t \in \mathbb{R} \). Each character is determined uniquely by its values on the set of primes \( \mathcal{P} = \{2, 3, 5, 7, 11, \ldots \} \), and the values at different primes may be chosen independently of each other. The set of all characters is denoted by \( \Xi \), and we realize that it can be equated with the infinite-dimensional polycircle \( \mathbb{T}^\infty \) by identifying each dimension with a prime number (see [7] for details). The polycircle \( \mathbb{T}^\infty \) has a natural product probability measure defined on it, denoted \( d\varpi \), the product of the normalized arc length measure \( d\sigma \) in each dimension. The set of characters \( \Xi \) constitutes the dual group of the multiplicative group of positive rationals \( \mathbb{Q}_+ \), if the latter is given the discrete topology. The Haar probability measure on the compact group \( \Xi \) coincides with \( d\varpi \). A natural question arises: Given \( f \in \mathcal{H} \), what is the almost sure convergence behavior of the series (1.3) for \( f_{\chi}(s) \), where \( s \) is a point in the complex plane, and \( \chi \) is a character? The words “almost sure” refer to the Haar probability measure \( d\varpi \) on \( \Xi \). It is mentioned in [7] that for almost all \( \chi \), \( f_{\chi}(s) \) extends to a holomorphic function on the right half plane \( \text{Re} s > 0 \), and that this is best possible. In fact, in [10] (see also [7], Theorem 4.4), Helson shows that for almost all \( \chi \), the Dirichlet series (1.3) actually converges in the half-plane \( \text{Re} s > 0 \). By Theorem 4.1 of [7], the function \( f_{\chi}(it) \) makes sense as a locally \( L^2 \) summable function on the real line, for almost all \( \chi \). This makes us suspect that we have convergence in (1.3) for almost all \( s \) on the line \( \text{Re} s = 0 \) and almost all \( \chi \). That is confirmed by the following theorem.

**Theorem 1.4.** Let \( f \in \mathcal{H} \) be of the form (1.1), and let \( f_{\chi} \in \mathcal{H} \) be defined by (1.3). Then the series

\[
\sum_{n=1}^{\infty} a_n \chi(n)n^{-it}
\]

converges for almost all characters \( \chi \) and almost all reals \( t \).

An equivalent formulation of this result reads as follows.

**Theorem 1.5.** Let \( f \in \mathcal{H} \) be of the form (1.1). Then the series

\[
\tilde{f}(\chi) = \sum_{n=1}^{\infty} a_n \chi(n), \quad \chi \in \Xi,
\]

converges almost everywhere.

Theorem 1.5 states that square summable infinite-dimensional Taylor series converge almost everywhere on the polycircle \( \mathbb{T}^\infty \) with respect to a certain order of summation (the space of such Taylor series is known as the Hardy space \( H^2(\mathbb{D}^\infty) \)). Here is how that works. Let \( p_j \) be the \( j \)-th prime,
and write \( z_j = \chi(p_j) \); then the infinite-dimensional power series

\[
F(z_1, z_2, \ldots) = \sum_{n=1}^{\infty} a_n z_{k_1}^{\nu_1} z_{k_2}^{\nu_2} \cdots z_{k_r}^{\nu_r},
\]

converges for almost all \((z_1, z_2, z_3, \ldots) \in T^\infty\), where \( k_j, \nu_j \), and \( n \) are related via the prime number factorization of \( n \): \( n = p_{k_1}^{\nu_1} p_{k_2}^{\nu_2} \cdots p_{k_r}^{\nu_r} \). The order of summation, of course, is dictated by the index \( n \): The condition of summing up to index \( N \) is expressed by

\[
\log n = \nu_1 \log p_{k_1} + \nu_2 \log p_{k_2} + \cdots + \nu_r \log p_{k_r} \leq \log N,
\]

so that the summation is cut by a single hyperplane in the index plane. This permits us to apply a technique devised by Fefferman \[3\] for finitely many cuts in a finite-dimensional setting to the infinite-dimensional case, and obtain the \( L^2 \) maximal function estimate here as well. Fefferman’s idea is to reduce the situation to the one-dimensional Fourier series case, where the Carleson-Hunt theorem (Theorem 1.2) applies.

Theorem 1.5 yields as a consequence nontrivial estimates for the almost sure growth behavior of partial sums of random characters, a question which was considered in \[7\]. It follows that almost surely,

\[
\sum_{n=1}^{N} \chi(n) = O \left( \sqrt{N} \log N \left( \log \log N \right)^{1/2+\varepsilon} \right), \quad \text{as } N \to +\infty.
\]

Finding the best possible growth bound for the almost sure behavior of these partial sums has an unmistakable Erdős-type flavor, in its combination of probability and number theory. And sure enough, in \[2\, pp. 251-252\], Erdős states as a problem to determine the almost sure growth of the analogous sums, where the \( \chi(p) \) for prime indices \( p \) are replaced by independent random variables assuming the values ±1 with equal probabilities \( \frac{1}{2} \). Erdős looks to compare the growth of the partial sums with the classical law of the iterated logarithm, where all the terms \( \chi(n) \) are independent and take values ±1 with equal probabilities \( \frac{1}{2} \). In Erdős’ problem, as in ours, the characters have the multiplicative property \( \chi(mn) = \chi(m) \chi(n) \), which reduces the randomness and introduces a number-theoretic ingredient. A complete solution should thus shed light on the multiplicative structure of the integers.

2. Carleson-Hunt’s theorem and duality.

The Hardy space \( H^2(\mathbb{D}) \) was introduced earlier in terms of Taylor coefficients. Here we mention that a function \( f \) holomorphic in \( \mathbb{D} \) is in \( H^2(\mathbb{D}) \) if and only if

\[
\|f\|_{H^2} = \sup_{0<r<1} \left( \int_{T} |f(r \zeta)|^2 \, d\sigma(\zeta) \right)^{\frac{1}{2}} < +\infty,
\]
and that the values of \( f \) are well-defined almost everywhere on the boundary \( \mathbb{T} \). The norm of \( f \) in \( H^2(\mathbb{D}) \) then equals the \( L^2(\mathbb{T}) \) norm of the boundary function. Moreover, a function in \( L^2(\mathbb{T}) \) is in \( H^2(\mathbb{D}) \) if and only if its harmonic extension (via the Poisson integral) to the interior is holomorphic.

To simplify the notation, we identify the unit circle \( \mathbb{T} \) with the interval \( [\pi, \pi] \), where topologically the endpoints are tied together. Our point of departure is the fundamental Carleson-Hunt estimate of the maximal function, as stated in Theorem 1.2. We intend to find a dual reformulation of this result. To that end, take an \( f \in H^2(\mathbb{D}) \), with Taylor series expansion

\[
f(z) = \sum_{n=0}^{\infty} \hat{f}(n) z^n, \quad z \in \mathbb{D},
\]

and observe that by the \( l^1 - l^\infty \) duality, we may write, for a given fixed \( x \in ]-\pi, \pi] \),

\[
Mf(x) = \sup_{\phi_j} \text{Re} \sum_{j=0}^{\infty} \sum_{n=j}^{\infty} \hat{f}(n) e^{inx} \overline{\phi_j} = \sup_{\phi_j} \sum_{n=0}^{\infty} \hat{f}(n) e^{inx} \sum_{j=0}^{n} \overline{\phi_j},
\]

where the supremum runs over all complex-valued sequences \( \{\phi_j\}_{j=0}^{\infty} \) with

\[
\sum_{j=0}^{\infty} |\phi_j| \leq 1.
\]

We now let \( x \) vary and allow the \( \phi_j \) to depend on \( x \). Keeping in mind that the function \( f \) is fixed, it is easy to check that given an \( \varepsilon > 0 \), we may find Borel measurable functions \( \phi_j \in L^\infty(] - \pi, \pi]) \) with the analog of (2.1),

\[
\sum_{j=0}^{\infty} |\phi_j(x)| \leq 1, \quad x \in ] - \pi, \pi],
\]

such that

\[
\text{Re} \sum_{n=0}^{\infty} \hat{f}(n) e^{inx} \sum_{j=0}^{n} \overline{\phi_j}(x) = \text{Re} \sum_{j=0}^{\infty} \overline{\phi_j}(x) \sum_{n=j}^{\infty} \hat{f}(n) e^{inx} \geq (1 - \varepsilon) Mf(x)
\]

almost everywhere on \( ] - \pi, \pi] \). It follows that we may calculate the \( L^2(] - \pi, \pi]) \)-norm of the maximal function in the following fashion:

\[
\|Mf\|_{L^2} = \sup_{g} \text{Re} \frac{1}{2\pi} \int_{-\pi}^{\pi} Mf(x) \overline{g(x)} \, dx
\]

\[
= \sup_{g, \phi_j} \text{Re} \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{n=0}^{\infty} \hat{f}(n) e^{inx} \sum_{j=0}^{n} \overline{\phi_j}(x) \overline{\phi_j}(x) \overline{g(x)} \, dx
\]

\[
= \sup_{g, \phi_j} \sum_{n=0}^{\infty} \hat{f}(n) \sum_{j=0}^{n} \phi_j g(n),
\]
where \( g \) ranges over all functions in \( L^2(\mathbb{R} - \pi, \pi) \) with norm \( \leq 1 \), and \( \phi_j \) ranges over all sequences of \( L^\infty(\mathbb{R} - \pi, \pi) \) functions with (2.2). To obtain the "norm" \( \|M\| \) of the nonlinear operator \( M \), which by the Carleson-Hunt theorem is bounded by an absolute constant \( \sqrt{A_{\text{CH}}} \), we form the supremum of \( \|Mf\|_{L^2} \) over all functions \( f \in L^2(\mathbb{D}) \) of the norm \( \leq 1 \). Taking the supremum over all such \( f \) of the last expression in (2.3),

\[
\text{Re} \sum_{n=0}^\infty \hat{f}(n) \sum_{j=0}^n \hat{\phi}_j g(n),
\]

we obtain

\[
\left( \sum_{n=0}^\infty \left| \sum_{j=0}^n \hat{\phi}_j g(n) \right|^2 \right)^{1/2},
\]

and hence the assertion of the Carleson-Hunt theorem reads

(2.4)

\[
\sup_{g, \phi_j} \sum_{n=0}^\infty \left| \sum_{j=0}^n \hat{\phi}_j g(n) \right|^2 \leq A_{\text{CH}},
\]

where the supremum is, as before, taken over all \( g \in L^2(\mathbb{R} - \pi, \pi) \) of unit norm and all sequences of bounded measurable functions \( \phi_j \) with (2.2). Now consider functions \( \psi_j \in L^2(\mathbb{R} - \pi, \pi) \), such that

(2.5)

\[
\int_{-\pi}^\pi \left( \sum_{j=0}^\infty \left| \psi_j(x) \right| \right)^2 \frac{dx}{2\pi} \leq 1.
\]

Put \( g = \sum_j |\psi_j| \in L^2(\mathbb{R} - \pi, \pi) \), which then has norm at most 1 in \( L^2(\mathbb{R} - \pi, \pi) \), and observe that the functions \( \phi_j \) defined by the relations \( \phi_j(x) = \psi_j(x)/g(x) \) if \( g(x) > 0 \), and \( \phi_j(x) = 0 \) otherwise, meet the requirement (2.1). It follows from (2.4) that we have

(2.6)

\[
\sup_{\psi_j} \sum_{n=0}^\infty \left| \sum_{j=0}^n \hat{\psi}_j(n) \right|^2 \leq A_{\text{CH}},
\]

the supremum being taken over all sequences \( \psi_j \) with (2.5). Clearly, (2.6) generalizes (2.4). By dropping the scaling restriction (2.5), we end up with the following reformulation of Theorem 1.2.

Theorem 2.1. For all sequences \( \psi_j \) of functions in \( L^2(\mathbb{R} - \pi, \pi) \), we have the inequality

\[
\sum_{n=0}^\infty \left| \sum_{j=0}^n \hat{\psi}_j(n) \right|^2 \leq A \int_{-\pi}^\pi \left( \sum_{j=0}^\infty \left| \psi_j(x) \right| \right)^2 \frac{dx}{2\pi},
\]
where $A = A_{\text{CH}}$ is the constant of the Carleson-Hunt theorem.

In the above theorem, the assertion is trivial if the right-hand side assumes the value $+\infty$. It is clear that we have obtained an honest reformulation of the Carleson-Hunt theorem, because we may run the argument backwards and obtain Theorem 1.2 out of Theorem 2.1. The only thing we were sloppy about was that we did not justify that we could change the order of summation at a certain point of the argument. However, by first restricting our attention to, say polynomials $f$, the general case follows by approximation.

We carry on to write the above theorem in the form mentioned in the introduction (Theorem 1.3). Recall that in the statement, $\varphi$ is a $C^\infty$-smooth compactly supported function on $[-\pi, \pi] \times [0, +\infty]$. For $j = 0, 1, 2, \ldots$, we let the functions $\psi_j$ be defined by

$$
\psi_j(x) = \int_j^{j+1} \varphi(x, \theta) \, d\theta, \quad x \in [-\pi, \pi],
$$

so that by Theorem 2.1,

(2.7) $$
\int [0, +\infty] \left| \int_{-\pi}^{\pi} e^{-itx} \int_0^{t+1} \varphi(x, \theta) \, d\theta \, dx \right|^2 \frac{2\pi}{d\#(t)} \leq A_{\text{CH}} \int_{-\pi}^{\pi} \left( \sum_{j=0}^{\infty} \int_j^{j+1} \varphi(x, \theta) \, d\theta \right)^2 \frac{2\pi}{d\#(t)}
$$

where $d\#$ stands for the counting measure on the nonnegative integers $\{0, 1, 2, \ldots \}$. By Minkowski’s inequality,

$$
\sum_{j=0}^{\infty} \left| \int_j^{j+1} \varphi(x, \theta) \, d\theta \right| \leq \int_0^{+\infty} |\varphi(x, \theta)| \, d\theta,
$$

so that (2.7) implies that

(2.8) $$
\int [0, +\infty] \left| \int_{-\pi}^{\pi} e^{-itx} \int_0^{t+1} \varphi(x, \theta) \, d\theta \, dx \right|^2 \frac{2\pi}{d\#(t)} \leq A_{\text{CH}} \int_{-\pi}^{\pi} \left( \int_0^{+\infty} |\varphi(x, \theta)| \, d\theta \right)^2 \frac{2\pi}{d\#(t)}.
$$

Potentially this statement might be weaker than the estimate of Theorem 2.1. However, by making the smooth function $\varphi(x, \theta)$ suitably approximate the sum of point masses

$$
\sum_{j=0}^{\infty} \psi_j(x) \delta_j(\theta),
$$

where $\delta_j$ is the unit point mass at $j$, we see that the two inequalities are of equal strength.
By expanding the square and changing the order of integration in (2.8),
the inequality may be written in the formally equivalent form
\[
\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{0}^{+\infty} \int_{0}^{+\infty} e^{-it(x-y)} d\#(t) \nonumber \\
\cdot \varphi(x, \theta_1) \overline{\varphi}(x, \theta_2) d\theta_1 d\theta_2 \frac{dxdy}{4\pi^2} 
onumber \\
\leq A_{CH} \int_{-\pi}^{\pi} \left( \int_{0}^{+\infty} |\varphi(x, \theta)| d\theta \right)^2 \frac{dx}{2\pi}. 
\]

However, it is not obvious that the change of order in the integration is
permitted. In addition, the innermost integral calls for an interpretation in
the sense of distribution theory. In order to avoid such complications, we
introduce a smoothing parameter \( \varepsilon, 0 < \varepsilon < \frac{1}{2} \). We write
\[
d\#_{\varepsilon}(t) = e^{-\varepsilon t} d\#(t) = \sum_{j=0}^{\infty} e^{-\varepsilon j} d\delta_j(t), 
\]
and observe that since \( d\#_{\varepsilon} \) is smaller than \( d\# \), (2.8) implies that
\[
\int_{[0, +\infty]} \left| \int_{-\pi}^{\pi} e^{-itx} \int_{0}^{t+1} \varphi(x, \theta) d\theta \frac{dx}{2\pi} \right|^2 d\#_{\varepsilon}(t) \nonumber \\
\leq A_{CH} \int_{-\pi}^{\pi} \left( \int_{0}^{+\infty} |\varphi(x, \theta)| d\theta \right)^2 \frac{dx}{2\pi}. 
\]
In this case, we may appeal to the Fubini theorem, and arrive at
\[
\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{0}^{+\infty} \int_{0}^{+\infty} \int_{\max(\theta_1, \theta_2) - 1, +\infty}^{e^{-it(x-y)} d\#_{\varepsilon}(t)} \nonumber \\
\cdot \varphi(x, \theta_1) \overline{\varphi}(y, \theta_2) d\theta_1 d\theta_2 \frac{dxdy}{4\pi^2} 
\leq A_{CH} \int_{-\pi}^{\pi} \left( \int_{0}^{+\infty} |\varphi(x, \theta)| d\theta \right)^2 \frac{dx}{2\pi}.
\]
We sum the geometric series which actually appears on the left-hand side of
the above inequality, and obtain
\[
\int_{\max(\theta_1, \theta_2) - 1, +\infty}^{e^{-it(x-y)} d\#_{\varepsilon}(t)} = \frac{e^{-[\max(\theta_1, \theta_2)](\varepsilon + i(x-y))}}{1 - e^{-(\varepsilon + i(x-y))}}, 
\]
where \([\cdot]\) denotes the operation of taking the integer part. Let the kernel
\( K_{\varepsilon}^{\#} \) be given by
\[
K_{\varepsilon}^{\#}(x, \theta) = \frac{e^{-|\theta|(\varepsilon + i x)}}{1 - e^{-(\varepsilon + i x)}}, 
\]
in terms of which (2.10) simplifies:

\[
\int_0^{+\infty} \int_0^{+\infty} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} K^\#_\varepsilon(x - y, \max(\theta_1, \theta_2)) \varphi(x, \theta_1) \overline{\varphi}(y, \theta_2) \frac{dxdy}{4\pi^2} \ d\theta_1 d\theta_2
\]

\[
\leq A_{CH} \int_{-\pi}^{\pi} \left( \int_0^{+\infty} |\varphi(x, \theta)| \ d\theta \right)^2 \frac{dx}{2\pi}.
\]

The left-hand side of (2.9) increases as \(\varepsilon\) decreases, so that the left-hand side of (2.11) must increase with decreasing \(\varepsilon\) as well. Letting \(\varepsilon\) tend to 0, we therefore obtain the strongest form of (2.11). The kernel \(K^\#_\varepsilon\) then tends to the distribution

\[
K^\#(x, \theta) + \pi \sum_{n=-\infty}^{\infty} \delta_{2\pi n}(x),
\]

where \(\delta_{2\pi n}\) is the unit point mass at \(2\pi n\), and

\[
K^\#(x, \theta) = \text{pv} \ e^{-i[\theta] x} \frac{e^{-i[\theta] x}}{1 - e^{-i\pi x}} = \text{pv} \ e^{-i[\theta] x} \frac{1}{1 - e^{-i\pi x}},
\]

the “pv” standing for the principal value operation (in the \(x\) variable). Given the smoothness of \(\varphi(x, \theta)\), and the fact that in the \(x\) variable, it is supported inside \([-\pi, \pi]\), the integral

\[
\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} K^\#(x - y, \max(\theta_1, \theta_2)) \varphi(x, \theta_1) \overline{\varphi}(y, \theta_2) \frac{dxdy}{4\pi^2}
\]

tends to the expression

\[
\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} K^\#(x - y, \max(\theta_1, \theta_2)) \varphi(x, \theta_1) \overline{\varphi}(y, \theta_2) \frac{dxdy}{4\pi^2}
\]

\[
+ \int_{-\pi}^{\pi} \varphi(x, \theta_1) \overline{\varphi}(x, \theta_2) \frac{dx}{4\pi}
\]

as \(\varepsilon \to 0\), where the singular integral is well-defined as either a distributional convolution in \(x\) first, and then an ordinary Lebesgue integral in \(y\), or we may reverse the order of \(x, y\); it doesn’t matter which we choose to do, because the integral comes out the same. The above convergence (as \(\varepsilon \to 0\)) is easily seen to be uniform in \((\theta_1, \theta_2)\), and because of the assumption that \(\varphi\) has compact support, the integral on the left-hand side of (2.11) tends to

\[
\int_0^{+\infty} \int_0^{+\infty} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} K^\#(x - y, \max(\theta_1, \theta_2)) \varphi(x, \theta_1) \overline{\varphi}(y, \theta_2) \frac{dxdy}{4\pi^2} \ d\theta_1 d\theta_2
\]

\[
+ \int_{-\pi}^{\pi} \left| \int_0^{+\infty} \varphi(x, \theta) \ d\theta \right|^2 \frac{dx}{4\pi}.
\]

It follows that we have obtained yet another reformulation of the Carleson-Hunt theorem.
Theorem 2.2. Suppose \( \varphi \) is a compactly supported \( C^\infty \)-smooth function on the set \( ]-\pi, \pi[ \times [0, +\infty[ \). Then the following estimate holds:

\[
\int_0^{+\infty} \int_0^{+\infty} \int_{-\pi}^{\pi} K \#(x - y, \max(\theta_1, \theta_2)) \varphi(x, \theta_1) \overline{\varphi}(y, \theta_2) \frac{dxdy}{4\pi^2} d\theta_1 d\theta_2 \\
\leq A \int_{-\pi}^{\pi} \left( \int_0^{+\infty} |\varphi(x, \theta)| d\theta \right)^2 \frac{dx}{2\pi},
\]

where \( A = A_{\text{CH}} \) is the absolute constant of Theorem 1.2.

We remark that when we derive the Carleson-Hunt theorem from Theorem 2.2, we get a slightly worse constant, because we dropped a term for aesthetic reasons.

The property of the kernel \( K\# \) which ensures that the integral expression on the left-hand side of the inequality of the above theorem is real-valued can be formalized: It is the symmetry requirement of a given kernel \( K(x, \theta) \) that

\[
\overline{K}(x, \theta) = K(-x, \theta).
\]

Let \( K(x, \theta) \) have the above symmetry property and suppose that it is uniformly within finite distance from \( K\# \) on a symmetric interval:

\[
|K(x, \theta) - K\#(x, \theta)| \leq C, \quad (x, \theta) \in ]-2\alpha, 2\alpha[ \times [0, +\infty[,
\]

where \( \alpha \) and \( C \) are a positive real numbers. Just as \( K\# \), we define \( K \) as a principal value distribution on \( ]-2\alpha, 2\alpha[ \). Let \( \alpha \) be confined to the interval \( 0 < \alpha < \frac{1}{2} \pi \), so that we have the above inequality fulfilled for the particular choice

\[
K(x, \theta) = \text{pv} \frac{e^{-i\theta x}}{ix} = e^{-i\theta x} \text{pv} \frac{1}{ix}, \quad x \in \mathbb{R} \setminus \{0\}, \quad \theta \in \mathbb{R}.
\]

Then since

\[
\int_0^{+\infty} \int_0^{+\infty} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |\varphi(x, \theta_1)| |\varphi(y, \theta_2)| \frac{dxdy}{4\pi^2} d\theta_1 d\theta_2 \\
= \left( \int_0^{+\infty} \int_{-\pi}^{\pi} |\varphi(x, \theta)| \frac{dx}{2\pi} d\theta \right)^2 \\
\leq \int_{-\pi}^{\pi} \left( \int_0^{+\infty} |\varphi(x, \theta)| d\theta \right)^2 \frac{dx}{2\pi},
\]

an application of Theorem 2.2 yields that

\[
(2.12)
\int_0^{+\infty} \int_0^{+\infty} \int_{-\pi}^{\pi} K(x - y, \max(\theta_1, \theta_2)) \varphi(x, \theta_1) \overline{\varphi}(y, \theta_2) \frac{dxdy}{4\pi^2} d\theta_1 d\theta_2 \\
\leq (A + C) \int_{-\pi}^{\pi} \left( \int_0^{+\infty} |\varphi(x, \theta)| d\theta \right)^2 \frac{dx}{2\pi},
\]
provided that the compact support of the function \( \varphi \) is contained in \( ]-\alpha, \alpha[ \times [0, +\infty[ \). An analysis of the derivation of Theorem 2.2 reveals that the estimate of that theorem, restricted to functions \( \varphi \) supported on \( ]-\alpha, \alpha[ \times [0, +\infty[ \), is equivalent to the following estimate for the maximal function:

\[
\int_{-\alpha}^{\alpha} |Mf(x)|^2 \frac{dx}{2\pi} \leq A' \int_{-\pi}^{\pi} |f(x)|^2 \frac{dx}{2\pi}, \quad f \in H^2(\mathbb{D}),
\]

for some absolute constant \( A' \). By rotation invariance, this estimate is equipotent with the estimate of the Carleson-Hunt theorem, albeit that it leads to a worse absolute constant. In other words, we have the following reformulation of Theorem 1.2, if we accept that the size of the absolute constant is unimportant.

**Theorem 2.3.** Suppose \( 0 < \alpha < \frac{1}{2}\pi \), and that \( \varphi \) is a compactly supported \( C^\infty \)-smooth function on the set \( ]-\alpha, \alpha[ \times [0, +\infty[ \), extended to vanish elsewhere. Then the following estimate holds:

\[
\int_{0}^{+\infty} \int_{0}^{+\infty} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{p}{i(x-y, \max(\theta_1, \theta_2))} \varphi(x, \theta_1) \overline{\varphi(y, \theta_2)} \, dx \, dy \, d\theta_1 \, d\theta_2 \leq B \int_{-\pi}^{\pi} \left( \int_{0}^{+\infty} |\varphi(x, \theta)| \, d\theta \right)^2 \, dx,
\]

where \( B \) is an absolute constant.

In the formulation in the introduction, \( \varphi \) was allowed to have compact support in \( ]-\pi, \pi[ \times [0, +\infty[ \). To cover that case, we recall the specific choice of \( K \) there, and make a suitable dilation in both variables \((x, \theta)\).

Theorem 2.3 suggests possible generalizations of the maximal function estimate of the Carleson-Hunt theorem. For instance, what if we replace the \( L^1 \) norm on the right-hand side by a slightly weaker norm expression?

**Remark 2.4.** In the sharp constant form, Hilbert’s inequality states that

\[
\left| \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \text{pv} \frac{1}{i(x-y)} \varphi(x) \overline{\varphi(y)} \, dx \, dy \right| \leq \pi \int_{-\pi}^{\pi} |\varphi(x)|^2 \, dx,
\]

and there is a variant which applies to the bilinear form with two functions \( \varphi, \psi \). As we apply that version to the left-hand side of the expression in Theorem 2.3, we obtain

\[
\left| \int_{0}^{+\infty} \int_{0}^{+\infty} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} K(x-y, \max(\theta_1, \theta_2)) \varphi(x, \theta_1) \overline{\varphi(y, \theta_2)} \, dx \, dy \, d\theta_1 \, d\theta_2 \right|
\leq \pi \left( \int_{0}^{+\infty} \left( \int_{-\pi}^{\pi} |\varphi(x, \theta)|^2 \, dx \right)^{\frac{1}{2}} \, d\theta \right)^2,
\]
which estimate differs from that of Theorem 2.3 in that the $L^2$ in the $x$ variable is taken first, and the $L^1$ norm in $\theta$ second. Of course this estimate is worse (if we forget about the size of the constant, that is), because the triangle inequality guarantees that taking the $L^1$ norm in $\theta$ first and then the $L^2$ norm in $x$ produces a smaller quantity.

3. Proof of Theorem 1.1.

We assume that

$$\sum_{n=1}^{\infty} |a_n|^2 < +\infty,$$

where $a_1, a_2, a_3, \ldots$ is a complex-valued sequence, and study the convergence properties of the Dirichlet series

$$f(t) = \sum_{n=1}^{\infty} a_n n^{-\frac{1}{2}-it}, \quad t \in \mathbb{R}. \quad (3.1)$$

We have changed the notation a little: Previously, the above function would have been denoted by $f(\frac{1}{2} + it)$. We mentioned in the introduction that $f$ makes sense as a locally square-integrable function on $\mathbb{R}$ (see [7]). The corresponding Dirichlet series maximal function is defined by

$$M_D f(t) = \sup_{j} \left| \sum_{n=j}^{\infty} a_n n^{-\frac{1}{2}-it} \right|, \quad t \in \mathbb{R},$$

where $j$ ranges over $\{1, 2, 3, \ldots \}$, and the infinite sum is interpreted as

$$\sum_{n=j}^{\infty} a_n n^{-\frac{1}{2}-it} = f(t) - \sum_{n=1}^{j-1} a_n n^{-\frac{1}{2}-it}.$$

By translation invariance, we need only prove the almost everywhere convergence on the interval $[-1, 1]$. We will do this by obtaining the following maximal function estimate:

$$\int_{-1}^{1} |M_D f(t)|^2 dt \leq B \sum_{n=1}^{\infty} |a_n|^2, \quad (3.2)$$

where $B$ is an absolute constant. The derivation of the convergence statement then follows from a standard argument (see [21, Proposition 6.2]), which was used for Taylor series as well: Approximate $f$ by a Dirichlet polynomial $g$ such that the norm of the difference is small in $\mathcal{H}$, and apply the maximal function estimate to the difference, whence the assertion follows by inspection.
We proceed exactly as in the proof of the first dual formulation (Theorem 2.1) of the Carleson-Hunt theorem. We obtain

\[
\left( \int_{-1}^{1} |M_D f(t)|^2 dt \right)^\frac{1}{2} = \sup_g \int_{-1}^{1} M_D f(t) \overline{g}(t) dt \\
= \sup_{g, \phi_j} \text{Re} \int_{-1}^{1} \sum_{n=1}^{\infty} a_n n^{-\frac{1}{2}} e^{it \log n} \sum_{j=1}^{n} \phi_j(t) \overline{g}(t) dt \\
= \sup_{g, \phi_j} \text{Re} \sum_{n=1}^{\infty} a_n \sum_{j=1}^{n} n^{-\frac{1}{2}} \overline{\hat{\phi}_j} g(\log n),
\]

where \( g \) is restricted by the condition that it is Lebesgue measurable and

\[
\int_{-1}^{1} |g(t)|^2 dt \leq 1,
\]

and \( \phi_j \) is a sequence of bounded Lebesgue measurable functions with

\[
\sum_{j=1}^{\infty} |\phi_j(t)| \leq 1, \quad t \in [-1, 1].
\]

These functions are extended to vanish off \([-1, 1]\), and the Fourier transform appearing in the above formula is given by

\[
\hat{\varphi}(\xi) = \int_{-\infty}^{+\infty} e^{-it\xi} \varphi(t) dt, \quad \xi \in \mathbb{R}.
\]

As in the previous Taylor series case, we have some difficulty motivating why we may change the order of summation, and we resolve the difficulty by assuming that only finitely many of the coefficients \( a_n \) are different from 0. Such functions (Dirichlet polynomials) are dense in \( \mathcal{H} \), and it suffices to obtain the maximal function estimate (3.2) for them. It follows from the Cauchy-Schwarz inequality applied to (3.3) that we just need to obtain the estimate

\[
\sum_{n=1}^{\infty} \frac{1}{n} \left| \sum_{j=1}^{n} \hat{\phi}_j g(\log n) \right|^2 \leq B,
\]

assuming that \( g \) and \( \phi_j \) are as above. As before we find that this is equivalent to establishing for sequences \( \psi_1, \psi_2, \psi_3, \ldots \) of Lebesgue measurable functions on \([-1, 1]\) the inequality

\[
\sum_{n=1}^{\infty} \frac{1}{n} \left| \sum_{j=1}^{n} \hat{\psi}_j(\log n) \right|^2 \leq B \int_{0}^{1} \left( \sum_{j=1}^{\infty} |\psi_j(t)| \right)^2 dt.
\]
We again proceed as in Section 3, with the counting measure $d\#$ on the nonnegative integers replaced by the sum of point masses
\[ \sum_{n=1}^{\infty} \frac{1}{n} \delta_{\log n}, \]
and obtain that the above estimate is equivalent to having, for $C^\infty$-smooth compactly supported functions $\varphi$ on $]-1, 1[\times[0, +\infty[$,
\begin{equation}
\int_{0}^{+\infty} \int_{0}^{+\infty} \int_{-1}^{1} \int_{-1}^{1} K^D(x - y, \max(\theta_1, \theta_2)) \varphi(x, \theta_1) \overline{\varphi}(y, \theta_2) \, dxdy \, d\theta_1 \, d\theta_2 \\
\leq B' \int_{0}^{1} \left( \int_{0}^{+\infty} |\varphi(x, \theta)| \, d\theta \right)^2 \, dx,
\end{equation}
where $B'$ is some absolute constant. The kernel $K^D$ is essentially the distributional limit of the kernels $K^D_\varepsilon$ as $0 < \varepsilon \to 0$ (we want $K^D$ to be a principal value distribution, and the limit of the kernels $K^D_\varepsilon$ contains a point mass at the origin, which needs to be removed), where, for $\varepsilon$, $0 < \varepsilon < +\infty$, $K^D_\varepsilon$ is given by the expression
\[ K^D_\varepsilon(x, \theta) = \sum_{n=[e^{\theta}]}^{\infty} n^{-1-\varepsilon-ix}. \]
We approximate the sum by the corresponding integral expressions,
\[
K^D_\varepsilon(x, \theta) - \frac{e^{-\theta(\varepsilon+ix)}}{\varepsilon + ix} \\
= \int_{[e^{\theta}, +\infty[} (d\#(t) - dt) \\
= e^{-(1+\varepsilon+ix)\theta} (e^{\theta} - [e^{\theta}]') - (1 + \varepsilon + ix) \int_{e^{\theta}}^{+\infty} t^{-2-\varepsilon-ix}(t - [t]) \, dt,
\]
where we rely on integration by parts. The notation $[x]'$ stands for the integer part of $x$, except that if $x$ is an integer, we get $x - 1$ instead of $x$. It follows that we have
\[
\left| K^D_\varepsilon(x, \theta) - \frac{e^{-\theta(\varepsilon+ix)}}{\varepsilon + ix} \right| \leq 4, \quad 0 \leq \theta < +\infty, \quad -2 \leq x \leq 2.
\]
As $0 < \varepsilon \to 0$, the expression
\[ \frac{e^{-\theta(\varepsilon+ix)}}{\varepsilon + ix} \]
tends to the distribution $K(x,\theta) + \pi \delta_0(x)$, where $K$ is as before:

$$K(x,\theta) = \text{pv} \frac{e^{-i\theta x}}{i x} = e^{-i\theta x} \frac{1}{i x}.$$  

Let $K^D_\varepsilon$ be the distributional limit of the kernels $K^D_\varepsilon$ as $0 < \varepsilon \to 0$, and put

$$K^D(x,\theta) = K^D_\varepsilon(x,\theta) - \pi \delta_0(x),$$

so that $K^D$ becomes a principal value distribution near $x = 0$. It suffices to obtain (3.4) for this particular kernel. By the above, $K^D$ is uniformly close to the kernel $K$:

$$|K^D(x,\theta) - K(x,\theta)| \leq 4, \quad (x,\theta) \in [-2,2] \times [0, +\infty[.$$  

The desired boundedness is now a consequence of Theorem 2.3.

4. Proof of Theorem 1.5.

As in the previous section, $a_1, a_2, a_3, \ldots$ is a sequence of complex numbers, subject to the square summability condition

$$\sum_{n=1}^{\infty} |a_n|^2 < +\infty.$$  

The function we shall study is the infinite power series

$$(4.1) \quad f(\chi) = \sum_{n=1}^{\infty} a_n \chi(n), \quad \chi \in \Xi,$$

which defines a square summable function on $\Xi$, by standard Fourier analysis (see [7]):

$$\|f\|_{L^2(\Xi)}^2 = \int_{\Xi} |f(\chi)|^2 d\varpi(\chi) = \sum_{n=1}^{\infty} |a_n|^2.$$  

The space of all such $f$ is denoted by $H^2(\Xi)$, and is in a natural way the Hardy space on the infinite-dimensional polydisk $\mathbb{D}^\infty$. The corresponding maximal function is

$$(4.2) \quad M_{\Xi} f(\chi) = \sup_j \left| \sum_{n=j}^{\infty} a_n \chi(n) \right|, \quad \chi \in \Xi,$$

where $j$ ranges over $\{1, 2, 3, \ldots \}$, and the infinite sum is interpreted as

$$\sum_{n=j}^{\infty} a_n \chi(n) = f(\chi) - \sum_{n=1}^{j-1} a_n \chi(n).$$
As before the claimed almost convergence follows as soon as we obtain the estimate
\[
\|M_\Xi f\|_{L^2(\Xi)}^2 \leq B \|f\|_{L^2(\Xi)}^2,
\]
for some finite absolute constant $B$, by a standard approximation argument: We take a Dirichlet polynomial which approximates $f$ in norm, apply the estimate (4.3) to the difference, and the almost everywhere convergence assertion follows.

For $N = 1, 2, 3, \ldots$, we denote by $\mathcal{N}_N$ the set of all positive integers whose prime factorizations contain only the first $N$ primes $p_1, p_2, \ldots, p_N$. We define the corresponding space $H^2_{\mathcal{N}_N}(\Xi)$ consisting of all functions $f \in H^2(\Xi)$ with series expansion (4.1) for which $a_n = 0$ unless $n \in \mathcal{N}_N$. If we put $z_j = \chi(p_j)$, and think of $f \in H^2_{\mathcal{N}_N}(\Xi)$ as a function of $(z_1, z_2, \ldots, z_N)$, the series expansion (4.1) can be written as
\[
f(z_1, z_2, \ldots, z_N) = \sum_{k_1, \ldots, k_N = 0}^{\infty} a_n z_1^{k_1} \cdots z_N^{k_N}, \quad n = p_1^{k_1} \cdots p_N^{k_N},
\]
so that $H^2(\Xi)$ may be identified with $H^2(\mathbb{D}^N)$, the Hardy space on the finite-dimensional polydisk $\mathbb{D}^N$. We shall obtain the maximal function estimate (4.3) for functions $f \in H^2_{\mathcal{N}_N}(\Xi)$, with a constant $B$ independent of $N$. The assertion (4.3) then follows in general by a standard approximation argument.

We write $z_j = e^{i\theta_j}$, where $\theta_j$ is a real parameter, and
\[
a_{k_1, \ldots, k_N} = a_n \quad \text{provided that} \quad n = p_1^{k_1} \cdots p_N^{k_N},
\]
and extend $a_{k_1, \ldots, k_N}$ to vanish whenever one of the indices $k_1, \ldots, k_N$ is negative. Modulo a slight abuse of notation, we then have
\[
f(\theta_1, \ldots, \theta_N) = \sum_{k_1, \ldots, k_N = 0}^{\infty} a_{k_1, \ldots, k_N} e^{ik_1 \theta_1 + \cdots + ik_N \theta_N}, \quad n = p_1^{k_1} \cdots p_N^{k_N}.
\]
The corresponding maximal function is
\[
M_N f(\theta_1, \ldots, \theta_N) = \sup_j \left| \sum_{(k_1, \ldots, k_N) \in \mathcal{G}(\log j, N)} a_{k_1, \ldots, k_N} e^{ik_1 \theta_1 + \cdots + ik_N \theta_N} \right|,
\]
where $\mathcal{G}(T, N)$ consists of all vectors $(k_1, \ldots, k_N)$ where the entries are nonegative integers, and the following inequality holds:
\[
T \leq k_1 \log p_1 + k_2 \log p_2 + \cdots + k_N \log p_N.
\]
We need to prove that there exists an absolute constant $B$ such that

$$
\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} |M_N f(\theta_1, \ldots, \theta_N)|^2 \frac{d\theta_1}{2\pi} \cdots \frac{d\theta_N}{2\pi} \leq B \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} |f(\theta_1, \ldots, \theta_N)|^2 \frac{d\theta_1}{2\pi} \cdots \frac{d\theta_N}{2\pi}
$$

$$
= B \sum_{k_1, \ldots, k_N = 0}^{\infty} |a_{k_1, \ldots, k_N}|^2.
$$

We turn to Fefferman’s idea \cite{3}. It suffices to obtain (4.4) for functions $f$ whose coefficients are nonzero for a finite number of indices only. Suppose we know that the above maximal function estimate (4.4) holds when $G(T, N)$ is replaced by $G_x(T, N)$, consisting of all vectors $(k_1, \ldots, k_N)$ with nonnegative integer entries subject to

$$
T \leq k_1 x_1 + k_2 x_2 + \cdots + k_N x_N,
$$

for a dense $(0, +\infty]^N$ collection of vectors $x = (x_1, \ldots, x_N)$ with positive rational entries. By approximation, then, (4.4) holds, with the original set $G(T, N)$, for all functions $f$, the coefficients of which are nonzero for a finite number of indices only.

We consider $x$ of the form $(q_1/Q, q_2/Q, \ldots, q_N/Q)$, where $q_j$ and $Q$ are positive integers, and the integers $q_1, q_2$ are relatively prime. We may pick $x_1$ and $x_2$ as arbitrarily positive rationals, and let $Q$ be the least possible positive integer such that $q_1 = Q x_1$ and $q_2 = Q x_2$ are integers: Then $q_1$ and $q_2$ are automatically relatively prime. The set of pairs $(x_1, x_2)$ for which $Q$ gets as large as we prescribe is dense in $[0, +\infty]^2$, and hence the rational vectors $x$ of the above form are dense in $[0, +\infty]^N$. The maximal function for slope $x$ is

$$
M_{N,x} f(\theta) = \sup_{0 < T < +\infty} \left| \sum_{k \in G(T, N)} a_k e^{i \langle k, \theta \rangle} \right|, \quad \theta \in \mathbb{R}^N,
$$

where $k = (k_1, \ldots, k_N)$, $\theta = (\theta_1, \ldots, \theta_N)$, and $\langle k, \theta \rangle$ is

$$
\langle k, \theta \rangle = k_1 \theta_1 + \cdots + k_N \theta_N.
$$

The criterion that $k \in G_x(T/Q, N)$ can be written

$$
T \leq k_1 q_1 + k_2 q_2 + \cdots + k_N q_N = \langle k, q \rangle,
$$

where $q = (q_1, \ldots, q_N)$, which suggests calling this set $G_q(T, N)$. The maximal function for slope $x$ can then be written as

$$
M_{N,x} f(\theta) = \sup_{0 \leq T < +\infty} \left| \sum_{k \in G_q(T, N)} a_k e^{i \langle k, \theta \rangle} \right|, \quad \theta \in \mathbb{R}^N.
$$
Since $q_1$ and $q_2$ are assumed relatively prime, we can find integers $r_1, r_2$ such that $q_1r_2 - q_2r_1 = 1$, so that the $N \times N$ matrix $A$,

$$A = \begin{pmatrix}
q_1 & q_2 & q_3 & q_4 & \cdots & q_n \\
r_1 & r_2 & 0 & 0 & \cdots & 0 \\
0 & 0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \cdots & 1
\end{pmatrix},$$

has determinant 1. The matrix $A$ has integer entries, and by Cramer’s rule, so does the inverse matrix $A^{-1}$. Consequently, the map $A: \mathbb{Z}^n \to \mathbb{Z}^n$ is one to one and onto ($\mathbb{Z}$ is the set of all integers). Putting $j = Ak$, where $j, k$ are thought of as column vectors, it follows that the maximal function for slope $x$ can be rewritten as

$$M_{N,x} f(\theta) = \sup_{0 \leq T < +\infty} \left| \sum_{j \in A\mathcal{G}_q(T,N)} a_{A^{-1}j} e^{i\langle j, (A^{-1})^* \theta \rangle} \right|,$$

where the superscript * indicates that the matrix is transposed. In view of the definition of the index set $\mathcal{G}_q(T,N)$, $j \in A\mathcal{G}_q(T,N)$ means that $T \leq j_1$, where $j = (j_1, \ldots, j_N)$. After the change of variables $\phi = (A^{-1})^* \theta$, we have, because this transformation is volume-preserving,

$$\int_{-\pi,\pi} |M_{N,x} f(\theta)|^2 \frac{d\theta}{(2\pi)^N} = \int_{-\pi,\pi} \sup_{0 \leq T < +\infty} \left| \sum_{j \in A\mathcal{G}_q(T,N)} a_{A^{-1}j} e^{i\langle j, \phi \rangle} \right|^2 \frac{d\phi}{(2\pi)^N},$$

where $d\theta = d\theta_1 \cdots d\theta_N$ and $d\phi = d\phi_1 \cdots d\phi_N$. Actually, the image of $]-\pi,\pi[^N$ under the linear transformation induced by $(A^{-1})^*$ probably is not $]-\pi,\pi[^N$, but in any case an equivalent domain in $\mathbb{R}^N$ modulo $(2\pi\mathbb{Z})^N$. We write $j = (j_1, j')$, where $j' = (j_2, \ldots, j_N) \in \mathbb{Z}^{N-1}$, and do the same for $\phi$. Then

$$\sum_{j \in A\mathcal{G}_q(T,N)} a_{A^{-1}j} e^{i\langle j, \phi \rangle} = \sum_{j_1 \geq T} \sum_{j' \in \mathbb{Z}^{N-1}} a_{A^{-1}j} e^{i\langle j', \phi' \rangle} e^{ij_1\phi_1},$$

where we are able to extend the latter summation over the whole lattice $\mathbb{Z}^{N-1}$ by our convention that $a_j = 0$ if any component of $j$ is negative. Applying the one-dimensional maximal function estimate of Theorem 1.2 (which holds for $L^2(\mathbb{T})$ functions, not just for those in $H^2(\mathbb{D})$) with respect
to the variable $\theta_1$, we find that for fixed $\phi'$,

$$
\int_{-\pi}^{\pi} \sup_{0 \leq T < +\infty} \left| \sum_{j_1 \geq T} \sum_{j' \in \mathbb{Z}^{N-1}} a_{A^{-1}j} e^{i(j',\phi')} e^{i j_1 \phi_1} \right|^2 \frac{d\theta_1}{2\pi} \leq A_{CH} \int_{-\pi}^{\pi} \left| \sum_{j' \in \mathbb{Z}^{N-1}} a_{A^{-1}j} e^{i(j',\phi')} \right|^2 \frac{d\phi_1}{2\pi}
$$

$$
= A_{CH} \int_{-\pi}^{\pi} \left| \sum_{j \in \mathbb{Z}^N} a_{A^{-1}j} e^{i(j,\phi)} \right|^2 \frac{d\phi_1}{2\pi},
$$

where $A_{CH}$ is the absolute constant of the Carleson-Hunt theorem. Integrating with respect to all the other parameters $\theta_2, \ldots, \theta_N$, we obtain

$$
\int_{-\pi}^{\pi} \sup_{0 \leq T < +\infty} \left| \sum_{j_1 \geq T} \sum_{j' \in \mathbb{Z}^{N-1}} a_{A^{-1}j} e^{i(j',\phi')} \right|^2 \frac{d\theta_1}{2\pi} \leq A_{CH} \int_{-\pi}^{\pi} \left| \sum_{j' \in \mathbb{Z}^{N-1}} a_{A^{-1}j} e^{i(j',\phi')} \right|^2 \frac{d\phi_1}{2\pi}
$$

$$
= A_{CH} \int_{-\pi}^{\pi} \left| \sum_{j \in \mathbb{Z}^N} a_{A^{-1}j} e^{i(j,\phi)} \right|^2 \frac{d\phi_1}{(2\pi)^N} = A_{CH} \sum_{j \in \mathbb{Z}^N} |a_{A^{-1}j}|^2
$$

as desired. Hence (4.4) holds with the constant $B = A_{CH}$, where $A_{CH}$ is the absolute constant of the Carleson-Hunt theorem for Fourier series (which may be slightly larger than for Taylor series). The proof is complete.

5. Comments on character sums.

We next comment on the relation of Theorem 1.5 to the question mentioned in the introduction — essentially the same as the problem raised by Erdős — concerning the almost sure growth of the character sums

$$
S_N(\chi) = \sum_{n=1}^{N} \chi(n).
$$

Erdős [2, pp. 251-252] studies the sum $S_N(\chi)$ in the analogous case where for prime indices $p$, the random variables $\chi(p)$ are independent and take on the values $\pm 1$ with equal probability $\frac{1}{2}$. At other positive integers, the value of the character is then determined by the multiplicative rule $\chi(mn) = \chi(m) \chi(n)$. He announced the existence of a positive constant $c$ such that for almost every character $\chi$, one has

$$
S_N(\chi) = O(\sqrt{N} \log^c N), \quad \text{as} \quad N \to +\infty.
$$
A somewhat unfortunate feature of the numbers ±1 is that the square of each is 1. That means that when \( n \) is a square, we know with certainty that \( \chi(n) = 1 \). This peculiarity will influence any number containing a square as a factor. This is in sharp contrast with the case when for prime indices \( p \) the character \( \chi(p) \) is uniformly distributed on \( \mathbb{T} \), because then \( \chi(n) \) keeps the same uniform distribution on \( \mathbb{T} \) for all \( n = 2, 3, 4, \ldots \). Instead of switching from ±1 to \( \mathbb{T} \), there is another way to alleviate the difficulty. Namely, we decide to sum only over the square-free numbers, defined by the property that they are not divisible by any square (other than 1); these are the numbers in whose factorization each prime occurs at most once. The square-free numbers have density \( 6/\pi^2 \) in the positive integers [18, p. 390]. Wintner opts to sum only over square-free numbers [24] (his reason is different, though: He studies the reciprocal of the zeta function, where the Dirichlet coefficients are given by the Mœbius function \( \mu \), and he wants to know what happens when \( \mu \) is replaced by a random function), but as his work predates that of Erdős, we cannot expect a better estimate. Indeed, he obtain the upper estimate \( O(N^{1/2+\varepsilon}) \), but he also shows that \( O(N^{1/2-\varepsilon}) \) is almost surely false (here, \( \varepsilon \) stands for an arbitrarily small positive number). Halász [6] makes a considerably deeper study of the same problem. He obtains the estimate from above

\[
S_N(\chi) = O(\sqrt{N} e^{c\sqrt{\log \log N \log \log \log N}}), \quad \text{as} \quad N \to +\infty,
\]

almost surely in \( \chi \), for some positive constant \( c \), and also finds that there is some negative \( c \) such that the above estimate fails almost surely. Although quite sharp, these estimates leave some room for improvement. For instance, if we — as does Erdős — compare with the law of the iterated logarithm, where the sharp growth is known to be \( O(\sqrt{N \log \log N}) \) (see [20, p. 397]), we cannot say whether the multiplicativity makes the sum behave better or worse than this. There is some reason to believe that it makes it better (that is, smaller), and indeed, in the appendix of Montgomery’s monograph [17], problem number 26, due to Halász, asks whether we have \( S_N(\chi) = O(\sqrt{N}) \) almost surely.

As the peculiar difficulties with squares encountered when summing over all the positive integers vanish when we turn to having \( \chi(p) \) uniformly distributed on \( \mathbb{T} \), we expect that Halasz’ methods will carry over and supply similar estimates in this setting. We do not pursue this idea further here. Instead, we restrict ourselves to recording a nontrivial upper bound of Erdős type as a corollary of Theorem 1.5.

Suppose \( \{a_n\}_{n=1}^\infty \) is an arbitrary positive, decreasing, square-summable sequence. Applying summation by parts to Theorem 1.5 (in the form known
as Kronecker’s lemma, see [20, p. 390]), we obtain the estimate
\[ S_N(\chi) = O\left(\frac{1}{a_N}\right) \quad \text{as} \quad N \to +\infty. \]

With a suitable choice of the sequence \( \{a_n\}_{n=1}^\infty \) we have the following:

**Corollary 5.1.** For any \( \varepsilon > 0 \), we have, for almost every character \( \chi \),
\[ \sum_{n=1}^N \chi(n) = O\left(\sqrt{N \log N \left( \log \log N \right)^{1+\varepsilon}}\right) \quad \text{as} \quad N \to +\infty. \]

This settles a question in [7, p. 22], which asked for the estimate \( O(\sqrt{N \log N}) \). It is better than the bounds obtained directly from the classical Menshov-Steinhaus theorem (see [13] or [19]).

We finally mention a lower bound that can be obtained in a relatively simple manner by observing the growth of the ‘random zeta function’
\[ \zeta_\chi(s) = \sum_{n=1}^\infty \chi(n) n^{-s} \]
as \( s \to \frac{1}{2} \) along the positive real line. The proof is omitted.

**Proposition 5.2.** For almost every character \( \chi \), the following estimate fails:
\[ \sum_{n=1}^N \chi(n) = O\left(\frac{\sqrt{N}}{\log N}\right) \quad \text{as} \quad N \to +\infty. \]

There is (so far quite weak) reason to believe that only the randomness really contributes to growth. For instance, in the the totally deterministic case of additive characters on the positive integers, the partial sums are \( O(1) \) almost everywhere on the unit circle. In contrast, the completely random case of the law of the iterated logarithm gives \( O(\sqrt{N \log \log N}) \). We are led to ask whether in the multiplicative case we might have almost everywhere
\[ \sum_{n=1}^N \chi(n) = O\left(\sqrt{N \log \log N \log N}\right) \quad \text{as} \quad N \to +\infty, \]
by counting the number of primes below \( N \) and taking into account the \( \log \log \) contribution of the law of the iterated logarithm. Of course, if this is true, then the question of Halász mentioned above automatically gets an affirmative answer (in the setting of uniformly distributed random variables on the unit circle, that is).

**Acknowledgement.** We wish to thank Bo Berndtsson, Lennart Carleson, Peter Lindqvist, Kristian Seip, and Sergei Shimorin for inspiring conversations.
References


Received May 25, 2001.

**Department of Mathematics**
Lund University, Box 118
S-22100 Lund
Sweden

E-mail address: haakan@maths.lth.se

**Department of Mathematics**
The Royal Institute of Technology
S-100 44 Stockholm
Sweden

**Department of Mathematics**
Department of Mathematics and Statistics
University of Jyväskylä
P.O. Box 35 (MaD)
40014 University of Jyväskylä
Finland

E-mail address: saksman@maths.jyu.fi
AN INVERSE PROBLEM FROM SUB-RIEMANNIAN GEOMETRY

THOMAS A. IVEY

The geodesics for a sub-Riemannian metric on a three-dimensional contact manifold $M$ form a 1-parameter family of curves along each contact direction. However, a collection of such contact curves on $M$, locally equivalent to the solutions of a fourth-order ODE, are the geodesics of a sub-Riemannian metric only if a sequence of invariants vanish. The first of these, which was first identified by Fels, determines if the differential equation is variational. The next two determine if there is a well-defined metric on $M$ and if the given paths are its geodesics.

Introduction.

In this note we discuss the problem of recovering the geometric structure of a three-dimensional contact manifold with a sub-Riemannian metric from the geodesics for the metric. (Sub-Riemannian metrics are also known as Carnot-Caratheodory metrics.) Since all the results herein will be local in nature, the manifold may be taken to be an open set $U \in \mathbb{R}^3$ with contact form $dy - zdx$, and we may assume that on contact planes the metric has the form

$$g = Edx^2 + Fdxdz + Gdz^2,$$

where $E, F, G$ are smooth functions on $U$ such that $g$ is positive definite. The geodesics form a collection of paths tangent to the contact structure, such that there is a 1-parameter family of distinct paths tangent to each contact direction at each point. Thus, part of the problem will be to determine which such collections of paths come from a sub-Riemannian metric.

As explained below, the paths are locally equivalent to the integral curves of a scalar fourth-order ODE. The variational multiplier problem for fourth-order ODE—i.e., the problem of characterizing equations which are, up to multiple, the Euler-Lagrange equations for a second-order Lagrangian—was solved by M. Fels [4]. Since sub-Riemannian geodesics arise as solutions of a variational problem, the present work is an extension of that of Fels; to avoid confusion, the notation of [4] will be used whenever possible.
1. Contact path geometries.

In this section we review the construction of sub-Riemannian geodesics, and define a $G$-structure canonically associated to the geodesics as paths.

Let $M$ be an oriented three-manifold with contact distribution $\mathcal{D}$ and sub-Riemannian metric $g$. It is standard that one can associate to $g$ a $SO(2)$-structure $\mathcal{N}$ inside the oriented coframe bundle $\mathcal{F}(M)$, such that, for any coframing which is a local section of $\mathcal{N}$, the forms $(\omega^1, \omega^2, \omega^3)$ of the coframing satisfy:

(i) $\omega^3$ annihilates the contact planes;
(ii) $(\omega^1)^2 + (\omega^2)^2$ coincides with the metric on the contact planes; and,
(iii) $\omega^1 \wedge \omega^2$ gives the induced orientation.

Furthermore, we may specify $\mathcal{N}$ uniquely by requiring that $d\omega^3 = \omega^1 \wedge \omega^2$. Then $\mathcal{N}$ has a connection form $\phi$ satisfying the following structure equations:

\[
\begin{align*}
    d\omega^1 &= \phi \wedge \omega^2 + (a_1 \omega^1 + a_2 \omega^2) \wedge \omega^3 \\
    d\omega^2 &= -\phi \wedge \omega^1 + (a_2 \omega^1 - a_1 \omega^2) \wedge \omega^3 \\
    d\omega^3 &= \omega^1 \wedge \omega^2 \\
    d\phi &= K \omega^1 \wedge \omega^2 \mod \omega^3.
\end{align*}
\]

The functions $a_1, a_2$ are components of the torsion of $g$ and $K$ is the curvature.\(^2\)

Every oriented contact curve in $M$ has a lift to $\mathcal{N}$ on which the forms $\omega^2$ and $\omega^3$ vanish. Applying the Griffiths formalism \([7]\) to find the integral curves in $\mathcal{N}$ of the Pfaffian system $\{\omega^2, \omega^3\}$ which are extremal curves for arclength $\int \omega^1$, we obtain the following characterization of sub-Riemannian geodesics:\(^3\)

**Proposition 1.1.** Let $Z$ be the rank one affine subbundle of $T^*N$ on which the canonical one-form is $\sigma = \omega^1 - x \omega^3$, $x \in \mathbb{R}$. (Forms on $Z$ are pulled back via $\pi : Z \to N$.) Then smooth geodesics are in 1-to-1 correspondence, via the submersion $Z \to M$, with integral curves of the Pfaffian system

---

\(^1\)This result appears in \([6]\), where it is attributed independently to Bryant-Hsu and to G. Wilkens. A detailed derivation can be found in \([9]\).

\(^2\) Clearly, taking the $\omega^i$ as an orthonormal coframe canonically associates to $g$ with Riemannian metric $\tilde{g}$ on $M$, which induces $g$ on $\mathcal{D}$, and defines a canonical foliation perpendicular to $\mathcal{D}$. The torsion tensor is the Lie derivative of $\tilde{g}$ along the leaves; if this vanishes, $g$ descends to any (locally defined) quotient surface by foliation, and $K$ is the Gauss curvature of the metric on that surface.

\(^3\)See \([9]\) and \([10]\) for derivations of the geodesics by this and other methods.
\[ F = \{ \theta_0, \theta_1, \theta_2, \theta_3 \} \text{ on } Z, \text{ where} \]

\[
\begin{align*}
\theta_0 &= \omega^3 \\
\theta_1 &= \omega^2 \\
\theta_2 &= \phi - x\omega^1 \\
\theta_3 &= dx - a_1\omega^1 - a_2\omega^2.
\end{align*}
\]

**Remark.** For variational problems with differential constraints, it is in general not known under what conditions all extremal curves arise as projections of integral curves of the differential system formulated by Griffiths. For example, for sub-Riemannian metrics on a generic two-plane distribution in dimension four, exceptional extremal curves exist which do not come from the Griffiths system. Essentially, this is because these *abnormal minimizers* \([2, 10]\) have few or no compactly supported variations that are tangent to the given distribution. However, by applying the regularity test given by Hsu \([8]\), one can show that for a sub-Riemannian metric on a contact manifold, all geodesics arise via the Griffiths formalism. (Intuitively, enough variations exist because contact curves can be locally expressed in terms of an arbitrary function and its derivatives.)

Returning to the system \(F\) given above, let \(L\) be the line field on \(Z\) which is annihilated by \(F\). Integral curves of this line field push down via \(\pi\) to give a 1-parameter family of curves through each point of \(N\), and push down to \(M\) to give a 1-parameter family of geodesics tangent to each contact direction.

Recall that the *prolongation* \([2]\) of a contact manifold \(M\) is the sub-bundle of the projectivization of \(TM\) whose fibre consists of all contact directions at the basepoint in \(M\). For a distribution \(D\) of two-planes on a three-manifold, the prolongation is a \(\mathbb{P}^1\)-bundle \(\mathbb{P}D\) over \(M\). This bundle carries a canonical smooth two-plane distribution \(D'\), defined as follows: For a nonzero vector \(v \in D\), we say that a vector \(w \in T_v\mathbb{P}D\) is tangent to \(D'\) if and only if \(\pi_*w\) is a multiple of \(v\), where \(\pi : \mathbb{P}D \to M\) is the fibration. It follows that \(D'\) is tangent to the fibres of \(\pi\).

For example, let \(D\) be the contact distribution on sub-Riemannian three-manifold \(M\). Then the canonical \(SO(2)\)-structure \(N\) is a double cover of \(\mathbb{P}D\), by the map that sends coframe \((\omega^1, \omega^2, \omega^3) \in T^*M\) to the line in \(T_\pi M\) annihilated by \(\omega^2\) and \(\omega^3\). It follows that \(D'\) lifts to \(N\) to be the two-plane field annihilated by \(\omega^2\) and \(\omega^3\).

The construction of the prolongation may be repeated, resulting each time in a \(\mathbb{P}^1\)-bundle over the previous space, carrying a canonical two-plane distribution. For example, \(Z\) can be embedded as an open subset of the prolongation of the two-plane field on \(N\) (the subset consisting of directions \(v\) such that \(\omega^1(v) \neq 0\)) by sending a point in the fibre \(Z_u\) to the line in \(T_u N\).
annihilated by \( \theta_0, \theta_1 \) and \( \theta_2 \). Under this embedding, the integral curves of the Pfaffian system \( \mathcal{F} \) defined above become tangent to the contact planes on the prolongation of \( N \). Moreover, under the natural lifting of the double cover \( N \to \mathbb{P}D \) to a double cover from \( Z \) to an open subset \( U \subset \mathbb{P}D' \), these integral curves descend to give a well-defined foliation of \( U \).

We will now generalize this situation by throwing away the metric.

**Definition 1.2.** Let \( M^3 \) be a contact manifold and let \( P^5 \) be the second prolongation of \( M \). Let \( L \) be a line field on \( P \) which is tangent to the canonical two-plane distribution \( \mathcal{D}' \) and everywhere transverse to the fibres of \( \rho : P \to M \). Let \( \mathcal{I} \) be the Pfaffian system on \( P \) which annihilates \( L \) and the fibres of \( \rho \), and let \( \mathcal{J} \) be the intersection of the retracting space [1] of \( \mathcal{I} \) with the annihilator of \( L \). Then \((P, L, \rho)\) defines a **contact path geometry** on \( M \) if:

(i) The first derived system \( \mathcal{I}' \) is one-dimensional at each point of \( P \);  
(ii) \( \mathcal{J}' = \mathcal{I} \) at each point of \( P \).

These two conditions need explaining. Because \( L \) is transverse to the fibres, \( \mathcal{I} \) is two-dimensional. If \( \mathcal{I} \) were integrable (i.e., \( \mathcal{I}' = \mathcal{I} \), instead of being one-dimensional) then all paths through a given fibre \( \rho^{-1}(x) \) would project down to a single contact curve on \( M \), so that there would be only one path through \( x \in M \). Condition (i) implies that \( \mathcal{J} \) is three-dimensional at each point; it is automatic that \( \mathcal{I} \subset \mathcal{J}' \). If \( \mathcal{J} \) were integrable, then integral surfaces of \( \mathcal{J} \) would intersect \( \rho^{-1}(x) \) in a 1-parameter family of curves; since each such surface would project down to a single contact curve in \( M \), this would imply that there was only a 1-parameter family of paths through \( x \).

Since \( \mathcal{I}' \) contains the pullback of a contact form on \( M \), condition (i) also implies that \( \mathcal{I}'' = 0 \). Thus, the three-dimensional distribution containing \( L \) and the kernel of \( \rho_* \) is bracket-generating. That in turn guarantees, by Chow’s theorem [3], that two arbitrary points in \( M \) can be connected by a piecewise smooth sequence of paths.

In practice, we will work locally, assuming that \( L \) is defined on an open subset \( U \subset P \). Now the question of which path geometries are sub-Riemannian becomes that of which such line fields \( L \) are locally diffeomorphic to the line field on \( Z \) associated to some sub-Riemannian metric on \( M \), under diffeomorphisms which respect the fibrations and two-plane distributions.

**Proposition 1.3.** Given a contact path geometry we can construct, in a neighbourhood of any point \( q \in P \), a coframe \((\sigma, \theta_0, \theta_1, \theta_2, \theta_3)\) such that:

1. \( v \in TP \) projects down to be a contact direction on \( M \) if and only if \( \theta_0(v) = 0 \)
2. \( \mathcal{I} = \{\theta_0, \theta_1\} \)
3. \( \mathcal{J} = \{\theta_0, \theta_1, \theta_2\} \)
4. \( L^\perp = \{\theta_0, \theta_1, \theta_2, \theta_3\} \).
Moreover, these forms satisfy

\[
(d\theta_i \equiv \theta_{i+1} \wedge \sigma \mod \theta_0, \ldots, \theta_i, \quad 0 \leq i \leq 2)
\]

These will be called 0-adapted coframes for the contact path geometry.

**Proof.** Let \( \rho(q) = x \in M \). On a neighbourhood \( V \) of \( x \), there exists a contact form \( \theta_0 \), and 1-forms \( \sigma, \theta_1 \) such that \( d\theta_0 \equiv \theta_1 \wedge \sigma \mod \theta_0 \). Pull these forms back to \( U = \rho^{-1}(V) \subset P \); we will shrink \( U \) when necessary. Since \( \sigma, \theta_1 \) both restrict to be zero along the fibres of \( \rho \), they cannot be independent modulo \( I \). Therefore we can arrange, by adding multiples of \( \sigma \), that \( \theta_1 \in I \). (Note that now \( \theta_1 \) is no longer the pullback of a form on \( V \).) Since \( \theta_0, \theta_1 \in L^\perp \), then \( \sigma \notin L^\perp \). Since \( \theta_0 \in I' \), then \( d\theta_1 \neq 0 \mod I \).

Since \( \theta_0, \theta_1, \sigma \) span an integrable system, then there will be a smooth 1-form \( \theta_2 \) on \( U \) that \( d\theta_1 \equiv \theta_2 \wedge \sigma \mod I \). (Since \( I' \) is one-dimensional at each point, \( \theta_2 \) is nonzero on \( U \).) By adding multiples of \( \sigma \), we can arrange that \( \theta_2 \in L^\perp \), giving condition 3. Because \( J' \neq J \), there must be a nonzero 1-form \( \theta_3 \) on \( U \) such that \( d\theta_2 \equiv \theta_3 \wedge \sigma \mod J \). We can similarly arrange that \( \theta_3 \in L^\perp \). \( \square \)

**Remark.** The above proposition could also be proved just using the assumption that \( L \) is a line field on a five-manifold \( P \), carrying a two-dimensional Pfaffian system \( I \) annihilating \( L \), such that \( I, J \) satisfy conditions (i, ii) in Defn. 1.2. The contact structure and the submersion to \( M \) can be recovered from \( I' \) and the retracting space \( C(I') \) respectively.

**Corollary 1.4.** In some neighbourhood \( U \) of any given point \( q \in P \), there exist coordinates \( x, y_0, y_1, y_2, y_3 \) such that, for some function \( F \) on \( U \),

\[
\begin{align*}
\sigma &= -dx \\
\theta_0 &= dy_0 - y_1 dx \\
\theta_1 &= dy_1 - y_2 dx \\
\theta_2 &= dy_2 - y_3 dx \\
\theta_3 &= dy_3 - F(x, y_0, y_1, y_2, y_3) dx
\end{align*}
\]

is a 0-adapted coframe. Consequently, paths in \( P \) are locally equivalent to the solutions of the fourth-order ODE

\[
y''' = F(x, y, y', y'', y''').
\]

**Proof.** The structure equations (1) enable us to apply the Goursat normal form theorem [1] to system \( J \). This gives \( J = \{ \theta_0, \theta_1, \theta_2 \} \), in terms of the forms defined here. Since \( I = \{ \theta_0, \theta_1 \} \), then \( dx \notin L^\perp \), and so there exists some function \( F \) such that \( dy_3 - F(x, y_0, y_1, y_2, y_3) dx = 0 \) along the paths in \( U \). \( \square \)
The set of 0-adapted coframes \((\sigma, \theta_0, \theta_1, \theta_2, \theta_3)\) for given contact path geometry forms a principal bundle \(B_0\) over \(P\), with ten-dimensional structure group \(G_0 \subset GL(5, \mathbb{R})\) consisting of matrices of the form
\[
\begin{pmatrix}
a & * & * & 0 & 0 \\
0 & b & 0 & 0 & 0 \\
0 & * & a^{-1}b & 0 & 0 \\
0 & * & * & a^{-2}b & 0 \\
0 & * & * & * & a^{-3}b
\end{pmatrix}
\]
(The stars indicate arbitrary entries.) This is precisely the \(G\)-structure that Fels associates to a fourth-order ODE up to contact transformation (cf. [4], Lemma 3.1). Since the path geometry can be recovered uniquely from the \(G\)-structure, we will treat the two notions as synonymous.

2. Variational and sub-Riemannian path geometries.

The goal of Cartan’s method of equivalence [5] is, for a given \(G\)-structure, to find a sub-bundle, with reduced structure group, on which there exists a unique connection. Like the Levi-Civita connection in Riemannian geometry, this is typically obtained by fixing the value of all or part of the torsion of the connection. Then, invariants may be extracted from the remaining torsion or the curvature of the connection.

We begin with Fels’ result for \(G_0\)-structures of coframes satisfying (1). This gives a reduction of structure to the subgroup \(G_1 \subset G_0\) consisting of matrices of the form
\[
\begin{pmatrix}
a & 0 & 0 & 0 & 0 \\
0 & b & 0 & 0 & 0 \\
0 & 0 & a^{-1}b & 0 & 0 \\
0 & 0 & 0 & a^{-2}b & 0 \\
0 & 0 & 0 & 0 & a^{-3}b
\end{pmatrix}
\cdot
\exp
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & c & 0 & 0 \\
0 & 0 & 2c & 0 & 0 \\
0 & 0 & 0 & c & 0
\end{pmatrix}
\]
In terms of path geometry, the result is:

**Theorem 1** (Fels [4]). Let \(B_0 \setminus P\) define a contact path geometry. Then there is a sub-bundle \(B_1\) with three-dimensional structure group \(G_1\), on which there exists a unique equivariant connection satisfying the following structure equations:

\[(4a) \quad d\sigma = \alpha \wedge \sigma + \theta_0 \wedge (T_1\theta_1 + T_2\theta_2 + T_3\theta_3) + \theta_1 \wedge (T_4\theta_2 + T_5\theta_3)\]
\[(4b) \quad d\theta_0 = \beta \wedge \theta_0 + \sigma \wedge \theta_1\]
\[(4c) \quad d\theta_1 = (\beta - \alpha) \wedge \theta_1 + \gamma \wedge \theta_0 + \sigma \wedge \theta_2\]
\[(4d) \quad d\theta_2 = (\beta - 2\alpha) \wedge \theta_2 + \frac{4}{3} \gamma \wedge \theta_1 + \sigma \wedge \theta_3\]
\[(4e) \quad d\theta_3 = (\beta - 3\alpha) \wedge \theta_3 + \gamma \wedge \theta_2 + \sigma \wedge (I_0\theta_0 + I_1\theta_1) + T_6\theta_0 \wedge \theta_1 + T_7\theta_0 \wedge \theta_2 + T_8\theta_1 \wedge \theta_2.\]
[The one-forms $\alpha, \beta, \gamma$ are connection forms, and $I_0, I_1, T_1, \ldots T_8$ are components of the torsion of the connection.]

Moreover, assuming $P$ is locally defined by a fourth-order ODE (3), solutions of that ODE are critical curves for a second-order Lagrangian if and only if the relative invariants $I_1$ and $T_5$ both vanish identically on $B_1$. In that case, $T_8$ also vanishes.

The essence of Fels’ proof of the second statement is exhibiting a two-form on $B_1$,

$$\omega = m (\theta_0 \wedge \theta_3 - \theta_1 \wedge \theta_2),$$

where $m$ is a nonzero function, such that $\omega$ is closed and $G_1$-invariant. (In fact, $d \log m = 3\alpha - 2\beta$, and, as Fels notes, the structure equations imply that that one-form is closed in the variational case.) It then follows that $\omega$ is the exterior derivative of the Poincaré-Cartan form associated to a Lagrangian on the space of 2-jets.

We will speak of a path geometry for which $I_1, T_5, T_8$ vanish identically as being variational.

**Example 1.** Consider the second-order Lagrangian $\int e^{-3y''} dx$, for which the Euler-Lagrangian equations are, up to multiple,

$$y'''' - 3(y'''^2) = 0.$$  

The coframe (2) gives a section of the bundle $B_0$ defining the corresponding $G_0$-structure on $J^3(\mathbb{R}, \mathbb{R})$. This coframe may be modified to give the following section of the reduced structure $B_1$:

\[
\begin{align*}
\theta_0 &= dy_0 - y_1 dx \\
\theta_1 &= dy_1 - y_2 dx \\
\theta_2 &= dy_2 - y_3 dx - y_3 \theta_1 + \frac{3}{10} y_3^2 \theta_0 \\
\theta_3 &= dy_3 - 3y_3 dy_2 - \frac{3}{10} y_3^2 \theta_1 + \frac{6}{5} y_3^3 \theta_0 \\
\sigma &= dx + \theta_1 - \frac{3}{5} y_3 \theta_0.
\end{align*}
\]

Of course, the torsion satisfies $I_1 = T_5 = T_8 = 0$, but one may also compute\(^4\) that $T_2 = \frac{12}{5} y_3$, $T_3 = \frac{3}{5}$ and $T_4 = -1$ along this section of $B_1$.

---

\(^4\)In order to evaluate the torsion components along a given section of $B_1$, one must determine the values of the connection forms in terms of the given coframe. To do this, begin with the $d\theta_0$ equation (4b), which determines $\beta$ modulo $\theta_0$. One may set $\beta = \beta_0 + b \theta_0$, where $\beta_0$ is any form satisfying (4b) and $b$ is not yet determined. Then (4c) determines $\alpha$ and $\gamma$ modulo $\theta_0, \theta_1$. In fact, one may set

$$\alpha = \alpha_0 + a \theta_0 + z \theta_1$$
$$\gamma = \gamma_0 - a \theta_1 + c \theta_0.$$  

Now (4a) determines $z$ while (4d), (4e) determine $a, b$ and $c$.  

Example 2 (sub-Riemannian geometry). Let $Z$ be the five-manifold of Proposition 1.1. It is easy to verify that the 1-forms given there, when rounded out by $\sigma = \omega^1 - x\omega^3$, form a 0-adapted coframe for the corresponding contact path geometry. We may adapt the coframe to obtain a section of the reduced bundle $B_1 \searrow Z$:

(6) \[
\begin{align*}
\theta_0 &= \omega^3 \\
\theta_1 &= \omega^2 \\
\theta_2 &= \phi - x\omega^1 + A\omega^3 \\
\theta_3 &= dx - a_1\omega^1 - (a_2 + A)\omega^2 + B\omega^3 \\
\sigma &= \omega^1 - \frac{3}{5}x\omega^3
\end{align*}
\]

with

\[
A = \frac{1}{10} \left( a_2 + 3x^2 - 3K \right) \\
B = \frac{1}{10} \left( s_2 - 3k_1 - 6a_1x - 21b_1 \right),
\]

where $K$ is the scalar curvature, and the $b_i$, $s_i$ and $k_i$ are defined on $N$ by

\[
d\alpha \equiv 2a_2\phi + (s_1 + b_2)\omega^1 + (s_2 + b_1)\omega^2 \\
d\beta \equiv 2a_1\phi + (s_2 - b_1)\omega^1 + (b_2 - s_1)\omega^2 \\
dK \equiv k_1\omega^1 + k_2\omega^2 \\
\sigma \equiv U_1\theta_1 + U_2\theta_2 - T_2\theta_3 + T_7\sigma.
\]

Again, one may compute that $I_1 = T_5 = T_8 = 0$, confirming that the path geometry is variational, while $T_2 = 0$, $T_3 = \frac{3}{5}$, and $T_4 = -1$ for this coframe.

The fact that we obtained the same values for $T_3$ and $T_4$ as those from a general second-order Lagrangian hints at further relations among the torsion components. One uncovers one of these by deriving the refined structure equations:

Proposition 2.1. Let $B_1$ be the canonical $G_1$-structure for a variational path geometry. Then there exist functions $U_1, U_2$ on $B_1$ such that the connection forms satisfy

(7) \[
\begin{align*}
d\alpha &= \frac{2}{3}d\beta \\
d\beta &= \sigma \wedge \gamma - \tau \wedge \theta_1 - 3\nu \wedge \theta_0 \\
d\gamma &= \phi \wedge \alpha - \tau \wedge \theta_2 - \nu \wedge \theta_1 \mod \theta_0
\end{align*}
\]

where

\[
\begin{align*}
\tau &= T_1\theta_1 + T_2\theta_2 + T_3\theta_3 \\
\nu &= U_1\theta_1 + U_2\theta_2 - T_2\theta_3 + T_7\sigma.
\end{align*}
\]
The torsion components satisfy \( T_3 = -\frac{2}{3} T_4 \) and

\[
\begin{align*}
(8) \quad dT_1 & \equiv T_1(2\alpha - 2\beta) - \frac{4}{3} T_2 \gamma - 2 U_1 \sigma & \mod \theta_0, \theta_1, \theta_2, \theta_3 \\
(9) \quad dT_2 & \equiv T_2(3\alpha - 2\beta) - \frac{5}{5} T_4 \gamma - (T_1 + 2 U_2) \sigma & \mod \theta_0, \theta_1, \theta_2, \theta_3 \\
(10) \quad dT_4 & \equiv T_4(4\alpha - 2\beta) - \frac{5}{3} T_2 \sigma & \mod \theta_0, \theta_1, \theta_2.
\end{align*}
\]

The above equations indicate that \( T_4 \) is a relative invariant on \( B_1 \), i.e., it varies along the fibres only by scaling. Moreover, they indicate that the quadratic form \( g = \sigma^2 - T_4 \theta_1^2 \) is well-defined, up to multiple and modulo \( \theta_0, \) on \( N \). For, suppose \( v \) is a vector field on \( B_1 \) which is annihilated by \( \sigma, \theta_0, \theta_1, \theta_2 \). Then computing the Lie derivative of \( g \) gives

\[
\mathcal{L}_v(g) = 2\sigma \circ (v \cdot d\sigma) - 2 T_4 \theta_1 \circ (v \cdot d\theta_1) - (v \cdot dT_4) \theta_1^2 \\
\equiv 2 (v \cdot \alpha) \left[ \sigma^2 - T_4 \theta_1^2 \right] \mod \theta_0.
\]

Example 2 shows that, if the contact path geometry comes from a sub-Riemannian metric, then this quadratic form must coincide, up to multiple, with the metric. Matters being so, we will say that a variational geometry is nondegenerate if \( T_4 \neq 0 \) everywhere, and definite if \( T_4 \) is negative everywhere. Assuming the latter is the case, then we may normalize \( T_2 \) and \( T_4 \) to have the same values as in Example 2.

**Proposition 2.2.** Let \( B_1 \) be the canonical \( G_1 \)-structure for a definite variational path geometry. Then there is a sub-bundle \( B_2 \subset B_1 \) on which

\[
T_2 = 0 \quad \text{and} \quad T_4 = -1.
\]

On \( B_2 \) there exist smooth functions \( W_0, W_1, W_2, G_0, G_1, G_2, G_3, \) and \( H \) such that

\[
(11) \quad \beta = 2 \alpha + W_0 \theta_0 + W_1 \theta_1 + W_2 \theta_2 \\
(12) \quad \gamma = H \sigma - 3 (G_0 \theta_0 + G_1 \theta_1 + G_2 \theta_2 + G_3 \theta_3).
\]

**Proof.** Structure equations (9) and (10) show that we may first pass to the sub-bundle where \( T_4 = -1 \) and then move along the fibres in a direction dual to \( \gamma \) to pass to the sub-bundle where \( T_2 = 0 \). Once there, these equations show that \( \beta - 2 \alpha \) and \( \gamma \) restrict to have the above form. Of course, (9) shows that \( \frac{2}{3} H = T_1 + 2 U_2 \).

---

\(^5\)Suppose a variational path structure has \( T_3 \) and \( T_4 \) identically zero; the refined structure equations show that \( T_2 = 0 \) also. Recall that the system which restricts to be zero along the fibres of \( \rho : P \rightarrow M^3 \) is spanned by \( \sigma, \theta_0, \theta_1 \). Since \( d\sigma \equiv T_1 \theta_0 \wedge \theta_1 \mod \sigma \), vectors that are in the kernel of \( \sigma \) push down to give a well-defined plane field on \( M \). These planes intersect the contact planes in a distinguished family of contact directions, which are null lines with respect to \( g \).
On $B_2$, the structure equations (4) take the form

\begin{align}
\sigma & = \alpha \land \sigma + \theta_0 \land (T_1 \theta_1 + \frac{2}{3} \theta_3) - \theta_1 \land \theta_2 \\
\theta_0 & = 2 \alpha \land \theta_0 + \sigma \land \theta_1 \\
\theta_1 & = (\beta - \alpha) \land \theta_1 + \gamma \land \theta_0 + \sigma \land \theta_2 \\
\theta_2 & = (\beta - 2 \alpha) \land \theta_2 + \frac{4}{3} \gamma \land \theta_1 + \sigma \land \theta_3 \\
\theta_3 & = (\beta - 3 \alpha) \land \theta_3 + \gamma \land \theta_2 + 1_0 \sigma \land \theta_0 + T_0 \theta_0 \land \theta_1 + T_7 \theta_0 \land \theta_2
\end{align}

with $\beta$ given by (11).

It’s clear that the fibres of $B_2$ are one-dimensional (with $\alpha$ as the only independent connection form), and the structure group of $B_2$ is simply $\mathbb{R}^x$.

A element $\lambda \neq 0$ of this group acts on sections of $B_2$ by

\[ g_\lambda \cdot (\sigma, \theta_0, \theta_1, \theta_2, \theta_3) = (\lambda \sigma, \lambda^2 \theta_0, \lambda \theta_1, \lambda \theta_2, \lambda^{-1} \theta_3). \]

Structure equations (7) show that $g_\lambda^* \alpha = \alpha$, $g_\lambda^* \beta = \beta$, and $g_\lambda^* (\gamma) = \lambda^{-1} \gamma$.

Then the action on the new torsion is clearly

\[ g_\lambda \cdot (W_0, W_1, W_2) = (\lambda^{-2} W_0, \lambda^{-1} W_1, W_2), \]
\[ g_\lambda \cdot (H, G_0, G_1, G_2, G_3) = (\lambda^{-2} H, \lambda^{-3} G_0, \lambda^{-2} G_1, \lambda^{-1} G_2, G_3). \]

In particular, $W_2$, $G_3$, and the ratios $G_1 : W_0$ and $G_2 : W_1$ are invariant under the scaling action.

We should expect this scaling to be present, since two sub-Riemannian metrics which differ by a constant factor have the same geodesics and hence define the same path geometry. For purposes of constructing a specific metric, we will need to choose a section of $B_2$. Since $3 \alpha - 2 \beta$ is closed, integrals of this one-form comprise a canonical codimension-one foliation of $B_2$ which is transverse to fibres and invariant under the scaling action.

**Definition 2.3.** A section of $B_2$ along which

\[ 3 \alpha - 2 \beta = 0 \]

will be called a **canonical section** of $B_2$, or a **canonical coframe** on $P$. It follows from (11) that

\[ \alpha = -2(W_0 \theta_0 + W_1 \theta_1 + W_2 \theta_2) \]

along a canonical section.

One can check that the coframing constructed in Example 2 is a canonical coframe. Since such coframings are unique up to scale, it follows that if a path geometry comes from a sub-Riemannian metric, then in terms of a canonical coframe that metric must be $g = \sigma^2 + (\theta_1)^2$.

**Proposition 2.4.** Let $P$ be a definite variational path geometry for contact manifold $M^3$ and $(\sigma, \theta_0, \theta_1, \theta_2, \theta_3)$ a canonical coframe on $P$. Then $g = \sigma^2 + (\theta_1)^2$ gives a well-defined metric on the contact planes of $M$ if and only if $W_2$ is identically zero on $P$. 
Proof. Let \( v \) be any vector field on \( P \) tangent to the fibres of the projection \( \rho : P \to M \). Since \( v \) is annihilated by \( \theta_0, \theta_1 \) and \( \sigma \),
\[
\mathcal{L}_v(g) \equiv (v \omega (\beta - \alpha))(\theta_1)^2 \quad \text{mod} \theta_0
\equiv -W_2(v \omega \theta_2)(2(\sigma)^2 + (\theta_1)^2).
\]
\( \square \)

Although the coframe (5) is not a section of \( B_2 \), it can be adjusted so that \( T_2 = 0 \), whereupon we see that \( W_2 \) is nonzero for Example 1.

For the rest of this section we will assume that \( W_2 \) is identically zero. It remains to be seen if the given paths on \( M^3 \)—which are projections of the integral curves of the line field \( L \)—are the geodesics of the sub-Riemannian metric we have constructed. To investigate this further, we will need the torsion identities
\[
G_3 = 0, \quad G_2 = W_1,
\]
which result from computing \( d(d\theta_1) = 0 \) using the structure equations (13) and equations (12), (14) and (15) with \( W_2 = 0 \).

Remark. One might wonder if other identities hold among the remaining torsion coefficients \( G_0, G_1, H, I_0, T_1, T_6, T_7, W_0, W_1 \) as a result of our assumption that \( W_2 = 0 \). However, no further identities arise, and this is proved by showing that the exterior differential system defining a \( G \)-structure satisfying the structure equations on \( B_2 \) with \( W_2 = 0 \) is involutive.

Theorem 2. Let \( P \) be a definite variational path geometry with \( W_2 \) identically zero, and let \( (\sigma, \theta_0, \theta_1, \theta_2, \theta_3) \) be a fixed canonical coframe on \( P \). Then the paths in \( P \) project to be geodesics in \( M \) for the sub-Riemannian metric of Proposition 2.4 if and only if \( G_1 = 2W_0 \) identically on \( P \).

Proof. Let \( N \) be the quotient of \( P \) by the foliation by integral curves of the system \( \mathcal{I}^{(1)} = \{\sigma, \theta_0, \theta_1, \theta_2\} \). Each contact curve in \( M \) has a unique lift to \( N \) as an integral curve of \( \mathcal{I} = \{\theta_0, \theta_1\} \). Clearly, arclength is measured along these lifts by the integral of \( \sigma \) modulo \( \mathcal{I} \). However, the form \( \sigma \) on \( P \) does not descend to be well-defined on \( N \), as shown by
\[
d\sigma = \alpha \wedge \sigma + \theta_0 \wedge (T_1\theta_1 + T_3\theta_3) + T_4\theta_1 \wedge \theta_2
\equiv \frac{3}{2}\theta_0 \wedge \theta_3 \quad \text{mod} \quad \Lambda^2\mathcal{I}^{(1)}.
\]
Computing \( d^2\theta_0 = 0 \) shows that
\begin{equation}
(16) \quad dW_1 \equiv (G_1 - W_0)\sigma + \frac{1}{2}\theta_3 \quad \text{mod} \quad \mathcal{I},
\end{equation}
and this, together with \( d\theta_0 \equiv 0 \mod \Lambda^2\mathcal{I}^{(1)} \), shows that the 1-form
\[
\tilde{\sigma} = \sigma + 3W_1\theta_0
\]
is well-defined on \( N \).
Now arclength with respect to the metric may be measured on the integral curves of $\mathcal{I}$ by the Lagrangian $\int \tilde{\sigma}$. We will apply the Griffiths formalism \([7]\) to investigate which of these are geodesics for $g$. Then, we will try to find conditions under which these curves coincide with the projections of the paths in $P$ under $\pi : P \to N$.

Let $\xi = \tilde{\sigma} + x\theta_0 + y\theta_1$ on $Y = N \times \mathbb{R}^2$. Then one finds that the two-form $d\xi$ is of full rank on $Y$, except where $y = 0$. Accordingly, let $\xi = \tilde{\sigma} + x\theta_0$ on $Z = N \times \mathbb{R}$. Now one computes that

$$d\xi \equiv (dx + (3G_1 - W_0)\sigma) \wedge \theta_0 + (\theta_2 + (x + 5W_1)\sigma) \wedge \theta_1 \mod \theta_0 \wedge \theta_1.$$ \hspace{1cm} (17)

Let $\mathcal{K}$ be the rank four Pfaffian system on $Z$ spanned by the four one-forms on the right in (17):

$$\mathcal{K} = \{\theta_0, \theta_1, \theta_2 + (x + 5W_1)\sigma, dx + (3G_1 - W_0)\sigma\}.$$ 

According to the Griffiths formalism, integral curves of $\mathcal{K}$ project to be extremal curves for $\int \tilde{\sigma}$ on $N$. These coincide with the projections of the paths in $P$ if and only if, in a neighbourhood $U$ of each point of $P$, there is a local diffeomorphism $\varphi : U \to Z$ such that $\varphi^*\mathcal{K}$ coincides with $L^\perp = \{\theta_0, \theta_1, \theta_2, \theta_3\}$, the Pfaffian system on $P$ which defines the paths. (The diffeomorphism would follow from the identification of paths with geodesics on $N$.) The form $\varphi^*(\theta_2 + (x + 5W_1)\sigma)$ belongs in $L^\perp$ if and only if $\varphi^*x = -5W_1$. Then, by (16),

$$\varphi^*(dx + (3G_1 - W_0)\sigma) \equiv (4W_0 - 2G_1)\sigma \mod L^\perp,$$

showing that $\varphi^*\mathcal{K} = L^\perp$ if and only if $G_1 = 2W_0$. \hfill \Box

### 3. More examples.

The results of the previous section may be surprising. For, one could reason that, once a path geometry is known to be variational, it must arise from a second-order Lagrangian, of the form

$$\int L(x, y, y', y'')dx,$$

satisfying the nondegeneracy condition $\partial^2 L/\partial (y'')^2 \neq 0$. Then $L$ is of the form

$$L(x, y, y', y'') = \sqrt{E + FY'' + G(y'')^2},$$

for some functions $E, F, G$ of $x, y, y'$, if and only if $L$, as a function of $y''$, satisfies a certain third-order ODE. In other words, it seems like only one extra condition must be satisfied in order for the Lagrangian to be length with respect to a sub-Riemannian metric. Instead, we find that two scalar conditions (in addition to the Fels variational condition) must hold in order for the metric to be well-defined and in order for its extremals to coincide...
with the given paths. (The reader should note that the above remark about involutivity implies that the condition $G_1 = 2W_0$ is independent from $W_2 = 0$.) It would be interesting to find examples of variational path geometries (equivalently, variational fourth-order ODE) which are nondegenerate, and for which $W_2 = 0$, but the extremals of the associated metric do not coincide with the given paths.

Fels [4] calculates values for relative invariants $I_1$ and $T_5$ for a general fourth-order ODE (3):

$$T_5 = \frac{1}{6} \frac{\partial^3 F}{\partial y_3^3},$$

$$I_1 = \frac{\partial F}{\partial y_1} + \frac{1}{2} \frac{\partial F}{\partial y_2} \frac{\partial F}{\partial y_3} + \frac{1}{8} \left( \frac{\partial F}{\partial y_3} \right)^3 - \frac{d}{dx} \left( \frac{\partial F}{\partial y_2} + \frac{3}{8} \left( \frac{\partial F}{\partial y_3} \right)^2 \right) + \frac{1}{2} \frac{d^2}{dx^2} \frac{\partial F}{\partial y_3}.$$  

(Here, we use the shorthand $y_k$ for $d^k y/dx^k$, and the total derivative $d/dx$ is computed using the ODE.) So, the right-hand side of a variational fourth-order ODE must be at most quadratic in $y_3$. In particular, all linear fourth-order ODEs are variational. However, $T_4 = 0$ for all of these, and so they are degenerate in the sense defined earlier.

Consider the class of nonlinear equations defined by

$$y'''' = f(x, y, y_1, y_2)y_3^2. \tag{18}$$

The Fels conditions $I_1 = T_5 = 0$ imply that such an equation is variational if and only if $f$ does not depend on $y$, and is of the form

$$f(x, y_1, y_2) = g_0(q, r) + x g_1(q, r), \quad q = y_2, \ r = y_1 - x y_2 \tag{19}$$

for functions $g_0$ and $g_1$ that satisfy

$$\frac{\partial g_1}{\partial q} + \frac{\partial g_0}{\partial r} = 0. \tag{20}$$

Using the same adapted coframe as Fels, we calculate that

$$T_4 = -\frac{1}{5}(f^2 + 3f q).$$

To obtain nondegenerate examples, we will confine ourselves to functions $f$ for which $T_4 \neq 0$. As in Proposition 2.2, we modify the coframe so as to get $T_2 = 0$ and $T_4 = -1$. After computing the new connection forms, we find that $W_2 = 0$ if and only if

$$9f_{qq} + 18ff_q + 4f^3 = 0. \tag{21}$$

This equation, when taken together with the partial differential equations for $f$ implied by the $I_1 = 0$, leads to two other second-order equations for $f$ and three third-order equations, forming a Frobenius system whose solutions depend on five constants of integration.

However, for these examples the solutions of the variational fourth-order ODE (18) are never the geodesics of the canonical sub-Riemannian metric.
given by Proposition 2.4. For, with (21) and its consequences taken into account, we find that $W_0$ is independent of $y_3$, but $G_1$ is a cubic polynomial in $y_3$ with leading coefficient $(f^2 + 3f_q)^{3/2}$. Thus, $G_1 = 2W_0$ cannot hold, because of our nondegeneracy condition.

Acknowledgements. I am grateful to Ian Anderson and Mark Fels, and to the referee, for valuable comments and suggestions. Most of the calculations in this paper have been made using Maple V and Maple 6.

References


Received May 31, 2001 and revised October 16, 2001.

DEPT. OF MATHEMATICS
COLLEGE OF CHARLESTON
CHARLESTON SC 29424
E-mail address: ivery@math.cofc.edu
SMALL CONTRACTIONS OF SMOOTH VARIETIES

GIANLUCA OCCHETTA

Let $\varphi : X \to W$ be a proper surjective map from a smooth complex projective variety $X$ to a normal variety $W$; if $\varphi$ has connected fibers and $-K_X$ is $\varphi$-ample, $\varphi$ is called a Fano-Mori contraction; if $\varphi$ is an isomorphism in codimension 2, then $\varphi$ is called a small contraction.

In this paper we study Fano-Mori contractions with fibers covered by large families of rational curves. After some general results we specialize to the case of small contractions, giving a complete description of small contractions of fivefolds with smooth fibers and relatively spanned anticanonical bundle.

1. Introduction.

A contraction $\varphi : X \to W$ is a proper surjective map of normal varieties with connected fibers; if $X$ is a smooth complex projective $n$-dimensional variety, as we will assume throughout the paper, the contraction $\varphi$ is called Fano-Mori, or extremal if the anticanonical divisor $-K_X$ is $\varphi$-ample (for terminology and general properties of these maps see Section 2). Fano-Mori contractions of smooth varieties are defined by linear systems $|m(K_X + rL)|$, with $L$ a $\varphi$-ample line bundle, $m \gg 0$ and $r$ a positive integer; $K_X + rL$ is called a good supporting divisor of the contraction.

An important property of a Fano-Mori contraction is the existence of rational curves in its fibers, as proved by Mori [Mo1], [Mo2] for contractions of smooth varieties, and by Kawamata [Ka1], who extended the result to the log-terminal case. Studying families of rational curves with good properties (the so called unsplit families) on the fibers Ionescu and Wiśniewski proved the fiber locus inequality (Proposition 2.3.6), an inequality which involves the dimensions of the fibers, of the exceptional locus and of the ambient variety and the minimum anticanonical degree of the contracted curves. In particular, if $K_X + rL$ is a good supporting divisor of the contraction and $S'$ is an irreducible component of a fiber of $\varphi$, we have

$$\dim S' \geq r + \operatorname{codim}(E(\varphi)) - 1.$$  

The border case has been studied in [ABW2, 1.1], where the authors proved that the normalization of $S'$ is a projective space; we will deal with the next
case, giving a description of $S'$ when $\dim S' = r + \text{codim}(E(\varphi))$; this is the content of Theorem 4.1.

This theorem is the starting point of the study of small contractions which we develop in the rest of the paper; our main result is Theorem 5.1, which gives a complete description of small contractions supported by $K_X + (n-4)L$ with smooth fibers and such that $-K_X$ is $\varphi$-spanned.

In particular we give a complete description of smooth fibers of small contractions of fivefolds with $-K_X$ $\varphi$-spanned:

\[
\begin{array}{|c|c|}
\hline
F & N_{F/X} \\
\hline
\mathbb{P}^3 & \mathcal{O}_{\mathbb{P}^3}(-1) \oplus \mathcal{O}_{\mathbb{P}^3}(-1) \\
\mathbb{Q}^3 & \mathcal{O}_{\mathbb{Q}^3}(-1) \oplus \mathcal{O}_{\mathbb{Q}^3}(-1) \\
\mathbb{P}^2 & \mathcal{O}_{\mathbb{P}^2} \oplus \mathcal{O}_{\mathbb{P}^2}(-1)^{\oplus 2} \\
\hline
\end{array}
\]

Finally let us mention that, as a preparatory step for the Proof of Theorem 4.1, we prove that smooth fibers of any Fano-Mori contraction are rationally connected (see Corollary 3.6).

2. Background material.

2.1. Fano-Mori contractions. Let $X$ be a smooth complex projective variety; let $N_1(X) = \{1\text{-cycles}\} \otimes \mathbb{R}$, $N^1(X) = \{\text{divisors}\} \otimes \mathbb{R}$, and $\overline{NE}(X) = \{\text{effective 1-cycles}\}$; the last is a closed cone in $N_1(X)$; we also set $\rho(X) = \dim_{\mathbb{R}} N^1(X)$.

Suppose that $K_X$ is not nef, that is there exists an effective curve $C$ such that $K_X.C < 0$; by the Kawamata-Shokurov Contraction Theorem there exist a normal projective variety $Y$ and a surjective morphism $\varphi : X \to Y$ with connected fibers which contracts curves whose numerical class lies in a face in the negative part of the cone $\overline{NE}(X)$ and such that $-K_X$ is $\varphi$-ample.

**Definition 2.1.1** ([AW2, (1.0)]). The map $\varphi$ is called a Fano-Mori contraction (or an extremal contraction); the set $F$ is an extremal face, while a nef Cartier divisor $H$ s.t. $F = H^\perp \cap \overline{NE}(X)$ is a good supporting divisor for the map $\varphi$ (or the face $F$). The contraction is of fiber type if $\dim Y < \dim X$, otherwise it is birational; a birational contraction which is an isomorphism in codimension 2 is called a small contraction. If $\dim_{\mathbb{R}} F = 1$ the face $F$ is called an extremal ray, while $\varphi$ is called an elementary contraction.

**Definition 2.1.2.** Let $\varphi : X \to W$ be a small elementary contraction; the flip of $\varphi$ is a commutative diagram
where $\varphi^+: X^+ \to W$ is a birational morphism from a normal projective variety with only terminal singularities such that the canonical divisor is $\varphi^+$-ample and $X^+$ is isomorphic to $X$ in codimension 1 via $\text{tr}_{\varphi}$.

**Remark 2.1.3.** We have also (see [Mo2]) that if $X$ has an extremal ray $R$ then there exists a rational curve $C$ on $X$ such that $0 < -K_X \cdot C \leq n + 1$ and $R = \mathbb{R}_+[C] := \{ D \in NE(X) : D \equiv \lambda C, \lambda \in \mathbb{R}^+ \}$. A rational curve in $R$ whose intersection number with the anticanonical bundle is minimal is called an extremal curve.

**2.2. Apollonius method.** In the study of Fano-Mori contractions it is often very useful to apply a kind of inductive method, called Apollonius method, slicing the fibers.

**Local setup.** We assume that $\varphi : Y \to W$ is a Fano-Mori contraction of a smooth variety $Y$ onto a normal variety $W$; we choose a fiber $F$ and an open affine subset $Z \subset W$ such that $\varphi(F) \in Z$. Let $X = \varphi^{-1}(Z)$; we will call $\varphi : X \to Z$ a local contraction around $F$.

If $L$ is a $\varphi$-ample line bundle and $r$ a rational number such that $K_X + rL$ is trivial on the fibers of $\varphi$, then $K_X + rL$ is a good supporting divisor for the local contraction $\varphi$.

The idea of the Apollonius method is to slice $X$ in order to obtain a new variety $X'$ with a contraction $\varphi : X' \to Z'$ with smaller fibers, which is easier to study, and then “ascend” the properties of $X'$ to $X$. The first step is given by the following two lemmata:

**Lemma 2.2.1 ([AW1, Lemma 2.6]; (horizontal slicing)).** Suppose that $\varphi : X \to W$ is a local contraction supported by $K_X + rL$, and let $X'$ be a general divisor in the linear system $|L|$. Then, outside of the base locus of $|L|$, the singularities of $X'$ are not worse than those of $X$ and any section of $L$ on $X'$ extends to $X$.

Moreover, if we set $\varphi' := \varphi|_{X'}$ and $L' = L_{X'}$, then $K_{X'} + (r - 1)L'$ is $\varphi'$-trivial. If $r \geq 1 + \epsilon (\dim X - \dim Z)$ then $\varphi'$ is a contraction, i.e., has connected fibers.

**Lemma 2.2.2 ([AW1, Lemma 2.5]; (vertical slicing)).** Assume that $\varphi : X \to W$ is a local contraction supported by $K_X + rL$ and let $X'' \subset X$ be a nontrivial divisor defined by a global function $h \in H^0(X, K_X + rL) = H^0(X, \mathcal{O}_X)$; then, for a general choice of $h$, $X''$ has singularities not worse than those of $X$ and any section of $L$ on $X'$ extends to $X$. 
We will be interested in the study of normal bundles, so it is very important to know how to ascend their properties. The following construction was developed in [AW2]:

Let \( \varphi : X \to Z \) be a Fano-Mori or crepant contraction of a smooth variety, \( L \) an ample line bundle on \( X \), \( F = \varphi^{-1}(z) \) a fiber of \( \varphi \) which is locally complete intersection and \( X' \in |L| \) a normal divisor which does not contain any component of \( F \). Considering the embeddings \( F' = F \cap X' \subset F \subset X \) and \( F' \subset X' \subset X \) we obtain

\[
N_{F'/X} = N_{F'/X'} \oplus L_{F'} = (N_{F/X})_{|F'} \oplus L_F
\]
and therefore \( N_{F'/X'} = (N_{F/X})_{|F'} \); this fact will be used together with the following:

**Lemma 2.2.4** ([AW2, 5.7.2]). If the bundle \( N_{F'/X}^* \) is spanned by functions of \( \Gamma(X', \mathcal{O}_{X'}) \) at a point \( x \in F' \), then the bundle \( N_{F/X}^* \) is spanned at \( x \) by functions of \( \Gamma(X, \mathcal{O}_X) \). If \( N_{F'/X}^* \) is spanned by functions from \( \Gamma(X', \mathcal{O}_{X'}) \) everywhere on \( F' \), then \( N_{F/X}^* \) is nef.

### 2.3. Unsplit families of rational curves.

Throughout this section our main reference is [Ko], with which our notation is coherent.

Let \( X \) be a smooth variety, and \( x \in X \) a point; consider the schemes \( \text{Hom}(\mathbb{P}^1, X) \), parametrizing morphisms from \( \mathbb{P}^1 \) to \( X \), and \( \text{Hom}(\mathbb{P}^1, X; 0 \to x) \), parametrizing morphisms sending \( 0 \in \mathbb{P}^1 \) to \( x \in X \).

Let \( V \subset \text{Hom}(\mathbb{P}^1, X) \) be a closed irreducible subvariety; we will call \( V \) a family of morphisms or, by abuse, a family of rational curves on \( X \).

The image of \( V \) via the restriction of the evaluation morphism \( F : \mathbb{P}^1 \times \text{Hom}(\mathbb{P}^1, X) \to X \) will be denoted by \( \text{Locus}(V) \) and called locus of the family; finally we will denote by \( \text{Locus}(V, 0 \to x) \) the locus of \( V \cap \text{Hom}(\mathbb{P}^1, X; 0 \to x) \), i.e., the locus of the curves in the family which pass through \( x \), and by \( \deg_{-K}(V) \) the anticanonical degree of a general curve in the family.

**Definition 2.3.1.** A family of rational curves is called unsplit if the image of \( V \) in \( \text{Chow}(X) \) is closed.

**Remark 2.3.2** ([Ko, IV.2.2]). Let \( W \subset \text{Chow}(X) \) be the image of \( V \) and \( [C_g] \) be the cycle corresponding to the generic point of \( W \); the points in \( W \setminus W \) correspond to degenerations of \( [C_g] \) into reducible cycles.

**Example 2.3.3.** The family of deformations of an extremal rational curve \( C \) is an unsplit family: If \( C \) degenerates into a reducible cycle, the components must belong to the ray \( R \) generated by \( C \), since \( R \) is extremal; but in \( R \) the curve \( C \) has the minimal intersection with the anticanonical bundle, hence this is impossible.
Now we recall some results which give useful properties of unsplit families that we will use in the rest of the paper.

**Lemma 2.3.4** (see [Wi1, Appendix]). Let $X$ be a smooth variety, $V$ an unsplit family of rational curves on $X$ and fix $x \in \text{Locus}(V)$; every morphism $h : \text{Locus}(V, 0 \to x) \to Z$ is either finite-to-one or takes $\text{Locus}(V, 0 \to x)$ to a point.

**Proposition 2.3.5** ([Ko, IV.2.6]). Let $X$ be a smooth and proper variety and $V$ an unsplit family of rational curves. Then, for $x \in \text{Locus}(V)$:

1) $\dim X + \deg_{-K}(V) \leq \dim \text{Locus}(V) + \dim \text{Locus}(V, 0 \to x) + 1 = \dim V$

2) $\dim X + \deg_{-K}(V) \leq 2 \dim \text{Locus}(V) + 1$

3) $\deg_{-K}(V) \leq \dim \text{Locus}(V, 0 \to x) + 1$.

This last proposition, in case of the unsplit family of deformations of an extremal rational curve, gives the fiber locus inequality:

**Proposition 2.3.6** ([Io], [Wi2]). Let $\varphi$ be a Fano-Mori contraction of $X$ and let $E = E(\varphi)$ be the exceptional locus of $\varphi$ (if $\varphi$ is of fiber type then $E := X$); let $S$ be an irreducible component of a (nontrivial) fiber $F$. Then

$$\dim S + \dim E \geq \dim X + l - 1$$

where

$$l = \min\{-K_X \cdot C : C \text{ is a rational curve in } S\}.$$  

If $\varphi$ is the contraction of a ray $R$, then $l$ is called the length of the ray.

The projective space can be characterized as a variety covered by a large unsplit family of low degree rational curves:

**Theorem 2.3.7** ([ABW2, 1.1]). Let $X$ be a normal projective variety, $L \in \text{Pic}(X)$ and $V$ an unsplit family of rational curves on $X$ such that, for a general $x \in X$, we have $\text{Locus}(V, 0 \to x) = X$ and $L.C = 1$ for a curve $C$ of $V$; then $(X, L) \simeq (\mathbb{P}^n, O_{\mathbb{P}^n}(1))$.

**Theorem 2.3.8** ([Ke, Theorem 3.6]). Let $X$ be a normal projective variety, $L \in \text{Pic}(X)$ and $V$ an unsplit family of rational curves on $X$ such that, for a general $x \in X$, we have $\text{Locus}(V, 0 \to x) = X$ and $L.C = 2$ for a curve $C$ of $V$; then $(X, L) \simeq (\mathbb{P}^n, O_{\mathbb{P}^n}(2))$.

3. Finding transverse rational curves.

In this section we will prove the following:

**Proposition 3.1.** Let $\varphi : X \to W$ be a Fano-Mori contraction of a smooth variety, and let $F$ be a fiber of $\varphi$. Let $\pi : U \to Z$ be a surjective proper morphism from an open dense subset of $F$ onto a quasi projective variety $Z$ of positive dimension. Then for any point $z \in Z$ there exists a rational curve on $F$ which meets $\pi^{-1}(z)$ but is not contracted by $\pi$. 

This result grew out of reading [KoMiMo] via the following:

**Remark 3.2.** Let \( \varphi : X \to W \) be a Fano-Mori contraction of a smooth variety, and let \( F \) be a fiber of \( \varphi \). The numerical class of any curve in \( F \) is contained in the extremal face contracted by \( \varphi \), and the same happens for the deformations of such curves. In particular, the deformations of a curve in the fiber are contained in the exceptional locus, and the pointed deformations of a curve in the fiber are contained in the fiber itself.

For the convenience of the reader, we recall some definitions and results from [KoMiMo]:

**Definition 3.3.** Let \( X, Y \) and \( Z \) be irreducible schemes. Let \( U \subset X \) be an open dense subset, \( \pi : U \to Z \) a morphism, \( f : Y \to X \) with \( f(Y) \) meeting \( U \). By a relative deformation of \( f \) over \( Z \), parametrized by a connected punctured scheme \((S, 0)\) with a base subscheme \( B \subset Y \), we mean a morphism \( F = \{ f_s \} : Y \times S \to X \) which satisfies the following conditions:

1) \( f_0 = f \)
2) \( F_{|B \times S} = (f \circ \text{pr}_Y)_{|B \times S} \)
3) \( \pi f_s = \pi f \) for every \( s \in S \).

If we are considering \( Y \) to be a curve of positive genus, we can replace our original morphism with a new one which has no relative deformations over \( Z \).

**Proposition 3.4 ([KoMiMo, 2.4]).** Assume that there is an open subset \( U \subset X \) such that \( \pi_{|U} \) is a proper morphism over an open subset of \( Z \). Let \( f : Y \to X \) be a morphism of a curve of positive genus. If \( Z \) is not a single point, then there exists a morphism \( f' : Y \to X \) such that the following hold:

1) \( \pi f' = \pi f \)
2) \( \text{Hom}_Z(Y, Z; f', B) \) is a zero-dimensional scheme, when \( B \) is nonempty.

We are going to apply this result to fibers of Fano-Mori contractions using the following:

**Lemma 3.5 ([KoMiMo, Lemma 2.5]).** Let \( \pi : X \to Z \) be a dominant rational map between projective varieties. Let \( H \) and \( D \) be ample divisors on \( X \) and \( Z \), respectively. Then there exists a constant \( \alpha \) which depends only on \( \pi : X \to Z \), \( H \) and \( D \) such that

\[
\deg f^* H \geq \alpha \deg(\pi f)^* D
\]

for any smooth projective curve \( Y \) and any morphism \( f : Y \to X \) whose image meets the domain \( U \) on which \( \pi \) is defined.

**Proof of Proposition 3.1.** The proof follows the line of proof of [KoMiMo, Theorem 2.1], using the fact that, in view of Remark 3.2:
• We can bound the dimension of the deformation space of a curve $C$ in $F$ using $-K_X C$ instead of $-K_F C$.

• $-K_X|_F$ is ample.

**Step I.** Char $\mathbb{K} > 0$.

The divisor $-K_X|_F$ is ample on $F$; fix an ample divisor $D$ on $Z$ and let $\alpha$ be the constant such that

$$\deg f^*(-K_X|_F) \geq \alpha \deg(\pi f)^* D$$

whose existence is assured by Lemma 3.5. Choose a point $z \in Z$ and a smooth curve $Y \subset F$ which intersects $\pi^{-1}(z)$; fix a point $P_0$ in $Y \cap \pi^{-1}(z)$ and choose a Frobenius morphism $f : Y \to Y \subset F$ such that $\deg \pi f > \frac{n}{\alpha} g(Y)$.

By Proposition 3.4 we can replace $f$ by $f' : Y \to X$ such that

$$\deg \pi f' = \deg \pi f > \frac{n}{\alpha} g(Y)$$

and $f'$ has no relative deformations over $Z$ with base point $P = f'^{-1}(P_0)$.

By the first condition we have

$$\chi(Y, \mathcal{I}_{P_0} f'^* T_X) = \deg f'^*(-K_X) - g(Y) \dim X > 0$$

so there exist absolute deformations of $Y_0$ and hence a rational curve through $P_0$; since $f'$ has no relative deformations over $Z$, its deformations induce nontrivial deformations of $\pi f'$ on $Z$, so there exists a rational curve on $F$ through $P_0$ which is mapped to a rational curve on $Z$; then, deforming this curve, we can split it into a union of rational curves of degree $\leq \dim X + 1$, one of which meets $\pi^{-1}(z)$.

**Step II.** Lifting to characteristic zero.

This is a standard construction, since the rational curve we found has bounded degree with respect to the ample divisor $-K_X|_F$.

The existence of transverse rational curves was proved in [KoMiMo] as a step in the proof of rational connectedness of smooth Fano varieties; an analogous result holds for smooth fibers of Fano-Mori contractions, while it fails if we consider singular fibers.

**Corollary 3.6.** Let $\varphi : X \to W$ be a Fano-Mori contraction of a smooth variety and $F$ a smooth fiber of $\varphi$. Then $F$ is rationally connected.

**Proof.** Consider the maximal rationally chain connected fibration $F \to Z$ (see [Ko, Theorem 5.2]); if $Z$ is not a point there exists a rational curve passing through a point of a very general fiber of $\varphi$ which is not contracted by $\varphi$ and this is a contradiction, by the maximality of $\varphi$ (see [Ko, 5.2.1]). Thus $Z$ is a point and $F$ is rationally chain connected, but, in characteristic zero, rational chain connectedness is equivalent to rational connectedness. $\square$
Corollary 3.7. Let $\varphi : X \to W$ be a Fano-Mori contraction of a smooth variety and $F$ a smooth fiber of $\varphi$. Then

$$H^i(F, O_F) = 0 \quad \text{for } i \geq 1.$$  

Example 3.8. A singular fiber of a Fano-Mori contraction which is not rationally connected (suggested by Jaroslaw Wiśniewski).

Let $W$ be the projectivization of the linear system $|O_{P^3}(3)|$, $X \subset P^3 \times W$ the incidence variety and $p$ the projection onto $W$; it is easy to check that $-K_X$ is $p$-ample (it restricts to $O(1)$ on the fibers of $p$), so that $p$ is a Fano-Mori contraction; among its fibers there is the cubic singular surface given by the cone over an elliptic curve, which is not rationally connected.

4. Deformations of curves and fibers of Fano-Mori contractions.

In this section we will prove the following:

Theorem 4.1. Let $\varphi : X \to W$ be a Fano-Mori contraction of a smooth variety, supported by $K_X + rL$. Let $S'$ be any component of a nontrivial fiber $F = \varphi^{-1}(w)$, let $S$ be its normalization and denote again by $L$ the pullback of $L$ to $S$. If

$$\dim S = r + \text{codim } (E(\varphi)),$$

then either the Fujita $\Delta$-genus of $(S, L)$ is zero or $S$ is singular and has a desingularization which is a $P^{s-1}$-bundle over a smooth curve.

Moreover, if $S'$ has rational singularities then only the first possibility can occur.

Proof. Let $C \subset S$ be an extremal rational curve, $V$ the unsplit family of deformations of $C$ and $x$ a general point in Locus($V$); we have

$$n + r = \dim E(\varphi) + \dim S$$

$$\geq \dim \text{Locus}(V) + \dim \text{Locus}(V, 0 \to x) \geq n + r - 1,$$

whence Locus($V$) = $E(\varphi)$ and dim Locus($V, 0 \to x$) = $s$, $s - 1$.

Let $S$ be the normalization of $S'$; normalizing also the family and its graph we obtain a new unspli family $V$ on $S$.

If deg$_{-K}(V) = r$ then the curves in this family are lines with respect to $L$, otherwise we have $r = 1$ and deg$_{-K}(V) = 2$; in this last case, applying Theorem 2.3.8 we get $(S, L) \simeq (P^s, O_{P^2}(2))$.

Let $\theta : M \to S$ be a desingularization of $S$, $L_M = \theta^*L$, and $m$ the inverse image of a smooth point $e$ of $S$. The curves of $V$ passing through $e$, since they are not entirely contained in the singular locus of $S$, can be lifted to curves on $M$ passing through the point $m$.

We claim that there is a family $V_M$ of these lifted up curves which has degree one with respect to $L_M$, and covers a dense subset of $M$.
Indeed we know that a neighbourhood of $m$ is covered by these curves, so, if such a family did not exist, the neighbourhood would be contained in a countable sum of nowhere dense subsets of $M$, obtained by deforming each of the lifted up curves, a contradiction.

By construction, for the family $\mathcal{V}_M$ we have
\[
\dim \text{Locus}(\mathcal{V}_M) = s
\]
\[
\dim \text{Locus}(\mathcal{V}_M, 0 \to m) \geq s - 1.
\]

Let $ev : \mathcal{V}_M \times \mathbb{P}^1 \to M$ be the evaluation map and let $f : \mathbb{P}^1 \to M$ be a general member of the family; we have (cf. [Ko, II.3])
\[
\begin{align*}
T_{[f,p]}(\mathcal{V}_M \times \mathbb{P}^1) & \xrightarrow{T(ev)} T_{f(p)}M \\
H^0(\mathbb{P}^1, f^*TM) \times T_{\mathbb{P}^1} & \xrightarrow{ev_p + T_p f} (f^*TM)_p
\end{align*}
\]
and
\[
\begin{align*}
T_{[f,p]}((\mathcal{V}_M, 0 \to m) \times \mathbb{P}^1) & \xrightarrow{T(ev)} T_{f(p)}M \\
H^0(\mathbb{P}^1, f^*TM \otimes I_m) \times T_{\mathbb{P}^1} & \xrightarrow{ev_p + T_p f} (f^*TM)_p.
\end{align*}
\]

In both cases we have $\text{rk}(T(ev)) = \text{rk}(ev_p)$ so, by the information on the dimensions of the loci, we deduce
\[
f^*(TM) \cong \oplus \mathcal{O}_{\mathbb{P}^1}(a_i)
\]
with $a_1 \geq 2$, $a_i \geq 1$ for $i = 2, \ldots, s - 1$, $a_s \geq 0$, implying
\[
(K_M + tL_M).C < 0 \quad \text{if} \quad t < s.
\]

So, since the family covers a dense subset of $M$, we have the following vanishing results:
\[
\begin{align*}
(4.2) \quad h^s(M, -tL_M) = h^0(M, K_M + tL_M) = 0 & \quad \text{for} \quad t \leq s - 1 \\
(4.3) \quad h^i(M, \mathcal{O}_M) = h^0(M, \wedge^i TM^*) = 0 & \quad \text{for} \quad i \geq 2.
\end{align*}
\]

\textbf{Case 1.} $h^1(M, \mathcal{O}_M) = 0$.

By the Kawamata-Viehweg vanishing theorem we have
\[
(4.4) \quad h^i(M, -tL_M) = 0 \quad \text{for} \quad i < s, \quad t > 0.
\]

Combining (4.2), (4.3) and (4.4) we see that the Hilbert polynomial of $L_M$, $\chi(t) = \chi(M, tL_M)$, vanishes for $t = -1, -2, \ldots, -s + 1$, and $\chi(M, \mathcal{O}_M) = 1$,
so we can write the polynomial in the form

$$
\chi(M, tL) = \frac{d}{s!} \left( \prod_{k=1}^{s-1} (t + k) \right) (t + \frac{s}{d}).
$$

This implies that $h^0(M, L) = \chi(1) = s + d$ and then, computing the $\Delta$-genus of the quasi-polarized variety $(M, L)$,

$$
\Delta(M, L) = \dim M + \deg L - h^0(M, L) = s + d - (s + d) = 0.
$$

Using [Fu, Theorem 1.1] we get a commutative diagram

\[
\begin{array}{ccc}
\overline{M} & \xrightarrow{\phi} & F \\
h \downarrow & & \downarrow \\
M & \xrightarrow{\theta} & \theta(M)
\end{array}
\]

where $\overline{M}$ is a smooth variety and $L\overline{M}$ is a very ample line bundle on $\overline{M}$ such that $(\overline{M}, L\overline{M})$ has $\Delta$-genus zero, so this is the case also for $(S, L_S)$.

**Case 2.** $h^1(M, \mathcal{O}_M) \neq 0$.

Suppose that $H^1(M, \mathcal{O}_M)$ does not vanish and consider the Albanese map of $M$, $\alpha : M \to \text{Alb}(M)$; since we know that $\text{Locus}(V_M) = M$ and that $\dim \text{Locus}(V_M, 0 \to x) \geq s - 1$, the image of the Albanese map of $M$, $\alpha(M) \subset \text{Alb}(M)$ must be a nonsingular curve $B$.

Let $G$ be the generic fiber of $\alpha$; $G$ is a quasi polarized variety of dimension $s - 1$ covered by a family of rational curves whose dimension at the generic point is $s - 1$. We have

$$
h^{s-1}(G, -tL_G) = h^0(f, K_G + tL_G) = 0 \quad \text{for} \quad t \leq s - 1.
$$

So, since $G$ is a desingularization of $\tilde{G}$, the normalization of $\theta(G)$, by [Fu, Theorem 2.2] we have that $\tilde{G}$ is a projective space $\mathbb{P}^{s-1}$.

Consider the map $\Theta : M \to S \times B$ induced by $\theta$, and the image $\Theta(M)$ of $M$, with the maps $p_1 : \Theta(M) \to S$ and $p_2 : \Theta(M) \to B$ induced by the projections of $S \times B$ on the factors:

\[
\begin{array}{ccc}
S & \xrightarrow{\theta} & M & \xrightarrow{\alpha} & B \\
\downarrow & & \downarrow & & \downarrow \\
\Theta & \downarrow & \Theta(M) & \downarrow & \Theta(M)
\end{array}
\]
Let $\tilde{M}$ be the normalization of $\Theta(M)$; we have the following commutative diagram:

\[\begin{array}{ccc}
S & \xrightarrow{\theta} & M \\
\downarrow & & \downarrow \\
\tilde{M} & \xrightarrow{\tilde{p}_2} & B
\end{array}\]

The generic fiber of $\tilde{p}_2$, which is a normal variety, is $\mathbb{P}^{s-1}$ since $\tilde{G}$ is so: The line bundle $L_{\tilde{M}} = \tilde{p}_1^*L$ is ample on the fibers of $\tilde{p}_2$, so, up to replace it with $L'_{\tilde{M}} = L_{\tilde{M}} + \tilde{p}_2^*H$ where $H$ is ample on $B$, we can assume it is ample; then we argue as in [BS, Proposition 3.2.1] to obtain that $(\tilde{M}, L_{\tilde{M}})$ is a $\mathbb{P}^{s-1}$-bundle over $B$; note that we can disregard the assumptions on the singularities of $\tilde{M}$ since $\tilde{p}_2$ is flat.

The last assertion follows from the fact that, in case of rational singularities, we can consider the Albanese map of $S'$ itself and obtain that, in Case 2 of the Proof, $S'$ is a $\mathbb{P}^{s-1}$-bundle over a smooth curve $B$, which has to be rational by Corollary 3.7. \hfill \Box

**Remark 4.5.** In the case of elementary divisorial contractions it is possible to replace the assumption $\dim S = r + \text{codim } (E(\varphi))$ with $\dim S = l(\varphi) + \text{codim } (E(\varphi))$, avoiding any assumption on $L$; in fact one can prove that $K_X - rE(\varphi)$ is a good supporting divisor for $\varphi$ (see [AO1]).

**5. Small contractions.**

In the second part of the paper, keeping Theorem 4.1 as a starting point, we will deal with small contractions. It is well-known that (locally) the only small contraction of smooth varieties of dimension four is the Kawamata small contraction [Ka2], i.e., the contraction of a projective plane with normal bundle $\mathcal{O}_{\mathbb{P}^2}(-1)^{\oplus 2}$; this contraction is supported by $K_X + L$, and one natural generalization of this situation is the study of contractions supported by $K_X + (n - 3)L$ on a smooth $n$-fold; this was done in [ABW2].

The next step is the study of small contractions of smooth varieties of dimension five; this contractions can be supported either by $K_X + 2L$, and this situation falls in the case treated in [ABW2], or by $K_X + L$; this (or more generally the case of small contractions of $n$-folds supported by $K_X + (n - 4)L$) will be the object of our study. Our results are summarized in the following:

**Theorem 5.1.** Let $K_X + (n - 4)L$ be a good supporting divisor of an elementary small contraction $\varphi : X \to Z$, let $F$ be a smooth fiber of $\varphi$ and
\(N = N_{F/X}\) its normal bundle. If \(L\) is \(\varphi\)-spanned then one of the following possibilities occurs:

<table>
<thead>
<tr>
<th>(n)</th>
<th>(F)</th>
<th>(N_{F/X})</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>(\mathbb{P}^3)</td>
<td>(\mathcal{O}<em>{\mathbb{P}^3}(-1) \oplus \mathcal{O}</em>{\mathbb{P}^3}(-1))</td>
</tr>
<tr>
<td>(\geq 5)</td>
<td>(\mathbb{P}^{n-2})</td>
<td>(\mathcal{O}<em>{\mathbb{P}}(-2) \oplus \mathcal{O}</em>{\mathbb{P}}(-1))</td>
</tr>
<tr>
<td>(\geq 5)</td>
<td>(\mathbb{P}^{n-2})</td>
<td>(\mathcal{O}<em>{\mathbb{Q}}(-1) \oplus \mathcal{O}</em>{\mathbb{Q}}(-1))</td>
</tr>
<tr>
<td>(\geq 5)</td>
<td>(\mathbb{P}^{n-3})</td>
<td>(\mathcal{O}<em>{\mathbb{P}} \oplus \mathcal{O}</em>{\mathbb{P}}(-1)^{\oplus 2})</td>
</tr>
</tbody>
</table>

Moreover, the formal neighborhood of \(F\) in \(X\) is determined uniquely and it is the same as the formal neighborhood of the zero section of the total space of the bundle \(N_{F/X}^*\), and the flip of \(\varphi\) exists.

The rest of the section is devoted to the proof of this theorem. We will proceed in several steps.

5.1. Geometric fiber. First of all we present another deformation argument which will help us to refine the description of the fibers given in Theorem 4.1 in the case of small contractions:

**Proposition 5.1.1.** Let \(\varphi : X \to Z\) be a Fano-Mori elementary small contraction of a smooth variety supported by \(K_X + (n-4)L\), and let \(F\) be a fiber of \(\varphi\). Then \(F\) cannot be a \(\mathbb{P}\)-bundle over a rational curve.

*Proof.* By Proposition 2.3.6 the dimension of \(F\) is \(n-2\) or \(n-3\); in the latter case the normalization of \(F\) is a projective space by [ABW2, 1.1], so we have to deal only with the case \(\dim F = n-2\).

Suppose that \(F\) is a \(\mathbb{P}^{n-3}\)-bundle over a rational curve:

\[
F = \mathbb{P}_{\mathbb{P}^1}(\mathcal{E}) = \mathbb{P}_{\mathbb{P}^1}(\mathcal{O}(a_1) \oplus \mathcal{O}(a_2) \oplus \cdots \oplus \mathcal{O}(a_{n-2}))
\]

with \(a_1 \leq a_2 \leq \cdots \leq a_{n-2}\).

Consider the section \(C \subset F\) corresponding to the surjection \(\mathcal{E} \longrightarrow \mathcal{O}_{\mathbb{P}^1}(a_1)\); since \(L.C_0 = 1\), the restriction of \(L\) to \(F\) is a twist of \(\xi_\mathcal{E}\) and so the intersection number \(d = L.C\) is minimal among horizontal sections; in \(N_1(X)\), since the contraction is elementary, we have \(C \equiv dC_0\).

Consider the family \(\mathcal{C}\) of deformations of \(C\) inside \(X\); by Remark 3.2 these deformations are contained in \(F\), and so they cannot break: Suppose, by contradiction, that a deformation \(\tilde{C}\) of \(C\) is reducible; let \(\tilde{C} = \sum_i \tilde{C}_i\) and let \(f\) be a fiber of the bundle projection \(p\). Since \(\tilde{C}.f = 1\) exactly one of the \(\tilde{C}_i\) must be horizontal, and the others contained in the fibers, but, by the minimality of \(L.C\) among horizontal sections, this is impossible, hence the family \(\mathcal{C}\) is an unsplit family. By Proposition 2.3.5

\[
\dim \text{Locus}(\mathcal{C}, 0 \to c) \geq \dim X - \dim \text{Locus}(\mathcal{C}) + \deg_{-K}(\mathcal{C}) - 1 \\
\geq 2 + d(n-4) - 1 \geq n - 3
\]
then, using Lemma 2.3.4 and Serre’s inequality
\[ \dim(\text{Locus}(C, 0 \to c) \cap f) \geq \dim \text{Locus}(C, 0 \to c) + \dim f - \dim F \]
\[ \geq 2(n - 3) - (n - 2) = n - 4, \]
a contradiction, since \( n \geq 5 \).

We can thus give a complete (and effective) description of smooth fibers:

**Proposition 5.1.2.** Let \( \varphi : X \to Z \) be a small elementary contraction of a smooth variety, supported by \( K_X + (n - 4)L \), and let \( F \) be a smooth fiber of \( \varphi \). Then \( F \) is either \( \mathbb{P}^{n-2} \), \( \mathbb{Q}^{n-2} \), or \( \mathbb{P}^{n-3} \).

**Proof.** Let \( E \) be an irreducible component of the exceptional locus of \( \varphi \) containing \( F \); the fiber locus inequality (Proposition 2.3.6) combined with our hypothesis gives
\[ 2n - 4 \geq \dim E + \dim F \geq n + (n - 4) - 1. \]
If \( \dim F = n - 2 \) then \( E = F \); the hypothesis of Theorem 4.1 are satisfied, the case of the \( \mathbb{P} \)-bundle over a rational curve is ruled out by Proposition 5.1.1. If \( \dim F = n - 3 \) then \( F \simeq \mathbb{P}^{n-3} \) by [ABW2, 1.1].

From now on we will restrict our study to the case \( n = 5 \); the general case will then follow by horizontal slicing (see 5.6).

**5.2. Scheme theoretic fiber.** Let \( \varphi : Y \to W \) be a small contraction of a smooth fivefold; we choose a smooth fiber \( F = \varphi^{-1}(w) \) and consider a local contraction \( \varphi : X \to Z \) around \( F \); by Proposition 5.1.2 we know the geometric structure of \( F \), i.e., the set theoretic pre-image \( \varphi^{-1}(w) \) equipped with the reduced scheme structure, which is one of the following: \( \mathbb{P}^3 \), \( \mathbb{Q}^3 \), or \( \mathbb{P}^2 \). The next step in understanding the map is the study of the conormal bundle of \( F \) in \( X \), defined as the quotient \( \mathcal{I}_F/\mathcal{I}_F^2 \), where \( \mathcal{I}_F \) is the ideal of \( F \) with the reduced structure. We first establish some cohomological conditions which the normal bundle has to satisfy.

**Lemma 5.2.1 ([AW2, 1.2.2]).** Let \( \varphi : X \to Z \) be a Fano-Mori or crepant birational contraction, \( z \in Z \), \( F = \varphi^{-1}(z) \), and \( F' \) any subscheme whose support is contained in \( F \). Then
\[ H^r(F', \mathcal{O}_{F'}) = 0 \quad \text{for} \quad r \geq \dim F. \]

**Corollary 5.2.2.** Let \( F \) be a \( s \)-dimensional fiber of a Fano-Mori or crepant contraction such that \( s > 1 \) and \( H^{s-1}(F, \mathcal{O}_F) = 0 \) and let \( N^* \) be its conormal bundle. Then
\[ H^s(F, N^*) = 0. \]
Proof. Let \( J \) be the ideal of \( F \) in \( X \) and consider the exact sequence
\[
0 \to J / J^2 = N^* \to \mathcal{O}_X / J^2 \to \mathcal{O}_X / J \to 0.
\]
Our claim follows by the long exact sequence
\[
\ldots \to H^{s-1}(F, \mathcal{O}_F) \to H^s(F, N^*) \to H^s(F, \mathcal{O}_X / J^2) \to \ldots
\]
and Lemma 5.2.1.

Another cohomological condition holds for normal bundles of fibers of dimension three.

Lemma 5.2.3. Let \( F \) be a 3-dimensional smooth fiber of a small contraction of a smooth variety of dimension five, and let \( N := N_{F/X} \) be its normal bundle. Then \( h^0(N) - h^1(N) \leq 0 \).

Proof. By Proposition 2.3.6 in this case \( F \) coincides with the irreducible component of the exceptional locus in which it is contained, and so \( F \) cannot move in \( X \); since the dimension of its deformation space is greater or equal than \( h^0(N) - h^1(N) \), the lemma follows.

As a last preparatory step we compute the first Chern class of the conormal bundles.

Lemma 5.2.4. Let \( \varphi : X \to Z \) be a local small contraction of a smooth variety of dimension five around a smooth fiber \( F \), and let \( N^* := N_{F/X}^* \) be its conormal bundle; the possibilities for the first Chern class of \( N^* \) are given in the following table, where \( l(\varphi) \) denotes the length of \( \varphi \):

<table>
<thead>
<tr>
<th>( F )</th>
<th>( l(\varphi) )</th>
<th>( c_1(N^*) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathbb{P}^3 )</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>( \mathbb{Q}^3 )</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>( \mathbb{P}^2 )</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Proof. The result follows from Proposition 5.1.2 observing that, by Proposition 2.3.6,
\[
6 \geq 2 \dim E \geq 5 + l(\varphi) - 1.
\]
We have \( l(\varphi) \leq 2 \), and equality holds if \( l(\varphi) = 2 \); in this case by [ABW2, Lemma 1.1] we have \( F \simeq \mathbb{P}^3 \). Then we apply the adjunction formula.
5.3. Nefness of the conormal bundle.

**Definition 5.3.1.** Let \( \varphi : X \to Z \) be a local contraction of a smooth variety around a smooth fiber \( F \) and let \( L \) be a \( \varphi \)-ample line bundle; we say that \( L \) is \( \varphi \)-spanned if the evaluation morphism \( \varphi^* \varphi_* L \to L \) is surjective at every point of \( F \).

We will now prove that the relative spannedness of the anticanonical bundle is a sufficient condition for the nefness of \( N_{F/X}^* \).

**Proposition 5.3.2.** Let \( \varphi : X \to Z \) be a local small contraction of a smooth variety of dimension five around a smooth fiber \( F \); if \( -K_X \) is \( \varphi \)-spanned then \( N_{F/X}^* \) is nef.

**Proof.** We divide our proof into four cases, according to the possibilities given by Lemma 5.2.4.

- **Case 1.** \( F \simeq \mathbb{P}^3 \) and \( c_1(N^*) = 3 \).

  Since \( (-K_X)_{|\mathbb{P}^3} \simeq \mathcal{O}_{\mathbb{P}^3}(1) \) and \( -K_X \) is \( \varphi \)-spanned, \( -K_X \) is also \( \varphi \)-very ample; this implies that, for \( l \) a line in \( F \), \( N_{l/X}^* \) is spanned by global sections ([AW2, Lemma 1.3.5]).

  Moreover we can find two sections \( H_1, H_2 \in |-K_X| \) such that \( l = F \cap H_1 \cap H_2 \), so that by formula (2.2.3)

  \[ (N_{F/X}^*)_{|l} \oplus (K_X)_{|l}^\oplus_2 = N_{l/X}^* \],

  allowing for \( N_{F/X}^* \) the following splitting types on lines: (4, -1), (3, 0) and (2, 1).

  If the general splitting of \( N_{F/X}^* \) were (4, -1), then \( N_{F/X}^* \simeq \mathcal{O}_{\mathbb{P}^3}(4) \oplus \mathcal{O}_{\mathbb{P}^3}(-1) \), but, in this case we would have \( h^0(N_{F/X}^*) - h^1(N_{F/X}^*) > 0 \) and this is impossible by Lemma 5.2.3.

  The restriction of \( N_{F/X}^* \) to the generic line is thus nef, hence spanned; applying Lemma 2.2.4 twice we have that \( N_{F/X}^* \) is generically spanned on \( F \), hence nef.

- **Case 2.** \( F \simeq \mathbb{P}^3 \) and \( c_1(N^*) = 2 \).

  In this case we have \( (-K_X)_F \simeq \mathcal{O}_{\mathbb{P}^3}(2) \).

  **Claim.** Either there exists a line bundle \( L \) on \( X \) such that \( L_F \simeq \mathcal{O}_{\mathbb{P}^3}(1) \), or \( N^* = \mathcal{O}_{\mathbb{P}^3}(1)^\oplus_2 \).

  **Proof of the claim.** Let \( A \) be a very ample divisor on \( Z \) and \( H = \varphi^* A \); for \( m \) large enough the linear system \( |mH + K_X| \) is nonempty, so there exists an effective divisor \( D' \) such that \( D'.l < 0 \) for \( l \) a line in \( F \); in particular there exists an irreducible reduced divisor \( D \) s.t. \( D.l < 0 \).

  If \( D.l \) is odd, then a combination of \( D \) and \( -K_X \) gives us a line bundle \( L \) such that \( L_F \simeq \mathcal{O}_{\mathbb{P}^3}(1) \).
So suppose that $D.l = -2t$ and consider $\beta : \hat{X} \to X$, the blow up $X$ along $F$ with exceptional divisor $\hat{F} = \mathbb{P}(N_{F/X}^*)$ and the strict transform of $D$, $\overline{D} = \beta^*D - 2tE$.

Let $l$ be a line in $\mathbb{P}^3$; an easy computation shows that the restriction of $\overline{D}$ to $\mathbb{P}_l(N_{F/X}^*) \simeq \mathbb{F}_e$ has a component corresponding to $t$-times the fundamental section (which has to be contained in $\overline{D}$ since it has negative intersection with it); hence there exist divisors $D_1, D_2$ in $| - tE + p^*\mathcal{O}_{\mathbb{P}^3}(a)|$, $| - tE + p^*\mathcal{O}_{\mathbb{P}^3}(b)|$, with $a + b = -2$ (we can possibly have $D_1 = D_2$).

Another straightforward computation on the ruled surfaces $\mathbb{P}_l(N_{F/X}^*)$ shows that we have $e = b - a$ for every line. It follows that the splitting type of $N^*$ is constant on lines, hence $N^*$ is decomposable and so, by Lemma 5.2.3 the only possibility is $N^* = \mathcal{O}_{\mathbb{P}^3}(1)^{\oplus 2}$. 

So we can suppose that there exists a line bundle $\mathcal{L}$ on $X$ such that $\mathcal{L}_F \simeq \mathcal{O}_{\mathbb{P}^3}(1)$; up to twist $\mathcal{L}$ with $\varphi^*A$ with $A$ very ample on $Z$ we can assume that $\mathcal{L}$ is ample. Let $X'$ be a generic divisor in $|\mathcal{L}|$ and $F' = F \cap X'$; by Lemma 2.2.1 $X'$ is smooth and $K_{X'} + \mathcal{L}_{X'}$ is a good supporting divisor of a small contraction of a smooth fourfold. By [Ka2] $F' \simeq \mathbb{P}^2$ and $N_{F'/X'} \simeq \oplus^2 \mathcal{O}_{\mathbb{P}^2}(-1)$; since by formula (2.2.3)

$$\left( N_{F/X} |_{F'} \right) = N_{F'/X'},$$

we have that $N_{F/X} \simeq \oplus^2 \mathcal{O}_{\mathbb{P}^3}(-1)$.

Case 3. $E \simeq \mathbb{P}^3$.

Let $X'$ be a general divisor in $| - K_X |$, $F' = F \cap X' \simeq \mathbb{Q}^2$ and $N := N_{F'/X'}$. By [AW2, Lemma 2.10.1], if the generic splitting type of $N$ on conics is not $(-2, -2)$ then there exist a zero dimensional subscheme $T \subset F'$ and an exact sequence:

$$0 \longrightarrow \mathcal{O}_{\mathbb{Q}^2}(a_1, a_2) \longrightarrow N \longrightarrow \mathcal{T}_T \otimes \mathcal{O}_{\mathbb{Q}^2}(b_1, b_2) \longrightarrow 0.$$ 

Twisting our sequence by $\mathcal{O}_{\mathbb{Q}^2}(-2)$ we have:

$$0 \longrightarrow \mathcal{O}_{\mathbb{Q}^2}(a_1 - 2, a_2 - 2) \longrightarrow N(-2) \longrightarrow \mathcal{T}_T \otimes \mathcal{O}_{\mathbb{Q}^2}(b_1 - 2, b_2 - 2) \longrightarrow 0.$$ 

By Serre duality and Lemma 5.2.1 we have

$$H^0(\mathbb{Q}^2, N(-2)) = H^0(\mathbb{Q}^2, K_{\mathbb{Q}^2} \otimes N) \simeq H^2(\mathbb{Q}^2, N^*) = 0;$$

since $H^0(\mathbb{Q}^2, N(-2)) = 0$, by the long cohomology exact sequence, we have $H^0(\mathbb{Q}^2, \mathcal{O}(a_1 - 2, a_2 - 2)) = 0$, yielding $a_1, a_2 \leq 1$ (and $b_1, b_2 \geq -3$).

Observe that, since $-K_X$ is $\varphi$-spanned and $-K_X|_{F'} \simeq \mathcal{O}_{\mathbb{Q}^2}(1)$, either $-K_X$ is $\varphi$-very ample or the image of the vector space $H^0(X', -K_X|_{X'})$ in the space $H^0(\mathbb{Q}^2, \mathcal{O}_{\mathbb{Q}^2}(1))$ corresponds to a double cover $\pi : \mathbb{Q}^2 \to \mathbb{P}^2$.

In both cases it is possible to find a section $X'' \in |(-K_X)|_{X'}|$ such that $F'' = X' \cap X''$ is a smooth conic passing through two points of $T$; the
splitting type of $N$ on such a conic would then be $(b - 4, a + 4)$ and $X''$ would contain an exceptional curve with normal bundle $(b - 4, a + 4)$; by [Na] this implies that either $(a, b) = (-1, -3), (-2, -2)$ or $T$ consists of a single point, but in this case we would have $h^0(N) - h^1(N) > 0$, and this is impossible by Lemma 5.2.3.

It follows that on the generic conic $N$ has splitting type $(-1, -3)$ or $(-2, -2)$, hence $N^*$ is spanned on the generic conic and we can apply Lemma 2.2.4.

Case 4. $E \simeq \mathbb{P}^2$.

Let $X' \in |-K_X|$ be any divisor, and $l = F \cap X'$; the first cohomology group of the conormal bundle of $l$ in $X'$ vanishes [AW2, Proposition 5.6.1], so, since by formula (2.2.3)

$$(N^*_{F/X})_l = N^*_{l/X'},$$

the possible splitting types of $N^*$ are $(1, -3), (2, 0)$, and $(1, 1)$; if the general splitting type is different from $(1, -3)$ then by Lemma 2.2.4 $N^*$ is nef and we are done. If the general splitting of $N^*$ is $(1, -3)$ then $N^* \simeq \mathcal{O}(1) \oplus \mathcal{O}(-3)$, but this is impossible by Lemma 5.2.3. \hfill $\square$

**Remark 5.3.3.** By results of Mella [Me] we know that, for the contractions that we are studying, the anticanonical bundle has finite base locus on $F$; we believe that, in the set up of Proposition 5.3.2 it should be true that the anticanonical bundle is always $\varphi$-spanned.

### 5.4. Description of the conormal bundle.

The nefness of the conormal bundles implies that they are Fano bundles; this fact allows us to give a description of them:

**Proposition 5.4.1.** Let $\varphi : X \to Z$ be a small contraction of a smooth variety of dimension 5 such that $-K_X$ is $\varphi$-spanned, and $F$ a smooth fiber of $\varphi$. Then $(F, N^*_{F/X})$ is one of the following:

<table>
<thead>
<tr>
<th>$F$</th>
<th>$N^*_{F/X}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{P}^3$</td>
<td>$\mathcal{O}<em>{\mathbb{P}^3}(-1) \oplus \mathcal{O}</em>{\mathbb{P}^3}(-1)$</td>
</tr>
<tr>
<td>$\mathbb{Q}^3$</td>
<td>$\mathcal{O}<em>{\mathbb{Q}^3}(-1) \oplus \mathcal{O}</em>{\mathbb{Q}^3}(-1)$</td>
</tr>
<tr>
<td>$\mathbb{P}^2$</td>
<td>$\mathcal{O}<em>{\mathbb{P}^2} \oplus \mathcal{O}</em>{\mathbb{P}^2}(-1) \oplus \mathcal{O}_{\mathbb{P}^2}(-2)$</td>
</tr>
</tbody>
</table>

**Proof.** The anticanonical bundle of the projectivization $\mathbb{P}(N^*)$ is given by the formula

$$-K_{\mathbb{P}(N^*)} = -p^*(K_F + c_1(N^*)) + 2\xi_{N^*},$$

where $p : \mathbb{P}(N^*) \to F$ is the bundle projection; so it is ample, being the sum of two nef divisors, the first of which vanishes only on the fibers of $p$, where
the second is positive. It follows that $N^*$ is a Fano bundle over $\mathbb{P}^3$, $\mathbb{Q}^3$ or $\mathbb{P}^2$ i.e., its projectivization is a Fano manifold. In the rest of the proof we will use the following:

**Lemma 5.4.2** ([AW2, 5.3 and 5.3.1]). Let $\varphi : X \to Z$ be a good or crepant contraction of a smooth variety and let $F$ be a smooth fiber of $\varphi$. Consider the blow-up of $X$ along $F$, $\beta : \hat{X} \to X$ and denote by $\hat{F}$ the exceptional divisor.

If the conormal bundle $N^*_F$ is nef then the line bundle $O_{\hat{X}/X}(1) = -\hat{F}$ is $\varphi \circ \beta$-nef, some positive multiple $O_{\hat{X}/X}(k) = -\hat{k}F$ is $\varphi \circ \beta$-spanned and it defines a good contraction $\hat{\varphi}$ over $W$. The restriction of $\hat{\varphi}$ to $\hat{F}$ is induced by the evaluation map $\bigoplus H^0(S^k(N^*)) \to \bigoplus S^k(N^*)$.

**Corollary 5.4.3.** Suppose that $N^*_F$ is a nonample Fano-bundle, that $\hat{\varphi}_F$ has one dimensional fibers and that either $\hat{\varphi}_F$ is birational or has length $\geq 2$; then $\varphi$ cannot be small.

**Proof.** The restriction of $\hat{\varphi}$ to $\hat{F}$ is an extremal contraction of $\hat{F}$ and we have $l(\hat{\varphi}) \geq l(f|_F)$. Proposition 2.3.6 yields

$$\dim E(\hat{\varphi}) \geq n + l(\varphi) - 2,$$

so that $E(\hat{\varphi}) = \hat{X}$ if $f|_F$ has length $\geq 2$, and $E(\hat{\varphi})$ is a divisor if $\varphi_F$ is birational; both the situations are impossible since, outside $\hat{F}$, the exceptional loci of $\varphi$ and $\hat{\varphi}$ coincide. □

Now we resume the proof of the proposition, starting with the case of a 3-dimensional fiber; by the classification of Fano bundles [SW1] we have the following possibilities:

<table>
<thead>
<tr>
<th>$\mathbb{P}^3, c_1(N^*) = 3$</th>
<th>$\mathbb{P}^3, c_1(N^*) = 2$</th>
<th>$\mathbb{Q}^3, c_1(N^*) = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_{\mathbb{P}^3}(2) \oplus O_{\mathbb{P}^3}(1)$</td>
<td>$O_{\mathbb{P}^3}(1) \oplus O_{\mathbb{P}^3}(1)$</td>
<td>$O_{\mathbb{Q}^3}(1) \oplus O_{\mathbb{Q}^3}(1)$</td>
</tr>
<tr>
<td>$O_{\mathbb{P}^3}(1) \oplus O_{\mathbb{P}^3}(3)$</td>
<td>$O_{\mathbb{P}^3}(2)$</td>
<td>$\pi^* N(1)$</td>
</tr>
<tr>
<td>$\mathcal{N}(1)$</td>
<td>$\mathcal{N}(1)$</td>
<td>$\mathcal{N}(1)$</td>
</tr>
</tbody>
</table>

where $\mathcal{N}$ is a null correlation bundle over $\mathbb{P}^3$ (see [SW1, Section 2]) and $\pi : \mathbb{Q}^3 \to \mathbb{P}^3$ is a double cover.

The bundles with a trivial summand have the property that $h^0(N) - h^1(N) \geq 0$, so they have to be excluded by Lemma 5.2.3, while the null correlation bundle over $\mathbb{P}^3$ and its pullback to $\mathbb{Q}^3$ are ruled out by Corollary 5.4.3.

Now we can study the case in which $F$ is two-dimensional; the classification of Fano bundles on $\mathbb{P}^2$ [SW2] gives us the following possibilities:
### Description of the bundle

| Description of the bundle | $\hat{\phi}|_{\hat{F}}$ | dim fibers $\hat{\phi}|_{\hat{F}}$ |
|---------------------------|--------------------------|-------------------------------|
| $\mathcal{O}(1) \oplus \mathcal{O}(1) \oplus \mathcal{O}$ | Small | 2 |
| $\mathcal{O}(2) \oplus \mathcal{O} \oplus \mathcal{O}$ | Divisorial | 2 |
| $TP^2(-1) \oplus \mathcal{O}(1)$ | Divisorial | 1 |
| $\mathcal{O} \oplus \mathcal{E}_2$ with $\mathcal{E}_2$ in | Divisorial | 1 |
| $0 \to \mathcal{O} \to \mathcal{E}_2(-1) \to J_x \to 0$ | | |
| $0 \to \mathcal{O}(-1)^{\oplus 2} \to \mathcal{O}^3 \to \mathcal{E} \to 0$ | Divisorial | 1 |
| $0 \to \mathcal{O} \to \mathcal{E}_3 \to \mathcal{E}_2$ | Divisorial | 1 |
| $0 \to \mathcal{O}(-2) \to \mathcal{O}^{\oplus 3} \to \mathcal{E} \to 0$ | Fiber type | 1 |

All the bundles with $\hat{\phi}$ divisorial with one-dimensional fibers are ruled out by Corollary 5.4.3; the bundle $\mathcal{O}(2) \oplus \mathcal{O}^{\oplus 2}$ has also to be excluded, because for this bundle $h^0(N) - h^1(N) \geq 2$ and $F$ would move too much.

The last bundle is excluded observing that $H^1(N^*) = 0$ and $H^0(S^k(N^*)) = S^k H^0(N^*)$, so that by [AW2, Proposition 2.4] the contraction of $F$ should be to a smooth 4-dimensional point.

Note that in every case the conormal bundle $N^*_F/X$ is spanned by global sections; using [AW2, Proposition 5.4] we can conclude that the scheme-theoretic structure of $F$, i.e., the closed subscheme defined by the ideal $I_{F^\phi^{-1}(m_z)}\hat{O}_X$, coincides with the geometric structure of $F$.

Moreover, the formal neighbourhood of $F$ in $X$ is determined uniquely and it is the same as the formal neighbourhood of the zero section of the total space of the bundle $N^*$, as follows from a criterion of Mori [Mo2, 3.33]:

**Proposition 5.4.4.** Suppose that $F$ is a smooth fiber of a Fano-Mori or crepant contraction $\varphi : X \to Z$, and assume that the conormal bundle $N^*_{F/X}$ is nef. If $H^1(F, F_\varphi \otimes S^i(N^*)) = H^1(F, N \otimes S^i(N^*)) = 0$ for $i \geq 1$ then the formal neighbourhood of $F$ in $X$ is determined uniquely and it is the same as the formal neighbourhood of the zero section in the total space of the bundle $N^*$.

### 5.5. Description of the flip.

In the case of a three dimensional fiber consider again $\beta : \hat{X} \to X$, the blow up of $X$ along $F$, and denote by $\hat{F}$ the exceptional divisor; since in our case $N^*_F$ is ample, the line bundle $-\hat{F}$ on $\hat{X}$ is $(\varphi \circ \beta)$ ample and hence $-K_{\hat{X}}$ is $(\varphi \circ \beta)$ ample.

Since $\rho(\hat{X}/Z) = 2$ there is a Fano-Mori contraction $\phi : \hat{X} \to X^+$ over $Z$, different from $\varphi$; easy computations show that the line bundle $K_{\hat{X}} - (l(\varphi) + 1)\hat{F}$ is $\varphi \circ \beta$ nef and is a good supporting divisor for the contraction $\phi : \hat{X} \to X^+$; if $N^*_F$ is $\mathcal{O}(1) \oplus \mathcal{O}(1)$ then $\phi$ is a divisorial contraction with...
fibers isomorphic to $\mathbb{P}^3$ and $\mathbb{Q}^3$ respectively, and the flip is given by the following diagram:

![Diagram](attachment:image.png)

Otherwise, if $(F, N_F^*) \simeq (\mathbb{P}^3, \mathcal{O}_F(2) \oplus \mathcal{O}_F(1))$ then $\phi$ is a small contraction of a $\mathbb{P}^3$ with normal bundle $\mathcal{O}_F(-1) \oplus \mathcal{O}_F(-1)$, which falls in the previous case, so that the flip is given by the following diagram:

![Diagram](attachment:image.png)

where $\Phi$ is the contraction of a $\mathbb{P}^4$ with normal bundle $\mathcal{O}_F(-2)$ to a point.

In the case of 2-dimensional fibers we blow up $X$ along $E(\varphi)$; the exceptional divisor of the blow up is a fiber product of a $\mathbb{P}^2$ and a $\mathbb{P}^1$-bundle over $B$; we can contract $\hat{X}$ to a smooth variety by contracting the exceptional divisor to the $\mathbb{P}^1$-bundle over $B$ (see [ABW1]).

**5.6. The case $n > 5$.** Take $n - 5$ generic sections $H_1, \ldots, H_{n-5}$ in $|L|$ and consider the variety $X' = X \cap (\cap_{i}^{n-5} H_i)$; by Lemma 2.2.1 $X'$ is smooth and $K'_{X'} + L_{X'}$ is a good supporting divisor of a small contraction of $X'$; since $F$ is a projective space or a three dimensional quadric and $L_F \simeq \mathcal{O}(1)$, we have that $F' = F \cap (\cap_{i}^{n-5} H_i)$ is smooth and is again a projective space or a quadric. Its conormal bundle is the restriction of the conormal bundle of $F$; moreover by Lemma 2.2.4 the bundle $N_{F/X}'$ is nef. Note that, if $n > 5$ then $l(\varphi) = n - 4$, so we cannot obtain the pair $(\mathbb{P}^3, \mathcal{O}_{\mathbb{P}^3}(1) \oplus \mathcal{O}_{\mathbb{P}^3}(1))$ as $(F', N_{F'/X'})$; the flip of $\varphi$ is constructed exactly as in the case $n = 5$.

**6. Examples.**

**Example 6.1** (small contractions with fibers $\simeq \mathbb{P}^3, \mathbb{Q}^3$). Let $(F', \mathcal{E})$ be a pair consisting of a smooth Fano variety $F'$ and a numerically effective vector bundle $\mathcal{E}$ such that $-K_{F'} - det\mathcal{E}$ is ample.
Consider the projectivization $X := \mathbb{P}(\mathcal{E} \oplus \mathcal{O})$ and the section of the projective bundle $\pi : X \to F'$ determined by the surjection $\mathcal{E} \oplus \mathcal{O} \to \mathcal{O} \to 0$; denote this section by $F$; it is easy to check that $N_{F/X}^* = \mathcal{E}$.

Let $\xi$ denote the tautological line bundle over $X$; since $\xi$ is nef and $\xi - K_{\mathbb{P}(\mathcal{E} \oplus \mathcal{O})}$ is ample, by the Contraction theorem it follows that the linear system $|m\xi|$ is base point free for $m \gg 0$ and defines a Fano-Mori contraction $\varphi : X \to W$ with $W = \mathbb{P}(\oplus_{m \geq 0} H^0(S^m(\mathcal{E} \oplus \mathcal{O})))$.

If $\mathcal{E}$ is ample the map $\varphi$ is just the contraction of $F$ to a point; considering as $(F', \mathcal{E})$ the pairs

<table>
<thead>
<tr>
<th>$F'$</th>
<th>$\mathcal{E}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{P}^3$</td>
<td>$\mathcal{O}<em>{\mathbb{P}^3}(1) \oplus \mathcal{O}</em>{\mathbb{P}^3}(1)$</td>
</tr>
<tr>
<td>$\mathbb{Q}^3$</td>
<td>$\mathcal{O}<em>{\mathbb{Q}^3}(1) \oplus \mathcal{O}</em>{\mathbb{Q}^3}(1)$</td>
</tr>
</tbody>
</table>

we construct examples of small contractions with smooth fibers of dimension three.

**Example 6.2** (a small contraction with fibers $\simeq \mathbb{P}^2$). Let $F' = \mathbb{P}^2$ and $\mathcal{E} = \mathcal{O}_{\mathbb{P}^2}(1) \oplus \mathcal{O}_{\mathbb{P}^2}(1)$; the construction of the previous example gives us a small contraction of a smooth fourfold $f : X' \to W'$ which contracts a projective plane to a point.

Let $B$ be a smooth curve; the map $\varphi := f \times 1_B : X' \times B \to W' \times B$ is a Fano-Mori contraction of a smooth variety of dimension five with fibers $\simeq \mathbb{P}^2$.

**Example 6.3** (small contractions with singular quadrics as fibers). (The following construction was inspired by [AW3], Section 4, and [Ka2].)

Let $S \subset \mathbb{C}^6$ be the hypersurface defined by the equation

$$z_1^2 + z_2^2 + z_3^2 + z_4^2 + z_5^3 = 0,$$

$C \subset S$ the curve $z_1 = z_2 = \cdots = z_5 = 0$, and $Z \subset S$ the threefold $\sum_{i=1}^4 z_i^2 = z_5 = z_6 = 0$.

Consider the blow-up of $S$ along $C$, $\beta : Y \to S$; every fiber of $\beta$ is a three dimensional cone with a singular point.

The strict transform $Z'$ of $Z$ is smooth and meets $F_0 = \beta^{-1}(0,0,0,0,0,0)$ along a smooth two-dimensional quadric; consider now $\sigma : X \to Y$, the blow up of $Y$ along $Z'$. The variety $X$ obtained in this way is smooth, its anticanonical divisor $-K_X$ is $(\sigma \circ \beta)$-ample, and $\rho(X/S) = 2$, so there exists an extremal ray on $X$ not contracted by $\sigma$; this ray is spanned by the class of a line in $F$, the strict transform of $F_0$; the contraction of this ray contracts $F$ to a point and it is a small contraction.

In the same way we can construct a small contraction with a fiber isomorphic to a singular three dimensional quadric with $\dim \text{Sing}(Q) = 1$, taking
the hypersurface
\[ z_1^2 + z_2^2 + z_3^2 + z_4^3 + z_5^2 z_6 + z_5^2 z_6 = 0 \]
as \( S \), the curve \( z_1 = z_2 = \cdots = z_5 = 0 \) as \( C \) and the threefold \( z_1^2 + z_2^2 + z_3^2 + z_4^3 = z_5 = z_6 = 0 \) as \( Z \).

**Example 6.4** (a double covering). Consider a small contraction \( \psi : Y \to Z \) of a smooth fivefold which contracts a \( \mathbb{P}^3 \) with normal bundle \( \mathcal{O}_{\mathbb{P}^3}(-1)^{\oplus 2} \) to a point (cf. Example 6.1); we have seen that such a contraction is supported by \( K_X + 2L \), with \( L \) a \( \psi \)-ample line bundle; in this example we can take \( L = \pi^* \mathcal{O}_{\mathbb{P}^3}(1) \).

Take a smooth divisor \( B \in |2L| \) and construct a double covering \( p : X \to Y \) branched along \( B \); the variety \( X \) is smooth and \( -K_X \) is \( (\psi \circ \pi) \)-ample. Let \( \varphi : X \to W \) be the connected part of the Stein factorization of \( (\psi \circ \pi) \); then \( \varphi \) is a Fano-Mori contraction which fits in a commutative diagram

\[
\begin{array}{ccc}
X & \xrightarrow{\varphi} & W \\
p & & p' \\
Y & \xrightarrow{\psi} & Z
\end{array}
\]

where \( p' : W \to Z \) is a double covering branched along \( \psi(B) \).

The contraction \( \varphi \) is a small covering branched along a three-dimensional quadric to a point.

**Example 6.5** (toric examples).

1) Let \( N \) be a lattice of rank five and \( \Delta' \) the cone generated by a base \( v_1, v_2, v_3, w_1, w_2 \) of \( N \) and by \( v_4 = w_1 + w_2 - v_1 - v_2 - v_3 \); consider the subdivision \( \Delta \) of \( \Delta' \) obtained considering the four cones

\[ \langle v_1, \ldots, \hat{v}_i, \ldots, v_4, w_1, w_2 \rangle. \]

The induced toric morphism \( \varphi : X(\Delta) \to X(\Delta') \) is the contraction of \( V(\langle w_1, w_2 \rangle) \simeq \mathbb{P}^3 \) with normal bundle \( \mathcal{O}_{\mathbb{P}^3}(-1)^{\oplus 2} \) to a point. The flip of \( \varphi \) is given by the following diagram:
where $\hat{\Delta}$ is the subdivision of $\Delta$ obtained introducing the vector $u = w_1 + w_2$ and the cones $\langle v_1, \ldots, \hat{v}_i, \ldots, v_j, w \rangle$, and $\Delta^+$ is the subdivision of $\Delta'$ obtained considering the two cones $\langle v_1, v_2, v_3, v_4, w_1 \rangle$ and $\langle v_1, v_2, v_3, v_4, w_2 \rangle$.

2) Again let $N$ be a lattice of rank five and $\Delta'$ the cone generated by a base $v_1, v_2, v_3, w_1, w_2$ of $N$ and by $v_4 = w_1 + 2w_2 - v_1 - v_2 - v_3$; consider the subdivision $\Delta$ of $\Delta'$ obtained considering the four cones $\langle v_1, \ldots, \hat{v}_i, \ldots, v_4, w_1 \rangle$ and $\langle v_1, v_2, v_3, v_4, w_2 \rangle$ (see [Re, Theorem 3.4]).

The induced toric morphism $\phi : X(\Delta) \to X(\Delta')$ is the contraction of $V(\langle w_1, w_2 \rangle) \cong \mathbb{P}^3$, with normal bundle $O_{\mathbb{P}^3}(-1) \oplus O_{\mathbb{P}^3}(-2)$ to a point, and the flip is given by the following diagram:

$$
\begin{array}{ccc}
X(\Delta) & \xrightarrow{\phi} & X(\Delta') \\
\downarrow{\phi} & & \downarrow{\phi'} \\
X(\Delta^+) & & 
\end{array}
$$

where $\Delta^+$ is the subdivision of $\Delta'$ obtained considering the two cones $\langle v_1, v_2, v_3, v_4, w_1 \rangle$ and $\langle v_1, v_2, v_3, v_4, w_2 \rangle$ (see [Re, Theorem 3.4]).

Note that, introducing in $\Delta$ the vectors $u = w_1 + w_2$ and $v = w_1 + 2w_2$ and the appropriate cones, we can recover from the toric point of view the construction in 5.5.

Acknowledgements. This work grew out of a part of the author’s Ph.D. thesis [Oc] at the Department of Mathematics of the University of Trento: I would like to thank Marco Andreatta who proposed me the problem treated in this paper and helped me with precious suggestions.

I’m also indebted to Andrew Sommese, who gave me the idea of using the Albanese map in the Proof of Theorem 4.1 and to Jaroslaw Wiśniewski for many valuable comments and discussions.

References


Received May 24, 2001 and revised October 22, 2001.

Dipartimento di Matematica
Università degli Studi di Trento
Via Sommarive 14
I-38050 Povo Trento
Italy
E-mail address: occhetta@science.unitn.it
ORBIFOLD HODGE NUMBERS OF CALABI–YAU HYPERSURFACES

Mainak Poddar

We identify the twisted sectors of a compact simplicial toric variety. We do the same for a generic nondegenerate Calabi-Yau hypersurface of an n-dimensional simplicial Fano toric variety and then explicitly compute $h_{orb}^{1,1}$ and $h_{orb}^{n-2,1}$ for the hypersurface. We give applications to the orbifold string theory conjecture and orbifold mirror symmetry.

1. Introduction.

The K-Orbifold string theory conjecture states that there is a natural isomorphism between the Orbifold K-theory of a Gorenstein orbifold and the ordinary K-theory of its crepant resolution (see [AR], [Ru]). To construct a natural isomorphism as the conjecture demands, is a very hard problem. But weaker versions of the conjecture that compare Euler numbers, Hodge numbers, etc. have been studied extensively in the literature in the case of orbifolds that are global-quotients. Batyrev [B2], and Batyrev and Dais [BD] proved, in particular, the equality of orbifold Hodge numbers and the Hodge numbers of smooth crepant resolutions for Gorenstein global-quotient orbifolds. But there were no results for nonglobal-quotient orbifolds.

In this paper, we show that the orbifold Hodge numbers of a generic Calabi-Yau hypersurface in a complex 4-dimensional simplicial Fano toric variety coincide with the Hodge numbers of its smooth crepant resolution. Besides being the first nonglobal-quotient example, this is also an important example in mirror symmetry. An immediate corollary of this is the pairing of orbifold Hodge numbers of Calabi-Yau 3-folds and their Batyrev mirrors.

While this paper was being refereed, extensive generalisations and related results were reported in [BoM], [P] and [Y]. [BoM] and [P] use the characterisation of twisted sectors presented here. [Y] uses the theory of algebraic stacks and achieves a deep result. The survey article [Re] nicely explains the heart of the matter.

Now we briefly describe how this article is organised. In Sections 2 and 3 we review relevant facts from orbifold cohomology and toric geometry respectively. In Section 4 we find characterisations for the twisted sectors of complete simplicial toric varieties and nondegenerate Calabi-Yau hypersurfaces of simplicial Fano toric varieties. In Section 5 we compute formulas for
some orbifold Hodge numbers of these hypersurfaces, state some corollaries and then give an example.

2. Orbifolds.

2.1. Orbifold structure. Let $U$ be a connected topological space, $V$ a connected $n$-dimensional smooth manifold and $G$ a finite group acting smoothly on $V$. Then an $n$-dimensional uniformising system of $U$ is a triple $(V, G, \pi)$, where $\pi : V \to U$ is a continuous map inducing a homeomorphism between the quotient space $V/G$ and $U$. Two uniformising systems $(V_i, G_i, \pi_i)$, $i = 1, 2$, are isomorphic if there is a diffeomorphism $\phi : V_1 \to V_2$ and an isomorphism $\lambda : G_1 \to G_2$ such that $\phi$ is $\lambda$-equivariant, and $\pi_2 \circ \phi = \pi_1$.

If $(\phi, \lambda)$ is an automorphism of $(V, G, \pi)$, then there is a $g \in G$ such that $\phi(x) = g.x$ and $\lambda(a) = gag^{-1}$, for any $x \in V$ and $a \in G$.

Let $i : U' \to U$ be a connected open subset of $U$. An uniformising system $(V', G', \pi')$ of $U'$ is said to be induced from $(V, G, \pi)$ if there is a monomorphism $\lambda : G' \to G$ and a $\lambda$-equivariant open embedding $\phi : V' \to V$ such that $i \circ \pi' = \pi \circ \phi$. The pair $(\phi, \lambda) : (V', G', \pi') \to (V, G, \pi)$ is called an injection. Two uniformising systems $(V_1, G_1, \pi_1)$ and $(V_2, G_2, \pi_2)$ of neighbourhoods $U_1$ and $U_2$ of a point $p$ are equivalent at $p$ if they induce isomorphic uniformising systems for a neighbourhood $U_3$ of $p$.

Let $X$ be a Hausdorff, second countable topological space. An $n$-dimensional orbifold structure on $X$ is given by the following data: For every point $p \in X$, there is an assigned neighbourhood $U_p$ of $p$ and an $n$-dimensional uniformising system $(V_p, G_p, \pi_p)$ of $U_p$. The assignment satisfies the condition that for any point $q \in U_p$, $(V_p, G_p, \pi_p)$ and $(V_q, G_q, \pi_q)$ are equivalent at $q$.

Two orbifold structures $\{(V_p, G_p, \pi_p) : p \in X\}$ and $\{(V'_p, G'_p, \pi'_p) : p \in X\}$ are equivalent if for any $p \in X$, $(V_p, G_p, \pi_p)$ and $(V'_p, G'_p, \pi'_p)$ are equivalent at $p$. With a given equivalence class of orbifold structures on it, $X$ is called an orbifold.

We call each $U_p$ a uniformised neighbourhood of $p$, and $(V_p, G_p, \pi_p)$ a chart at $p$. In fact we choose $U_p$ to be small enough that $G_p$ has the minimum possible order; that is, every element of $G_p$ fixes the preimage of $p$ in $V_p$. In what follows, this choice is assumed. Then a point $p$ is called smooth if $G_p$ is trivial; otherwise, it is called singular. $X$ is called a global-quotient orbifold if $X$ itself is an uniformised open set.

An orbifold $X$ is called reduced if $G_p$ acts effectively on $V_p$. Furthermore if a group element acts nontrivially, we require the fixed-point set to be of at least (real) codimension two, so that the complement is locally connected. We will deal with reduced orbifolds only. Note that even a reduced non-smooth orbifold can have a smooth underlying variety because of examples...
with complex reflections. Gorenstein orbifolds do not present this problem as they do not admit such complex reflections.

### 2.2. Orbifold (Chen-Ruan) cohomology

First we will describe the so-called twisted sectors. Consider the set of pairs:

$$\tilde{X} = \{(p, (g)_{G_p}) \mid p \in X, g \in G_p\},$$

where \((g)_{G_p}\) denotes the conjugacy class of \(g\) in \(G_p\). Then Kawasaki showed (see \([CR]\)) that \(\tilde{X}\) has a natural orbifold structure. We will describe the connected components of \(\tilde{X}\). Recall that each point \(p\) has a local chart \((V_p, G_p, \pi_p)\) which gives a local uniformised neighbourhood \(U_p = \pi_p(V_p)\). If \(q \in U_p\), up to conjugation, there is an injective homomorphism \(G_q \to G_p\). For \(g \in G_q\), the conjugacy class \((g)_{G_p}\) is well-defined. We define an equivalence relation \((g)_{G_q} \sim (g)_{G_p}\). Let \(T\) denote the set of equivalence classes. By an abuse of notation, we use \((g)\) to denote the equivalence class to which \((g)_{G_q}\) belongs. \(\tilde{X}\) is decomposed as a disjoint union of connected components

$$\tilde{X} = \bigsqcup_{(g) \in T} X_{(g)},$$

where

$$X_{(g)} = \{(p, (g')_{G_p}) \mid g' \in G_p, (g')_{G_p} \in (g)\}.$$

**Definition 1.** \(X_{(g)}\) is called a **twisted sector** if \(g \neq 1\). We call \(X_{(1)} = X\) the **nontwisted sector**.

Assume that \(X\) is an almost complex orbifold with an almost complex structure \(J\) (see \([CR]\)). Then for a singular point \(p\), \(J\) gives rise to an effective representation \(\rho_p : G_p \to GL(n, \mathbb{C})\). For any \(g \in G_p\) we write \(\rho_p(g)\), up to conjugation, as a diagonal matrix \(\text{diag}\left(e^{2\pi i \frac{m_{i,g}}{m_g}}, \ldots, e^{2\pi i \frac{m_{n,g}}{m_g}}\right)\), where \(m_g\) is the order of \(g\) in \(G_p\), and \(0 \leq m_{i,g} < m_g\). Define a function \(\iota : \tilde{X} \to \mathbb{Q}\) by

$$\iota(p, (g)_{G_p}) = \sum_{i=1}^{n} \frac{m_{i,g}}{m_g}.$$

This function \(\iota : \tilde{X} \to \mathbb{Q}\) is locally constant. Denote its value on \(X_{(g)}\) by \(\iota_{(g)}\). \(\iota_{(g)}\) is called the **degree shifting number** of \(X_{(g)}\). It has the following properties:

1. \(\iota_{(g)}\) is integral iff \(\rho_p(g) \in SL(n, \mathbb{C})\).
2. \(\iota_{(g)} + \iota_{(g^{-1})} = \text{rank}(\rho_p(g) - I) = n - \dim(X_{(g)})\).

**Definition 2.** An almost complex orbifold is called Gorenstein if \(\iota_{(g)}\) is integral for all \((g)\).
Remark. An almost complex, complex or Kähler structure on $X$ induces a corresponding similar structure on each $X_{(g)}$.

**Definition 3.** Let $\mathbb{F}$ be any field containing $\mathbb{Q}$ as a subfield. We define the *orbifold chomology groups* of $X$ with coefficients in $\mathbb{F}$ by

$$H^d_{\text{orb}}(X; \mathbb{F}) = \oplus_{(g)\in T} H^{d-2\iota(g)}(X_{(g)}; \mathbb{F}).$$

**Definition 4.** Let $X$ be a closed complex orbifold. We define, for $0 \leq p, q \leq \dim_{\mathbb{C}} X$, *orbifold Dolbeault cohomology groups*

$$H^{p,q}_{\text{orb}}(X; \mathbb{C}) = \oplus_{(g)\in T} H^{p-\iota(g)-q+\iota(g)}(X_{(g)}; \mathbb{C}).$$

Remark. When $X$ is a closed Kähler orbifold (so is each $X_{(g)}$), these Dolbeault groups are related to the singular cohomology groups of $X$ and $X_{(g)}$ as in the manifold case, and the Hodge decomposition theorem holds for these cohomology groups.

**Definition 5.** We define *orbifold Hodge numbers* by

$$h^{p,q}_{\text{orb}}(X) = \dim H^{p,q}_{\text{orb}}(X; \mathbb{C}).$$

3. **Facts from toric geometry.**

3.1. **Orbits, divisors and polytopes.** A complex $n$-dimensional toric variety $X_\Xi$ is constructed from an $n$-dimensional lattice $N$ and a fan $\Xi$ in $\mathbb{R}^n = N\otimes_{\mathbb{Z}} \mathbb{R}$. We will write $X$ for $X_\Xi$ when there is no confusion. Let $M = \text{Hom}(N, \mathbb{Z})$ denote the dual lattice, with dual pairing denoted by $\langle , \rangle$.

If $\sigma$ is a cone in $N$, the dual cone $\check{\sigma}$ in $M\mathbb{R}$ determines a finitely generated commutative semigroup $R_\sigma = \check{\sigma}\cap M$. $\mathbb{C}[R_\sigma]$ is the $\mathbb{C}$-algebra with generators $\chi^m$ for each $m \in R_\sigma$ and relations $\chi^m\chi^{m'} = \chi^{m+m'}$. It gives an open affine subset $U_\sigma := \text{spec}(\mathbb{C}[R_\sigma])$ of $X$. A face $\tau$ of $\sigma$ gives an inclusion $U_\tau \hookrightarrow U_\sigma$.

$\Xi(d)$ denotes the set of $d$-dimensional cones of $\Xi$. We reserve the letter $\eta$ to denote elements of $\Xi(1)$. For each $\eta$, let $v_\eta$ denote the unique generator of the semigroup $\eta\cap N$. The $v_\eta \in \sigma$ are the *generators* of $\sigma$. If $r = |\Xi(1)|$ is the number of 1-dimensional cones, we sometimes write the $v_\eta$’s as $v_1, \ldots, v_r$.

$X$ is nonsingular iff for every cone in $\Xi$, its generators are part of a $\mathbb{Z}$-basis of $N$. Such a fan is called *smooth*. $X$ is an orbifold iff the generators for every cone in $\Xi$ are linearly independent over $\mathbb{R}$; and we say $X$ and $\Xi$ are *simplicial*.

The action of the torus $T_N = U_{\{0\}} = N\otimes \mathbb{C}$ on $X$ has exactly one orbit $O_\tau$ corresponding to each cone $\tau \in \Xi$. Each $O_\eta$ is an irreducible $T_N$-invariant Weil divisor denoted $D_\eta$. If $X$ is complete, these generate the Chow group $A_{n-1}(X)$. Two $T_N$-invariant Weil divisors are linearly equivalent iff they differ by $\text{div}(\chi^m) = \sum_\eta \langle m, v_\eta \rangle D_\eta$ for some $m \in M$. A Weil divisor $D = \sum_\eta a_\eta D_\eta$ is Cartier iff for each $\sigma \in \Xi$, there is $m_\sigma \in M$ such that...
\langle m_\sigma, v_\eta \rangle = -a_\eta \text{ whenever } \eta \subset \sigma. \text{ A Cartier divisor } D \text{ is ample iff } \langle m_\sigma, v_\eta \rangle > -a_\eta \text{ whenever } \eta \text{ is not in } \sigma \text{ and } \sigma \text{ is } n\text{-dimensional.}

If \( X \) is complete and \( D = \sum_\eta a_\eta D_\eta \) is Cartier, then \( \Delta_D = \{ m \in M_R : \langle m, v_\eta \rangle \geq -a_\eta \forall \eta \} \) is a polytope. A polytope is called \textit{integral} if its vertices are integral. \( \Delta_D \) is integral if \( D \) is ample. Conversely, given any \( n\)-dimensional integral polytope \( \Delta \) one can canonically associate a projective toric variety \( \mathbb{P}_\Delta \) to it. See [CK], Section 3.2.2 for details. It comes with a specific choice of ample divisor \( D_\Delta \) such that \( \Delta_{D_\Delta} = \Delta \). The \( T_N \) orbit closures of \( \mathbb{P}_\Delta \) are in one-to-one correspondence with the nonempty faces \( F \) of \( \Delta \). There is a natural inclusion of toric varieties \( \mathbb{P}_F \hookrightarrow \mathbb{P}_\Delta \).

Choose a basis for \( M \). This corresponds to picking coordinates \( t_1, \ldots, t_n \) for the torus \( T_N \). Then, if \( m \in M \) is written \( m = (a_1, \ldots, a_n) \), we have \( \chi^m = \prod_{i=1}^n t_i^{a_i} \), so we can write \( t^m \) instead of \( \chi^m \). For any \( k \geq 0 \), we have the space of Laurent polynomials \( \mathcal{L}(k\Delta) = \{ f : f = \sum_{m \in \Delta \cap M} \lambda_m t^m, \lambda_m \in \mathbb{C} \} \). Each \( f \in \mathcal{L}(k\Delta) \) gives the affine hypersurface \( Z_f \subset T_N \) defined by \( f = 0 \). There is a \( T_N \)-equivariant map \( \mathcal{H}^0(X, \mathcal{O}(D)) \cong \bigoplus_{m \in \Delta \cap M} \mathbb{C} \chi^m \). So \( \mathcal{L}(k\Delta) \simeq \mathcal{H}^0(\mathbb{P}_\Delta, \mathcal{O}_{\mathbb{P}_\Delta}(kD_\Delta)) \). Under this isomorphism, \( f \) corresponds to an effective divisor \( Z_f \subset \mathbb{P}_\Delta \). \( Z_f \) is a hypersurface. It is a compactification of \( Z_f \) for generic \( f \).

We will use the following notation:
(a) \( l(k\Delta) = |k\Delta \cap M| = \dim(L(k\Delta)) \),
(b) \( l^*(k\Delta) = \{ m \in k\Delta \cap M : m \text{ is not in any facet of } k\Delta \cap M \} \).

### 3.2. Homogeneous coordinate ring

We introduce a variable \( x_\eta \) for each \( \eta \in \Xi(1) \) and consider the polynomial ring \( S = \mathbb{C}[x_\eta : \eta \in \Xi(1)] \). A monomial in \( S \) is written \( x^D = \prod_\eta x_\eta^{a_\eta} \), where \( D = \sum_\eta a_\eta D_\eta \) is an effective torus-invariant divisor on \( X \). We say that \( x^D \) has degree \( \deg(x^D) = [D] \in A_{n-1}(X) \). Thus, \( S \) is graded by \( A_{n-1}(X) \). Given a divisor class \( \alpha \in A_{n-1}(X) \), \( S_\alpha \) denotes the graded piece of \( S \) of degree \( \alpha \). We often write the variables as \( x_1, \ldots, x_r \), where \( x_i \) corresponds to the cone generator \( v_i \) and \( r = |\Xi(1)| \). The ring \( S \), together with the grading defined above is called the \textit{homogeneous coordinate ring} of \( X \). See [CK], Chapter 3.2 for more details.

If \( \tau \) is any cone of \( \Xi \) then the orbit closure \( \overline{O}_\tau \) is given by the ideal \( (x_i : v_i \text{ is a generator of } \tau) \) of \( S \). Also the graded pieces of \( S \) have nice cohomological interpretation. We noted that \( L(\Delta) \cong \mathcal{H}^0(\mathbb{P}, \mathcal{O}_{\mathbb{P}}(D_\Delta)) \). Now the map sending the Laurent monomial \( t^m \) to \( \prod_\eta x_\eta^{(m, v_\eta) + a_\eta} \) induces an isomorphism \( \mathcal{H}^0(X, \mathcal{O}_X(D)) \cong S_\alpha \), where \( \alpha = [D] \in A_{n-1}(X) \).

### 3.3. Fano toric varieties

For any toric variety \( X \), the anticanonical divisor \( -K_X = \sum_\eta D_\eta \). A complete toric variety \( X \) is called \textit{Fano} if \( -K_X \) is Cartier and ample.

The anticanonical divisor of a Fano toric variety \( X \) determines a \textit{reflexive} polytope \( \Delta \). An integral polytope is called reflexive if:
(a) All facets $\Gamma$ of $\Delta$ are supported by an affine hyperplane of the form 
\[
\{ m \in M : \langle m, v \rangle = -1 \} \text{ for some } v \in N.
\]
(b) $\text{Int}(\Delta) \cap M = \{0\}$.

The polar polytope $\Delta^\circ$ of the reflexive polytope $\Delta$ is obtained by $\Delta^\circ = \{ v \in N : \langle m, v \rangle \geq -1 \text{ for all } m \in \Delta \} \subset N$. The fan $\Xi$ of $X$ can be retrieved by coning over the proper faces of $\Delta^\circ$. This fan is called the normal fan of $\Delta$ and $X = \mathbb{P}_\Delta^\circ$. $\Delta^\circ$ is also reflexive and $(\Delta^\circ)^\circ = \Delta$. The Fano toric variety constructed from the normal fan of $\Delta^\circ$ is denoted by $\mathbb{P}_{\Delta^\circ}$.

We shall use $F$ and $F^\circ$ to denote a face of $\Delta$ and $\Delta^\circ$ respectively. There exists an inclusion reversing duality between the faces of $\Delta$ and $\Delta^\circ$. For instance, the face of $\Delta$ dual to the face $F^\circ$ of $\Delta^\circ$ is defined as 
\[
\hat{F}^\circ := \{ m \in \Delta : \langle n, m \rangle = -1 \forall n \in F^\circ \}.
\]
Let $\tau$ be the cone in the normal fan of $\Delta$ associated to the face $F^\circ$, then the orbit closure $O_{\tau} = \mathbb{P}_{\hat{F}^\circ}$.

Generic anticanonical hypersurfaces $V$ in $\mathbb{P}_\Delta$ and $V^\circ$ in $\mathbb{P}_{\Delta^\circ}$ constitute two families of Calabi-Yau varieties, which are conjectured to be mirror families in the sense of Conformal Field Theory and called Batyrev mirrors in the literature. These varieties are orbifolds if the corresponding ambient toric variety is simplicial. Let $\tilde{V}$ and $\tilde{V}^\circ$ denote the MPCP resolutions (see [B1] or [CK]) of $V$ and $V^\circ$ respectively. These are again Calabi-Yau. These are smooth if $n = 4$.

**Remark.** A simplicial Fano toric variety or its Calabi-Yau hypersurfaces are Gorenstein orbifolds, the orbifold structures arising naturally from the algebraic structures. In particular, all the degree shifting numbers are integers and the singular locus is of at least complex codimension two.

4. **Twisted sectors.**

We claim that the twisted sectors of a toric variety or a Calabi-Yau hypersurface, up to reduction of orbifold structure, can be identified with subvarieties. Note that in general a twisted sector could be a multiple cover of the corresponding singular locus even if the group actions are all Abelian.

4.1. **Twisted sectors in simplicial toric variety.** Let $\Xi$ be any complete simplicial fan. Then the orbifold structure of the toric variety $X_\Xi$ can be described as follows. Let $\sigma$ be any $n$-dimensional cone of $\Xi$. Let $v_1, \ldots, v_n$ be the generators of $\sigma$. These are linearly independent in $N_\mathbb{R}$. Let $N_\sigma$ be the sublattice of $N$ generated by $v_1, \ldots, v_n$. Let $G_\sigma := N/N_\sigma$ be the quotient group. $G_\sigma$ is finite and abelian.

Let $\sigma'$ be the cone $\sigma$ regarded in $N_\sigma$. Let $\hat{\sigma}'$ be the dual cone of $\sigma'$ in $M_\sigma$, the dual lattice of $N_\sigma$. $U_{\hat{\sigma}'} = \text{spec}(\mathbb{C}[\hat{\sigma}' \cap M_\sigma])$. Note that $\sigma'$ is a smooth cone in $N_\sigma$. So $U_{\hat{\sigma}'} \cong \mathbb{C}^n$.

There is a canonical dual pairing $M_\sigma/M \times N/N_\sigma \to \mathbb{Q}/\mathbb{Z} \to \mathbb{C}^*$, the first map by the pairing $\langle , \rangle$ and the second by $q \mapsto \exp(2\pi i q)$. Now $G_\sigma$ acts on
$\mathbb{C}[M_{\sigma}]$, the group ring of $M_{\sigma}$, by $v(\chi^u) = \exp(2\pi i (u, v))\chi^u$, for $v \in N$ and $u \in M_{\sigma}$. Note that

\begin{equation}
(\mathbb{C}[M_{\sigma}])^{G_\sigma} = \mathbb{C}[M].
\end{equation}

Thus $G_\sigma$ acts on $U_{\sigma'}$. Let $\pi_\sigma$ be the quotient map. Then $U_\sigma = U_{\sigma'}/G_\sigma$. So $U_\sigma$ is uniformised by $(U_{\sigma'}, G_\sigma, \pi_\sigma)$. For any $\tau < \sigma$, the orbifold structure on $U_\tau$ is same as the one induced from the uniformising system on $U_\sigma$. Then in the absence of complex reflections, toric gluing implies that $\{(U_{\sigma'}, G_\sigma, \pi_\sigma) : \sigma \in \Xi(n)\}$ defines a reduced orbifold structure on $X$. We show this in the general case.

Let $B$ be the nonsingular matrix with generators $v_1, \ldots, v_n$ of $\sigma$ as rows. Then $\sigma'$ is generated in $M_\sigma$ by the the column vectors $v_1^\tau, \ldots, v_n^\tau$ of the matrix $B^{-1}$. So $\chi^v_1, \ldots, \chi^v_n$ are the coordinates of $U_{\sigma'}$. For any $\kappa = (k_1, \ldots, k_n) \in N$, the corresponding coset $[\kappa] \in G_\sigma$ acts on $U_{\sigma'}$ in these coordinates as a diagonal matrix: $\text{diag}(e^{2\pi i k_1}, \ldots, e^{2\pi i k_n})$ where $c_i = (\kappa, v^i)$. Such a matrix is uniquely represented by an $n$-tuple $a = (a_1, \ldots, a_n)$ where $a_i \in \{0, 1\}$ and $c_i = a_i + b_i, b_i \in \mathbb{Z}$. In matrix notation, $\kappa B^{-1} = a + b \iff \kappa = aB + bB$. We denote the integral vector $aB$ in $N$ by $\kappa_a$ and the diagonal matrix corresponding to $a$ by $g_a$. $\kappa_a \leftrightarrow g_a$ gives a one to one correspondence between the elements of $G_\sigma$ and the integral vectors in $N$ that are linear combinations of the generators of $\sigma$ with coefficients in $\{0, 1\}$.

Now let us examine the orbifold chart induced by $(U_{\sigma'}, G_\sigma, \pi_\sigma)$ at any point $x \in U_\sigma$. By the orbit decomposition of $U_\sigma$, there is a unique face $\tau$ of $\sigma$ such that $x \in O_\tau$. Without loss of generality assume that $\tau$ is generated by $v_1, \ldots, v_j, j \leq n$. Then any preimage of $x$ with respect to $\pi_\sigma$ has coordinates $\chi^v_i = 0$ if $i \leq j$. Let $z = (0, \ldots, 0, z_{j+1}, \ldots, z_n)$ be one such preimage. Let $G_\tau := \{g_a \in G_\sigma : a_i = 0 \text{ if } j + 1 \leq i \leq n\}$. We can find a small neighbourhood $W \subset (\mathbb{C}^*)^{n-j}$ of $(z_{j+1}, \ldots, z_n)$ such that the inclusions $(\mathbb{C}^j \times W) \hookrightarrow (\mathbb{C}^n \times W)$ and $G_\tau \hookrightarrow G_\sigma$ induces an injection of uniformising systems $(\mathbb{C}^j \times W, G_\tau, \pi) \hookrightarrow (U_{\sigma'}, G_\sigma, \pi_\sigma)$ on some small open neighbourhood $U_\delta$ of $x$. So we have $G_x = G_\tau$, and an orbifold chart $(\mathbb{C}^j \times W, G_\tau, \pi)$. Note that $G_\tau$ can be constructed from the set $\{\kappa_a = \sum_{i=1}^{j} a_i v_i : \kappa_a \in N, a_i \in \{0, 1\}\}$ which is completely determined by $\tau$ and hence is independent of $\sigma$.

Now we determine the twisted sectors. Take any $x \in X$. $x$ belongs to a unique $O_\tau$. Assume the generators of $\tau$ are $v_1, \ldots, v_j$. Consider any $n$-dimensional $\sigma > \tau$. Assume $v_1, \ldots, v_n$ generate $\sigma$. First suppose there is a $g_a$ in $G_\tau$ such that $a_i \neq 0, \forall i \leq j$ i.e., $\kappa_a$ lies in the interior of $\tau$. We want to find the twisted sector $X_{(g_a)}$. Consider $g_a$ as an element of $G_\sigma$. It is clear that $g_a$ fixes $z \in U_{\sigma'}$ iff $z_1 = \cdots = z_{j+s} = 0$, for some $s \geq 0$. Hence $\pi_\sigma(z) \in O_\tau$ or $\pi_\sigma(z) \in O_\delta$ for some $\delta > \tau$. So $X_{(g_a)} \cap U_\sigma = \overline{O_\tau \cap U_\sigma}$. Since a twisted sector is connected, $X_{(g_a)} = \overline{O_\tau}$. If $g_a \in G_x$ is such that (without loss of generality) only $a_1, \ldots, a_k \neq 0, k < j$, then $g_a \in G_\delta$ where $\delta$ is the
cone generated by \( v_1, \ldots, v_k \) and by the above argument \( X_{(g_\sigma)} = \mathcal{O}_\delta \). Thus we have proved the following theorem:

**Theorem 1.** A twisted sector of any complete simplicial toric variety \( X_\mathbb{Z} \) is isomorphic to a subvariety \( \overline{\sigma} \tau \) of \( X_\mathbb{Z} \) for some cone \( \tau \in \mathbb{Z} \). Moreover, there is a one-to-one correspondence between the set of twisted sectors of the type \( \overline{\sigma} \tau \) and the set of integral vectors in the interior of \( \tau \) which are linear combinations of the 1-dimensional generators of \( \tau \) with coefficients in \( (0, 1) \).

Note that the degree shifting number \( \delta_{(g_\sigma)} = \sum a_i \). Now if \( X_\mathbb{Z} \) is Fano, i.e., \( \mathbb{Z} \) is obtained by coning over the faces of a reflexive polytope \( \Delta^0 \), then the twisted sectors with \( \delta = 1 \) are in one to one correspondence with the integral interior points of faces of \( \Delta^0 \).

### 4.2. Twisted sectors of a hypersurface of a Fano variety.

We identify the twisted sectors of a generic nondegenerate anticanonical (Calabi-Yau) hypersurface \( V \) of a simplicial Fano toric variety \( X = \mathbb{P}_\Delta \). Nondegenerate means that \( V \cap O_\tau \) is either empty or a smooth subvariety of codimension one in \( O_\tau \), for each torus orbit \( O_\tau \) in \( X \). Then \( V \) turns out to be a suborbifold of \( X \). Also nondegeneracy is a generic condition. We show that \( V = \overline{Z_f} \), for a generic \( f \in L(\Delta) \), is nondegenerate and a suborbifold of \( X \). For a different treatment of this, see [BC].

Consider any \( n \)-dimensional cone \( \sigma \) with generators \( v_1, \ldots, v_n \). For notational simplicity set \( \chi^{v_i} = z_i \). Then \( z_1, \ldots, z_n \) are the coordinates of \( U_{\sigma'} \). Let \( Y \) be the preimage of \( V \cap U_\sigma \) in \( U_{\sigma'} \) with respect to \( \pi_\sigma \). Then \( Y \) is defined by the equation \( \sum_{m \in \Delta \cap M} \lambda_m \prod_{i=1}^n z_i^{(m, v_i)} + 1 = 0 \). This is because, \( t^m = \prod_{i=1}^n z_i^{q_i} \iff m = \sum q_i v_i = B^{-1} q \iff q = m B \iff q_i = \langle m, v_i \rangle \). The one is added to ensure that \( V \) is anticanonical. Note that by definition of \( \Delta \), \( \langle m, v_i \rangle + 1 \geq 0 \). If \( \lambda_{m_\sigma} \neq 0 \) then \( Y \) does not pass through the origin. It can be checked from this description using Bertini’s theorem that for generic values of the coefficients \( \lambda_{m_\sigma} \), \( Y \) is a smooth submanifold of \( U_{\sigma'} \) that intersects the coordinate planes \( z_{i_1} = \cdots = z_{i_j} = 0 \) transversely.

\( Y \) is \( G_\sigma \)-stable by (4.1.1). When \( Y \) is smooth, all singularities of \( V \cap U_\sigma \) are quotient singularities induced by action of \( G_\sigma \) on \( Y \). Since there are only finitely many \( n \)-dimensional cones, \( V \) is nondegenerate and a suborbifold of \( X \). \( (Y, G_\sigma, \pi_\sigma) \) is an uniformising system for \( V \cap U_\sigma \). Let \( \tau \) be the face of \( \sigma \) obtained by coning over the face \( F^0 \) of \( \Delta^0 \). Without loss of generality let \( v_1, \ldots, v_j : j \geq 2 \) be the generators of \( \tau \). (We remarked in Section 3.3 that there is no codimension one singularity.) We want to find a chart for any point \( x \in V \cap O_\tau \). By our earlier remark that \( Y \) misses the origin of \( U_{\sigma'} \), \( V \cap O_\tau \) is empty. So we need only consider proper faces \( \tau \) of \( \sigma \). First assume that \( F^0 \) has codimension 2. This means that \( O_\tau \) has dimension 1. Then the corollary on page 112 of [F] implies that the number of points in \( V \cap \mathcal{O}_\tau \) is the normalised volume of \( \mathcal{F}^0 \), which equals \( \lambda^* (\mathcal{F}^0) + 1 \) since \( \mathcal{F}^0 \)
has dimension 1. Since the only other points in $\overline{O_{\tau}}$ in this case are $O_{\sigma}$ for $n$-dimensional cones $\sigma > \tau$, all the intersection points actually lie in $O_{\sigma}$. If codimension $F^\circ$ is bigger than 2, then $V \cap \overline{O_{\tau}}$ is irreducible by Bertini.

Following the same notation as before, $x$ has a small neighbourhood $U_x \cap Y$ such that $((\mathbb{C}^l \times W) \cap Y, G_{\tau}, \pi)$ is a chart for $V$ at $x$. $\mathbb{C}^l \times W$ is, as before, a suitable neighbourhood of some preimage $z$ of $x$ in $U_{\sigma}$'. The tangent space $TY_z$ is a $G_{\tau}$-stable subspace of $T\mathbb{C}^n$. Any $g_a \in G_{\tau}$ acts trivially on $TW_z = \text{span}\{\partial/\partial z_i, i = j + 1, \ldots, n\}$. By transversality, we can choose basis $\{\xi_1, \ldots, \xi_n\}$ of $T\mathbb{C}^n$ such that $\xi_i \in TY_z$ for $i \leq n - 1$ and $\xi_n \in TW_z$. $g_a$ acts trivially on $\xi_n$. This implies that the degree shifting number of $g_a|_{TY_z}$ is still $\sum_{i=1}^n a_i$.

From the description of the charts, it is clear that twisted sectors of $V$ are isomorphic to $V \cap \overline{O_{\tau}}$ where $2 \leq \dim(\tau) \leq n - 1$. Recall that $\overline{D} = \mathbb{P}_{F^\circ}$ where $F^\circ$ is the face of $\Delta$ dual to $F^\circ$. In particular we have the following theorem:

**Theorem 2.** Let $V$ be a generic nondegenerate anticanonical hypersurface of an $n$-dimensional simplicial Fano toric variety $\mathbb{P}_\Delta$. Then the twisted sectors of $V$ are isomorphic to $V \cap \overline{O_{\tau}}$ for some face $F^\circ$ of $\Delta$ such that $1 \leq \dim(F^\circ) \leq n - 2$. There is exactly one twisted sector of this type having $i = 1$, corresponding to each integral interior point of $F^\circ$ if $\dim(F^\circ) < n - 2$. If $\dim(F^\circ) = n - 2$, then there are exactly $l^*(F^\circ) + 1$ twisted sectors of this type having $i = 1$, corresponding to each integral interior point of $F^\circ$.

5. Orbifold Hodge numbers.

5.1. $h^{1,1}_{\text{orb}}(V)$. Let $V_{(g)}$ denote a twisted sector of the hypersurface $V$ and $\iota_{(g)}$ its degree shifting number. $h^{1,1}_{\text{orb}}(V) = h^{1,1}(V) + \sum_{\iota_{(g)} = 1} h^{0,0}(V_{(g)})$. Since $h^{0,0}(V_{(g)}) = 1$ for each twisted sector, by Theorem 2 we obtain

$$\sum_{\iota_{(g)} = 1} h^{0,0}(V_{(g)}) = \sum_{1 \leq \dim(F^\circ) \leq n - 2} l^*(F^\circ) + \sum_{\dim(F^\circ) = n - 2} l^*(F^\circ) l^*(F^\circ)$$

$$= l(\Delta^\circ) - r - 1 - \sum_{\dim(F^\circ) = n - 1} l^*(F^\circ) + \sum_{\dim(F^\circ) = n - 2} l^*(F^\circ) l^*(F^\circ).$$

To compute $h^{1,1}(V)$ we invoke the following Lefschetz hyperplane theorem ([BC, Proposition 10.8]):

**Lemma 1.** Let $V$ be a nondegenerate ample hypersurface of an $n$-dimensional complete simplicial toric variety $X$. Then the natural map induced by inclusion $j^*: H^i(X) \to H^i(V)$, is an isomorphism for $i < n - 1$ and an injection for $i = n - 1$. 
In our case $V$ is anticanonical, and since the anticanonical divisor of a Fano variety is ample, $V$ is ample. Also it is well-known ([F, Section 5.1]) that for any simplicial toric variety $X$, $H^2(X, \mathbb{R}) = H^{1,1}(X, \mathbb{R}) = A_{n-1}(X) \otimes \mathbb{R}$. So for $n \geq 4$, $h^{1,1}(V) = h^{1,1}(\mathbb{P}_\Delta) = \text{rank} A_{n-1}(\mathbb{P}_\Delta) = r - n$.

Thus we have the following theorem:

**Theorem 3.** For any generic nondegenerate anticanonical hypersurface $V$ of an $n$-dimensional simplicial Fano toric variety $\mathbb{P}_\Delta$, $n \geq 4$,

$$h_{\text{orb}}^{1,1}(V) = l(\Delta^\circ) - n - 1 - \sum_{\dim(F^n) = n-1} l^*(F^n) + \sum_{\dim(F^n) = n-2} l^*(F^n)l^*(\widehat{F^n}).$$

**5.2. $h^{n-2,1}(V)$.** Next we compute $h^{n-2,1}(V)$ for $n \geq 4$. For this we use the homogeneous coordinate ring $S$ of $X = \mathbb{P}_\Delta$. Let $v_1, \ldots, v_r$ be the one-dimensional cones of the normal fan of $\Delta$. Let $x_1, \ldots, x_r$ be the corresponding homogeneous coordinates. Let $\beta_0 = [-K_X] = [\sum_{i=1}^r D_i] \in A_{n-1}(X)$. Then $S_{\beta_0} \simeq L(\Delta)$. And the divisor $f \in L(\Delta)$ corresponding to $V$ can be written in the homogeneous coordinates as $f = \sum_{m \in M \cap \Delta} \lambda_m \prod_{i=1}^r x_i^{\langle m, v_i \rangle + 1}$. For notational simplicity, we will denote $\prod_{i=1}^r x_i^{\langle m, v_i \rangle + 1}$ by $x^m$ for any $m \in M$. Define the Jacobian ideal of $f$ to be $J(f) = (\partial f/\partial x_1, \ldots, \partial f/\partial x_r)$.

First we quote the following theorem ([BC, Theorem 10.13]):

**Lemma 2.** Let $X$ be an $d$-dimensional complete simplicial toric variety and $V \subset X$ be a quasi-smooth (i.e., suborbifold) ample hypersurface defined by $f \in S_{\beta}$. Then for any $\lambda \neq (d/2) + 1$, there exists a canonical isomorphism

$$(S/J(f))_{k\beta - \beta_0} \simeq PH^{d-k,k-1}(V).$$

**Remark.** The primitive cohomology $PH^{d-1}(V) := H^{d-1}(V)/(\text{im } H^{d-1}(X))$. This coincides with the usual cohomology if $d$ is even.

For our application of Lemma 2, set $d = n$, $k = 2$ and $\beta = \beta_0$. With $n \geq 4$ these choices imply that $k \neq (d/2) + 1$ in the lemma. Also, $H^{n-2,1}(X) = 0$ if $n \geq 4$, so that $PH^{n-2,1}(V) = H^{n-2,1}(V)$. Thus $h^{n-2,1}(V) = \text{rank } (S/J(f))_{\beta_0}$.

**Lemma 3.** $x_i \partial f/\partial x_i \in J(f)_{\beta_0}, i = 1, \ldots, r$, and the space of complex linear relations among these has dimension $r - (n + 1)$.

**Proof.**

$$x_i \frac{\partial f}{\partial x_i} = \sum_m \lambda_m (\langle m, v_i \rangle + 1)x^m,$$

$$\sum_i c_i x_i \frac{\partial f}{\partial x_i} = \sum_m \lambda_m \left( \left\langle m, \sum_i c_i v_i \right\rangle + \sum_i c_i \right) x^m.$$

For a generic $f$ we can assume that $\lambda_m \neq 0$ for each $m \in \Delta \cap M$. Hence

$$\sum_i c_i x_i \partial f/\partial x_i \equiv 0 \iff \langle m, \sum_i c_i v_i \rangle + \sum_i c_i = 0 \forall m \in \Delta \cap M.$$


In particular, taking \( m = 0 \) we get \( \sum_i c_i = 0 \). Therefore \( \langle m, \sum_i c_i v_i \rangle = 0 \) \( \forall m \in \Delta \cap M \) and since \( \Delta \) is \( n \)-dimensional we have \( \sum_i c_i v_i = 0 \).

Thus \( \sum_i c_i x_i \partial f / \partial x_i = 0 \iff \sum_i c_i v_i = 0 \) and \( \sum_i c_i = 0 \).

Now let \( \tilde{v}_i = (v_i, 1) \in \mathbb{R}^{n+1} \subset \mathbb{C}^{n+1} = \mathbb{R}^{n+1} \otimes \mathbb{C} \).

Since the \( v_i \)'s are vertices of \( \Delta^o \), the \( \tilde{v}_i \)'s are generators of the \( (n+1) \)-dimensional cone \( \{ q \in \mathbb{R}^{n+1} : q t \in \Delta^o \times \{1\} \text{ for some } t \in \mathbb{R}^{n+1} \} \). So the \( \tilde{v}_i \)'s span \( \mathbb{R}^{n+1} \) over \( \mathbb{R} \), and hence they span \( \mathbb{R}^{n+1} \otimes \mathbb{C} \) over \( \mathbb{C} \). Note that \( (\sum c_i v_i, \sum c_i) = \sum c_i \tilde{v}_i \). Hence the lemma follows. \( \square \)

Without loss of generality assume that the \( x_k \partial f / \partial x_k, k = 1, \ldots, n + 1 \), are linearly independent. In other words \( \tilde{v}_k, k = 1, \ldots, n + 1 \) are linearly independent. We will consider monomials \( \prod_{j \neq i} x_j^{p_j} \) that have same degree in \( S \) as \( x_i \). So we want \( m^* \in M \) such that \( \langle m^*, v_j \rangle = p_j \geq 0 > -1 \) if \( j \neq i \) and \( \langle m^*, v_i \rangle = -1 \). Such \( m^* \) is given by interior lattice points of \( F_i \), the \( (n-1) \)-dimensional face of \( \Delta \) that is dual to the 0-dimensional face \( \{ v_i \} \) of \( \Delta^o \). Then for each \( m^* \in \text{Int}(F_i) \cap M \), \( \prod_{j \neq i} x_j^{\langle m^*, v_j \rangle} \partial f / \partial x_i \) belongs to \( J(f)_{\beta_0} \). Together with the \( x_i \partial f / \partial x_i \), these generate \( J(f)_{\beta_0} \) as we vary over all \( i \).

In the following computation, we denote the characteristic function of a set by \( I(\cdot) \). For instance, \( I(\{ m' \in \Delta \}) = 1 \) when \( m' \in \Delta \) and 0 otherwise. Also, recall that \( f = \sum_{m' \in M \cap \Delta} \lambda_{m'} x^{m'} \).

\[
\left( \prod_{j \neq i} x_j^{\langle m^*, v_j \rangle} \right) \partial f / \partial x_i \\
= \left( \prod_{j=1}^{r} x_j^{\langle m^*, v_j \rangle} \right) x_i \partial f / \partial x_i \\
= \sum_{m' \in \Delta \cap M} \lambda_{m'}(\langle m', v_i \rangle + 1) x^{m' + m^*} \\
= \sum_{m'+m^* \in \Delta \cap M} \lambda_{m'}(\langle m', v_i \rangle + 1) I(m' \in \Delta) x^{m' + m^*} \\
= \sum_{m \in \Delta \cap M} \lambda_{m-m^*}(\langle m - m^*, v_i \rangle + 1) I(m - m^* \in \Delta) x^m.
\]

To justify the fourth line in the above calculation, note that given \( m' \in \Delta \cap M \), either \( m' + m^* \in \Delta \cap M \) or \( (m', v_i) + 1 = 0 \). Then setting \( m = m' + m^* \) leads to the last line.

Let \( \text{Int}(F_i) \cap M = \{ m_{i,s} : 1 \leq s \leq t_i; t_i \geq 0 \} \). Then \( J(f)_{\beta_0} = \text{span}\{ x_k \partial f / \partial x_k, \prod_{j=1}^{r} x_j^{\langle m_{i,s}, v_j \rangle} x_i \partial f / \partial x_i : 1 \leq k \leq n+1, 1 \leq i_s \leq t_i, i = 1, \ldots, r \} \). We want to find the dimension of this complex vector
space. So we study the space of linear relations:

\[ \sum_{k} c_k x_k \frac{\partial f}{\partial x_k} + \sum_{i,s} d_{i,s} \prod_{j=1}^{r} x_j^{(m_{i,s}, v_j)} x_i \frac{\partial f}{\partial x_i} \equiv 0 \]

\[ \Leftrightarrow \]

\[ \sum_{m} \left\{ \sum_{k} c_k \lambda_m (\langle v_k, m \rangle + 1) + \sum_{i,s} d_{i,s} \lambda_{m-m_{i,s}} I(m-m_{i,s}, v_i) + 1) \right\} x^m \equiv 0 \]

\[ \Leftrightarrow \]

\[ \sum_{k} c_k \lambda_m (\langle v_k, m \rangle + 1) + \sum_{i,s} d_{i,s} \lambda_{m-m_{i,s}} I(m-m_{i,s}, \Delta) \langle m, v_i \rangle + 1) \equiv 0 \]

for each \( m \in \Delta \cap M, \) [note: \( \langle m_{i,s}, v_i \rangle = -1 \).

This is a system of \( l(\Delta) \) number of linear equations in \( \gamma = n + 1 + \sum_{i=1}^{r} l^*(F_i) \) variables namely \( c_k, d_{i,s} \). Note that \( l(\Delta) \geq \gamma \). We shall find a nonsingular subsystem of rank \( \gamma \).

To do so pick \( n \) linearly independent vertices \( m_1, \ldots, m_n \) of \( \Delta \) and let \( m_{n+1} = 0 \), the origin. Then from the above system we pick the equations corresponding to \( m = m_1, \ldots, m_{n+1} \) and \( m = m_{i,s} : i = 1, \ldots, r; 0 \leq i_s \leq t_i \). Denote this \( \gamma \times \gamma \) system by \( (***) \). It can be written as:

\[
\begin{bmatrix}
P & A \\
B & Q
\end{bmatrix}
\begin{bmatrix}
c \\
d
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
0
\end{bmatrix}
\]

where

\[
P = 
\begin{bmatrix}
\lambda_{m_1}(\langle m_1, v_1 \rangle + 1) & \ldots & \lambda_{m_1}(\langle m_1, v_{n+1} \rangle + 1) \\
\vdots & \ddots & \vdots \\
\lambda_{m_n}(\langle m_n, v_1 \rangle + 1) & \ldots & \lambda_{m_n}(\langle m_n, v_{n+1} \rangle + 1) \\
0 & \ldots & 0
\end{bmatrix}
\]

\[
Q = 
\begin{bmatrix}
\lambda_{m_{1,1}-m_{1,1}} I(\langle m_{1,1}-m_{1,1}, v_1 \rangle + 1) & \ldots & \lambda_{m_{1,1}-m_{r,1}} I(\langle m_{1,1}-m_{r,1}, v_1 \rangle + 1) \\
\vdots & \ddots & \vdots \\
\lambda_{m_{r,1}-m_{1,1}} I(\langle m_{r,1}-m_{1,1}, v_1 \rangle + 1) & \ldots & \lambda_{m_{r,1}-m_{r,1}} I(\langle m_{r,1}-m_{r,1}, v_1 \rangle + 1)
\end{bmatrix}
\]

Observe that all the diagonal entries of \( Q \) are \( \lambda_0 \), and none of its off-diagonal entries has \( \lambda_0 \). Also any entry of \( A \) is of the form \( \lambda_{m_k-m_{i,s}} I(\cdot) \) and hence does not involve \( \lambda_0 \). Similarly an entry of \( B \) is of the form \( \lambda_{m_{i,s}} (\langle m_{i,s}, v_k \rangle + 1) \) and so does not have \( \lambda_0 \).

Consider the determinant of the coefficient matrix \( \begin{bmatrix} P & A \\ B & Q \end{bmatrix} \) as a polynomial in the \( \lambda_i \)’s. Then the term of this determinant having the highest power of \( \lambda_0 \) is \( (\lambda_0)^{\sum l^*(F_i)} \det P \). We will show below that \( \det P = \text{nonzero} \).
constant times $\lambda_{m_1}\ldots\lambda_{m_n} \lambda_0$. Thus the determinant of the coefficient matrix of the system $(**)$ is a nontrivial polynomial in the $\lambda$’s and is therefore nonzero for generic choice of the $\lambda$’s. Hence $J(f)_{\beta_0}$ has rank $\gamma$ as a complex vector space, for a generic $f \in L(\Delta)$. Since $S_{\beta_0} \simeq L(\Delta)$, so $(S/J(f))_{\beta_0}$ has rank $l(\Delta) - \gamma$, for a generic $f$.

**Lemma 4.** The $(n+1) \times (n+1)$ matrix $P = (( P_{i,j} = \lambda_{m_i}(\langle m_i, v_j \rangle + 1 )))$ is nonsingular for generic choice of $\lambda$’s.

**Proof.** Let $E$ be the $(n+1) \times (n+1)$ matrix $(( E_{i,j} = (\langle m_i, v_j \rangle + 1 )))$. Then $\det P = \lambda_{m_1} \ldots \lambda_{m_{n+1}} \det E$. We claim that $E$ is nonsingular. Otherwise there exists a nonsingular vector $(c_1, \ldots, c_{n+1})$ such that $\sum_{k=1}^{n+1} c_k (\langle m_i, v_k \rangle + 1) = 0$ for all $i = 1, \ldots, n+1$. In particular, for $i = n+1$ we get $\sum_{k=1}^{n+1} c_k = 0$. This implies $\sum_{k=1}^{n+1} c_k (m_i, v_k) = 0$ for all $i = 1, \ldots, n$. Since $m_1, \ldots, m_n$ are linearly independent, this would imply that $\sum_{k=1}^{n+1} c_k (m, v_k) = 0$ for all $m \in \Delta$. Therefore $\sum_{k=1}^{n+1} c_k v_k = 0$. This combined with $\sum c_k = 0$ implies that $\sum_{k=1}^{n+1} c_k \tilde{v}_k = 0$ which contradicts the linear independence of $\tilde{v}_1, \ldots, \tilde{v}_{n+1}$. Thus the lemma holds. □

So we have the following theorem:

**Theorem 4.** For any generic nondegenerate anticanonical hypersurface $V$ of an $n$-dimensional simplicial Fano toric variety $\mathbb{P}_\Delta$, and $n \geq 4$,

\[
h^{n-2,1}(V) = l(\Delta) - n - 1 - \sum_{\dim(F) = n-1} l'(F).
\]

5.3. Cohomology of the twisted sectors of $V$. We now want to compute $h^{n-3,0}(V_{(g)})$ for any twisted sector $V_{(g)} \cong V \cap \mathbb{P}_{\widetilde{F}_0}$. This is obviously zero if $\dim(F^0) > 1$, since $\dim(\mathbb{P}_{\widetilde{F}_0}) = n - 1 - \dim(F^0)$. So we will only consider the case $\dim(F^0) = 1$. Let $\tau$ be the 2-dimensional cone obtained by coning over $F^0$. As noted earlier $\overline{\tau} = \mathbb{P}_{\overline{F}_0}$. The restriction of $V$ to $\overline{\tau}$ gives a quasi-smooth ample hypersurface of $\overline{\tau}$, which we shall identify with $V_{(g)}$. So we are again in a situation where we can invoke Lemma 2.

For this we need to understand the homogeneous coordinate ring $S'$ of $\overline{\tau}$. According to Fulton [F], Section 3.1, a fan for $\overline{\tau}$ can be constructed from the fan $\Xi$ of $X$ as follows.

Let $N_\tau$ be the sublattice of $N$ generated by the primitive one dimensional generators of $\tau$. Let $N(\tau) = N/N_\tau$. The dual lattice of $N(\tau)$ is given by $M(\tau) = \tau^\perp \cap M$. The star of a cone $\tau$ can be defined abstractly as the set of cones $\sigma$ in $\Xi$ that contain $\tau$ as a face. Such cones $\sigma$ are determined by their images in $N(\tau)$ i.e., by $\overline{\sigma} = (\sigma + (N_\tau)) / (N_\tau) \subset N(\tau) / (N_\tau) = N(\tau)_\mathbb{R}$. These cones $\{ \overline{\sigma} : \tau \prec \sigma \}$ form a fan $\overline{\text{Star}}(\tau)$ in $N(\tau)$. $\overline{\tau}$ is the toric variety corresponding to this fan. Without loss of generality, let $v_1, v_2$ be the generators of $\tau$. The corresponding Weil divisors in $X$ are $D_1$ and $D_2$. 


Assume that $v_j, j = 3, \ldots, l$ are the 1-dimensional cones of $\Xi$ such that 
\{v_1, v_2, v_j\} generate a 3-dimensional cone of $\Xi$. In other words, $\bar{\sigma}_j, j = 3, \ldots, l$ are the 1-dimensional cones of $\text{Star}(\tau)$. Let $\bar{\sigma}_j := D_1D_2D_j$ for $j = 3, \ldots, r$. Note that $\bar{\sigma}_j = 0$ if $j > l$. The divisor of $\bar{\sigma}_r$ corresponding to $\bar{\sigma}_j$ is $\bar{\sigma}_j$ for $j = 3, \ldots, l$. So the homogeneous coordinate ring $S'$ is generated by variables $y_j$ corresponding to $\bar{\sigma}_j$ for $j = 3, \ldots, l$. Denote by $\alpha_0$ the anticanonical class $\sum_{j=3}^l \bar{\sigma}_j$ in $S'$.

On the other hand the divisor $V$ restricts to $-K_XD_1D_2 = (D_1 + \cdots + D_r)D_1D_2 = \sum_{j=3}^l \bar{\sigma}_j + (D_1 + D_2)D_1D_2$. To see what $(D_1 + D_2)D_1D_2$ is in terms of the $\bar{\sigma}_j$s, we can pick a point $m \in \bar{F}^0 \cap M$ and let $b_i := \langle m, v_i \rangle$, $1 \leq i \leq r$. Then $\sum_{i=1}^r b_iD_i$ is linearly equivalent to zero. Note that $b_1 = b_2 = -1$. Hence we have $D_1 + D_2 = \sum_{i=3}^l b_iD_i$. So, $(D_1 + D_2)D_1D_2 = \sum_{i=3}^l b_iD_i$. Let $\alpha$ be the class in $S'$ representing the effective ample divisor $-K_XD_1D_2$. Let $f'$ be the associated homogeneous polynomial in the $y_j$s. Now we can apply Lemma 2 to the $(n - 2)$-dimensional variety $\bar{\sigma}_r$ and the ample hypersurface $V(g)$. Choose $k = 1$ in the lemma to get

$$(S'/J(f'))_{\alpha - \alpha_0} \simeq PH^{n-3,0}(V(g)) = H^{n-3,0}(V(g))$$

since $H^{n-3,0}(\bar{\sigma}_r) = 0$.

Now $\alpha - \alpha_0 = [\alpha - \sum_{j=3}^l \bar{\sigma}_j]$. A typical generator $\partial f'/\partial y_i \in J(f')$ has degree $[\alpha - \bar{\sigma}_j]$. There are no nonconstant regular functions on the projective variety $\bar{\sigma}_r$. So any nontrivial effective divisor, and in particular $\sum_{j=3,\ldots, l} \bar{\sigma}_j$ and $\sum_{j=3,\ldots, l} \bar{\sigma}_j$ are not linearly equivalent to zero. This implies that $J(f')_{\alpha - \alpha_0} = 0$. Hence we obtain that

$$(S')_{\alpha - \alpha_0} \simeq H^{n-3,0}(V(g)).$$

Now $\alpha - \alpha_0 = [\sum_{j=3}^l b_j \bar{\sigma}_j]$. We want to identify the effective divisors in this class. So we want $m_* \in M(\tau)$ such that $\sum_{j=3}^l (b_j + \langle m_*, \bar{\sigma}_j \rangle) \bar{\sigma}_j$ is effective. This is if and only if $(b_j + \langle m_*, \bar{\sigma}_j \rangle) \geq 0$ for all $j = 3, \ldots, l$ $\iff (\langle m + m_*, v_j \rangle) \geq 0 > -1$ for $j = 3, \ldots, l$ $\iff m + m_* \in \text{Int}(\bar{F}^0) \cap M$.

To justify the last step note that $\langle m + m_*, v_i \rangle = -1$ for $i = 1, 2$.

Since $m$ is fixed, the required effective divisors are in one-to-one correspondence with the interior lattice points of $\bar{F}^0$. Hence $H^{n-3,0}(V(g)) = l^*(\bar{F}^0)$. Since there are $l^*(\bar{F}^0)$ twisted sectors isomorphic to $V \cap \mathbb{P}_{\bar{F}^0}$ we have the following:

$$h_{\text{orb}}^{n-2,1}(V) = h^{n-2,1}(V) + \sum_{i(g)=1} h^{n-3,0}(V(g))$$
\[ l(\Delta) - n - 1 - \sum_{\dim(F) = n-1} l^*(F) + \sum_{\dim(F^\circ) = 1} l^*(F^\circ)l^*(\hat{F}^\circ) \]

\[ = l(\Delta) - n - 1 - \sum_{\dim(F) = n-1} l^*(F) + \sum_{\dim(F) = n-2} l^*(F)l^*(\hat{F}). \]

For the last step we used the one-to-one correspondence between faces of \( \Delta \) and \( \Delta^\circ \).

5.4. Main results. Combining the above formula with Theorem 4 we have the following theorem:

**Theorem 5.** For any generic nondegenerate anticanonical hypersurface \( V \) of an \( n \)-dimensional simplicial Fano toric variety \( \mathbb{P}_\Delta \), \( n \geq 4 \),

\[ h^{n-2,1}_{\text{orb}}(V) = l(\Delta) - n - 1 - \sum_{\dim(F) = n-1} l^*(F) + \sum_{\dim(F) = n-2} l^*(F)l^*(\hat{F}). \]

**Corollary 1.** If \( \hat{V} \) is an MPCP desingularisation of any generic nondegenerate anticanonical hypersurface \( V \) of an \( n \)-dimensional simplicial Fano toric variety \( \mathbb{P}_\Delta \), \( n \geq 4 \), then \( h^{p,1}_{\text{orb}}(V) = h^{p,1}_{\text{orb}}(\hat{V}) \) for \( p = 1 \) and \( p = n - 2 \).

**Proof.** The formulas for \( h^{p,1}_{\text{orb}}(\hat{V}) \) for \( p = 1, n - 2 \) computed in [B1] by Batyrev match the orbifold Hodge numbers for \( V \) obtained in Theorem 3 and Theorem 5. \( \square \)

**Corollary 2.** In the case \( n = 4 \), \( h^{p,q}_{\text{orb}}(V) = h^{p,q}_{\text{orb}}(\hat{V}) \) for any \( p \) and \( q \).

**Proof.** We need only consider \( p, q \leq 3 \). Also \( h^{p,0}_{\text{orb}} \equiv h^{p,0} \) by definition since \( \iota \) is nonnegative. So, by Serre duality for ordinary and orbifold cohomologies, it is enough to consider just the cases \( p = 1, q = 1 \) and \( p = 2, q = 1 \). These are addressed by Corollary 1. (We should remark here that in this case \( \hat{V} \) is actually smooth.) \( \square \)

**Corollary 3.** If \( \mathbb{P}_{\Delta^\circ} \) is also simplicial, and \( V^\circ \) is a generic nondegenerate anticanonical hypersurface of \( \mathbb{P}_{\Delta^\circ} \), then \( h^{1,1}_{\text{orb}}(V) = h^{n-2,1}_{\text{orb}}(V^\circ) \) and vice versa.

**Proof.** Follows from interchanging the roles of \( \Delta \) and \( \Delta^\circ \) in the formulas. \( \square \)

**Remark.** In particular, for the \( n = 4 \) case, we have \( h^{p,q}_{\text{orb}}(V) = h^{3-p,q}_{\text{orb}}(V^\circ) \). This is an example of ‘mirror symmetry’ of orbifold hodge numbers.

5.5. An example. This example first appeared in the Greene-Plesser mirror construction [GP] and was also studied in [COFKM] in the context of mirror symmetry.

Consider the complex 4-dimensional weighted projective space \( X = \mathbb{P}(1,1,2,2,2) \). It is a simplicial Fano toric variety. Its fan \( \Xi \) has the following
1-dimensional cones in $N \cong \mathbb{Z}^4$: $v_1 = (-1, -2, -2, -2)$, $v_2 = (1, 0, 0, 0)$, $v_3 = (0, 1, 0, 0)$, $v_4 = (0, 0, 1, 0)$, $v_5 = (0, 0, 0, 1)$. $\Xi$ has five 4-dimensional cones, obtained by dropping one of the $v_i$’s at a time and taking the cone generated by the remaining four.

Let $D_i$ denote the torus-invariant divisor given by the orbit closure $\overline{O_{v_i}}$.

It is easy to check that in $A_3(X)$, $[D_2] = [D_1]$ and $[D_i] = 2[D_1]$ for $i \geq 3$.

Construct the homogeneous coordinate ring of $X$ by introducing variables $x_i$ corresponding to $v_i$. Then $\deg(x_1) = \deg(x_2) = 1$ ($= [D_1]$) and $\deg(x_i) = 2$ for $i \geq 3$.

This leads to the more familiar description of $\mathbb{P}(1,1,2,2,2)$ as $(\mathbb{C}^5 - \{0\})/\mathbb{C}^*$. The action of any $\alpha \in \mathbb{C}^*$ on $\mathbb{C}^5 - \{0\}$ is as follows:

$$\alpha \cdot [x_1 : x_2 : x_3 : x_4 : x_5] = [\alpha x_1 : \alpha x_2 : \alpha^2 x_3 : \alpha^2 x_4 : \alpha^2 x_5].$$

In this description, $D_i$ corresponds to the hyperplane $\{x_i = 0\}$ and the 4-dimensional cones of $\Xi$ correspond to the open sets $\{x_i \neq 0\}$. It is also easily seen that the singular locus of $X$ is precisely the surface $\{x_1 = x_2 = 0\}$. In fact, this represents the only twisted sector of $X$. The $v_i$’s are the vertices of a reflexive polytope $\Delta^\circ$. The faces of $\Delta^\circ$ have only one interior lattice point: $(0, -1, -1, -1) = \frac{1}{2}(v_1 + v_2)$. This lattice point corresponds to the twisted sector and the local isotropy group is $\mathbb{Z}_2$.

The dual reflexive polytope $\Delta$ in $M_\mathbb{R}$ has the following vertices:

$$w_1 = (-1, -1, -1, -1), \quad w_2 = (7, -1, -1, -1), \quad w_3 = (1, 3, -1, -1),$$
$$w_4 = (-1, -1, 3, -1), \quad w_5 = (-1, -1, -1, 3).$$

$\Delta$ is the polytope corresponding to the anticanonical divisor $-K_X = \sum_{i=1}^{5} D_i$ of $X$, and $X = \mathbb{P}_\Delta$. If $V$ is a generic nondegenerate Calabi-Yau (anticanonical) hypersurface of $X$, then $V$ has just one twisted sector namely $C = V \cap \{x_1 = x_2 = 0\}$. One can directly compute the genus of this curve $C$ by using the Riemann-Hurwitz formula. It turns out to be 3. $V$ has the Hodge numbers: $h^{1,0} = h^{2,0} = 0$, $h^{3,0} = 1$, $h^{1,1} = 1$, $h^{2,1} = 83$. Since the degree shifting number of the twisted sector $C$ is 1, we compute $h^{1,1}_{\text{orb}}(V) = h^{1,1}(V) + h^{0,0}(C) = 1 + 1 = 2$, and $h^{2,1}_{\text{orb}}(V) = h^{2,1}(V) + h^{1,0}(C) = 83 + 3 = 86$.

The dual Fano variety $\mathbb{P}_\Delta^\circ$ is also simplicial. This is easily checked since its fan is is obtained by coning over the faces of $\Delta$. In fact, $\mathbb{P}_\Delta^\circ = \mathbb{P}_\Delta / \mathbb{Z}_4^\circ$. This is also shown easily. First, observe that $w_1 = -w_2 - 2w_3 - 2w_4 - 2w_5$.

Acknowledgments. This work was done under the supervision of my doctoral advisor at UW-Madison, Yongbin Ruan. I am indebted to him for his guidance. My special thanks are also due to Lev Borisov, David Cox and Igor Dolgachev for answering my questions. I thank Prof. Cox for suggesting many improvements, and Anvar Mavlyutov for his comments.
References


[Re] M. Reid, La correspondance de McKay, in Séminaire Bourbaki, exposé, 867, Novembre 1999, also at math.AG/9911165, CMP 1 886 756.

[Ru] Y. Ruan, Stringy geometry and topology of orbifolds, math.AG/0011149.


Received October 18, 2001 and revised May 6, 2002.

DEPARTMENT OF MATHEMATICS
MICHIGAN STATE UNIVERSITY
E. LANSING, MI 48824-1027
E-mail address: poddar@math.msu.edu
The aim of this paper is to discuss some applications of the relation between Seiberg-Witten theory and two natural norms defined on the first cohomology group of a closed 3-manifold $N$ — the Alexander and the Thurston norm. We will start by giving a “new” proof, applying SW theory, of McMullen’s inequality between these two norms, and then use these norms to study two problems related to symplectic 4-manifolds of the form $S^1 \times N$. First we will prove that — as long as $N$ is irreducible — the unit balls of the Thurston and Alexander norms are related in a way that is similar to the case of fibered 3-manifolds, supporting the conjecture that $N$ has to be fibered over $S^1$. Second, we will provide the first example of a 2-cohomology class on a symplectic manifold (of the form $S^1 \times N$) that lies in the positive cone and satisfies Taubes’ “more constraints”, but cannot be represented by a symplectic form, disproving a conjecture of Li and Liu (Li-Liu, 2001, Section 4.1).

1. Introduction.

It has been proven, in [McM], that two natural (semi) norms defined on the first cohomology group of a 3-manifold, namely the Alexander norm $\| \cdot \|_A$, defined from the Alexander polynomial of the manifold, and the Thurston norm $\| \cdot \|_T$, defined in terms of the minimal genus of the representatives of the Poincaré dual two dimensional homology class, satisfy a relation expressed in the following:

**Theorem 1.1** (McMullen). Let $N$ be a compact, connected, oriented 3-manifold (eventually with boundary a union of tori); then the Alexander and Thurston norm satisfy

\[
\| \cdot \|_A \leq \| \cdot \|_T + \begin{cases} 
(1 + b_3(N)) \text{div}(\cdot) & \text{if } b_1(N) = 1 \\
0 & \text{if } b_1(N) > 1,
\end{cases}
\]

where $\text{div}(\cdot)$ denotes the divisibility of an element in $H^1(N, \mathbb{Z})$. 

169
This inequality, applied to the particular case where the three manifold is the exterior of a knot \( K \), reduces to the well-known fact that the degree of the Alexander polynomial of the knot (i.e., the difference between highest and lowest power) is bounded from above by twice the genus of the knot, i.e., the lowest value of the genus of a Seifert surface of the knot.

These two norms turn out to be strictly related to the 3-dimensional Seiberg-Witten theory.

The proof of Theorem 1.1 given in [McM] is purely topological, but it is suggested the existence of a proof based on SW theory. (It appears that the first one to observe this has been D. Kotschick; P. Kronheimer previously proved the inequality in the case of \( N \) obtained as 0-surgery of a knot, in [K2]; the first detailed proof of the general case appeared in a preprint of the author ([V]), on which this paper is partly based.)

Our first aim will be to write the two norms in terms of SW basic and monopole classes for \( N \). This allows, as mentioned, an alternative proof of Theorem 1.1, that we will work out for the case, for us more interesting, of a closed manifold with \( b_1(N) > 1 \).

Then we will use these results to study symplectic 4-manifolds of the form \( S^1 \times N \), for an irreducible \( N \). We will prove the following:

**Theorem 1.2.** Let \( N \) be an irreducible 3-manifold with \( b_1(N) > 1 \) such that \( S^1 \times N \) admits a symplectic structure \( \omega \); then there exists a face \( F_T \) of the unit ball of the Thurston norm contained in a face \( F_A \) of the unit ball of the Alexander norm.

This quite peculiar property is satisfied by fibered 3-manifolds, and supports the conjecture that a 3-manifold \( N \) such that \( S^1 \times N \) admits a symplectic structure must in fact be fibered.

Theorems 1.1 and 1.2 hold true also in the case of \( b_1(N) = 1 \). It is not surprising that the proof is technically quite longer, due to the chamber structure of the SW invariants in that case. We will omit this case here, referring the interest reader to [V], where it is treated in detail.

We will then address the problem of determining the constraints for a cohomology class \( \alpha \) on a symplectic 4-manifold to be represented by a symplectic form \( \omega \). It is clear that such an \( \alpha \) must have positive square, and its pairing with the SW basic classes must satisfy the constraints determined in [T2]. Li-Liu have conjectured in [LL]\(^1\) that these are sufficient conditions. We will prove the following:

**Theorem 1.3.** There exists a symplectic 4-manifold of the form \( S^1 \times N \) and a cohomology class \( \alpha \) of positive square satisfying Taubes’ “more constraints” which can not be represented by a symplectic form.

\(^1\) Added in proof: The conjecture appears in the preliminary version of [LL] (see http://www.arxiv.org/abs/math.SG/0012048v1).
Hence, Li-Liu conjecture is false.

2. Alexander and Thurston norms.

We start by briefly recalling the definition of Alexander and Thurston norms on the first cohomology group of a closed, oriented 3-manifold $N$. Denote by $F$ the free abelian group $F := H_1(N, \mathbb{Z})/\text{Tor}$; by definition, $\text{rk}(F) = b_1(N)$. The Alexander polynomial of $N$ is an element of the group ring $\mathbb{Z}[F]$, i.e., a finite sum

$$\Delta_N = \sum_i a_i t^i$$

where $i = (i_1, \ldots, i_{b_1(N)})$ is a multi-index of cardinality $b_1(N)$, $t = (t_1, \ldots, t_{b_1(N)})$ with $\{t_i\}$ a basis of $F$ and $a_i$ are integer coefficients. The Alexander polynomial is well-defined up to multiplication by units of $\mathbb{Z}[F]$. For any element $\phi \in H^1(N; \mathbb{Z})$ we define the norm

$$\|\phi\|_A := \max_{ij} \phi(t^i \cdot t^{-j}),$$

where the indexes run over all $i, j$ such that $a_i, a_j$ are nonzero. It is clear that this definition is unaffected by the indeterminacy in the Alexander polynomial and does not depend on the coefficients.

The Thurston norm, described in [Th2], is defined as follows: For any Riemann surface $\Sigma$ embedded in $N$ denote

$$\chi_-(\Sigma) = \sum_{\Sigma_i, \gamma(\Sigma_i) \geq 1} (-\chi(\Sigma_i)),$$

where $\Sigma$ is the disjoint union of the $\Sigma_i$; we then define the norm

$$\|\phi\|_T = \min\{\chi_-(\Sigma) | \Sigma \hookrightarrow N, \text{PD}[\Sigma] = \phi\}.$$

It is not difficult to verify that both norms are linear on rays and satisfy the triangle inequality. It is possible to continuously extend these norms to cohomology with real coefficients. The unit ball of these norms is then a finite, convex (possibly noncompact) polyhedron. In particular, the unit ball of the Alexander norm is by construction dual (up to a factor 2) to the Newton polyhedron of $\Delta_N$.

3. Basic classes and monopole classes.

In this section we will discuss the way the Alexander and Thurston norms are related to Seiberg-Witten theory. Essentially the relation between Alexander norm and SW theory will be deduced from Meng-Taubes proof of the equivalence of a SW invariant of a 3-manifold and the Alexander polynomial of the manifold. The relation of Thurston norm and SW theory, instead, has been analyzed in [KM].
We start with a brief review of SW theory in dimension three, in order to have a formulation which is the suitable for our purposes. Let \((N, g)\) be a smooth, closed, oriented, riemannian three dimensional manifold. We will assume that \(b_1(N) > 1\). We equip \(N\) with the canonical homology orientation induced by a basis of \(F\). Once \(N\) is endowed with a spin\(^c\)-structure \(P_N\), i.e., a \(U(1)\)-lifting of the \(SO(3)\) frame bundle, we can consider the three dimensional SW equations

\[
F_A = q(\psi) - i\eta, \quad \bar{\partial}_A \psi = 0,
\]

where \(A\) is a connection on the determinant bundle of the spin\(^c\)-structure, \(q(\cdot)\) is an \(\Omega^2(N; i\mathbb{R})\)-valued bilinear form on the sections of the spinor bundle associated to \(P_N\), \(\eta\) is a perturbation term that lives in \(\Omega^2(N; \mathbb{R}) \cap \ker d\), and \(\bar{\partial}_A\) is the Dirac operator that acts on spinors. These equations are invariant under the gauge group of those automorphisms of \(P_N\) which act trivially on the frame bundle. This group acts freely away from reducible couples, that we can remove suitably choosing good perturbations. It is possible to prove (see e.g., \([MT]\)), using standard techniques, that choosing a generic nonexact perturbation the moduli space of solutions of Equation (6), modulo gauge equivalence, is a 0-dimensional compact, oriented, smooth manifold; under change of the metric and perturbation (as \(b_1 > 1\)) different moduli spaces are moreover cobordant. We denote by \(\mathcal{M}(P_N, g, \eta)\) the moduli space of solutions of Equations (6), omitting the arguments whenever unnecessary.

We define the SW invariant for \(P_N\) as the algebraic sum of the oriented points of \(\mathcal{M}(P_N, g, \eta)\) for \(\eta\) a good perturbation. We have the following definition:

**Definition 3.1.** Let \(c \in H^2(N; \mathbb{Z})\) be an integral cohomology class that arises as first Chern class of a spin\(^c\)-structure \(P_N\) such that the invariant \(\text{SW}(P_N)\) is nonzero. Then \(c\) is called a basic class of \(N\).

It is quite clear from this definition that the SW equations for a basic class admit a solution for any metric and a generic perturbation. Moreover, as the compactness of the equations implies that non-emptiness is an open condition, also the unperturbed equations have a solution for any metric, i.e., \(\mathcal{M}(P_N, g, 0) \neq \emptyset\) (note that this space can be nonsmooth). This makes it natural to introduce the:

**Definition 3.2** (Kronheimer-Mrowka). Let \(c \in H^2(N; \mathbb{Z})\) an integral cohomology class that arises as Chern class of a spin\(^c\)-structure \(P_N\) such that \(\mathcal{M}(P_N, g, 0) \neq \emptyset\) for any metric \(g\). Then \(c\) is called a monopole class.

From the previous observation, the set of monopole classes, that we denote by \(\mathcal{C}(N)\), contains all the basic classes.

We now introduce, following ref. \([MT]\), an element in \(\mathbb{Z}[[F]]\), defined from the family of SW invariants of the spin\(^c\)-structures.
The set \( S \) of \( \text{spin}^c \)-structures on \( N \) is an affine \( H^2(N;\mathbb{Z}) \). There is a natural way to define a map from \( S \) to \( F \), constructed as follows: Fix a reference \( \text{spin}^c \)-structure \( Q_N \), that we choose to be the product structure. Any other structure \( P_N \) differs from it by the action of an element of \( H^2(N;\mathbb{Z}) \).

Consider now the composed map

\[
H^2(N,\mathbb{Z}) \xrightarrow{\text{PD}} H_1(N,\mathbb{Z}) \xrightarrow{\pi} F.
\]

Using this map we can construct a map \( s \) which goes from \( S \) to \( F \). The fiber of this map is given by the order of the torsion of \( H_1(N,\mathbb{Z}) \), that we will denote now on by \( \text{ord}(N) \). Note that twice this map gives, up to torsion, the Poincaré dual of the Chern classes of \( P_N \). Consider for any \( t_i \in F \) the set \( s^{-1}(t_i) \subset S \). These are the \( \text{spin}^c \)-structures that have the same real Chern class. Define now

\[
\text{SW}(t_i) := \sum_{s^{-1}(t_i)} \text{SW}(P_N);
\]

we can now define from this the function

\[
\text{SW}(N) = \sum_i \text{SW}(t_i)t_i \in \mathbb{Z}[[F]].
\]

Well-known facts of SW theory are that the number of \( \text{spin}^c \)-structures for which unperturbed SW equations admit solutions is bounded, and that the invariant \( \text{SW}(t^i) \) is symmetric under the natural involution of \( F \). These observations, together with the definition of the function \( \text{SW} \), yield the fact that \( \text{SW}(N) \) is a symmetric element of \( \mathbb{Z}[F] \).

The previous definition, in the case where \( \text{ord}(N) = 1 \), is a simple reformulation of SW theory. In the other cases, instead, they define a kind of “average” over all structures which have the same real Chern class. We can introduce a new definition that is quite practical for treating the information on \( \text{spin}^c \)-structures contained in the SW functions of Equation (9). For any element \( \gamma \in H^2(N,\mathbb{Z}) \), we denote by \( \gamma^F \) its projection to \( H^2(N,\mathbb{Z})/\text{Tor}(= F^{\text{PD}}) \).

**Definition 3.3.** Let \( c \in H^2(N,\mathbb{Z})/\text{Tor} \) be a cohomology class such that

\[
\sum_{c^F(P_N) = c} \text{SW}(P_N) \neq 0.
\]

Then \( c \) is called an a-basic class (where the “a” stands for averaged).

We have the following inclusions:

\[
\mathcal{A}(N) = (\text{a-basic classes}) \subset (\text{basic classes})^F \subset (\text{monopole classes})^F = \mathcal{C}(N)^F.
\]

We want to relate now a-basic classes with the SW function \( \text{SW}(N) \): Let \( c \) be an a-basic class; then the sum appearing in Equation (10) coincides with
SW(t^i) where t^i is defined by the relation t^{2i} = PD(c), and the invariant SW(t^i) is nonzero.

4. Relation between the norms.

Our aim now is to relate a-basic classes of $N$ with its Alexander polynomial, and then to the Alexander norm. In this section we will give a proof of the following:

**Proposition 4.1.** Let $N$ be a closed three manifold with $b_1(N) > 1$; then the Alexander norm of an element $\phi \in H^1(N; \mathbb{Z})$ is given by

\[
\|\phi\|_A = \max_{A(N)} (c \cdot \phi)
\]

where the maximum is taken over all a-basic classes of $N$.

**Proof.** The basic ingredient for the proof is provided by the theorem of Meng and Taubes which identifies the SW function with the (sign-refined) Reidemeister-Franz torsion introduced by Milnor. This is related, on its own, to the (sign-refined) symmetrized Alexander polynomial, denoted by $\Delta^s_N$. More precisely, we have:

**Lemma 4.2** (Meng-Taubes, Turaev). Let $N$ be a closed three manifold equipped with its canonical homology orientation, with $b_1(N) > 1$; then we have, in $\mathbb{Z}[F]$,

\[
\text{SW}(N) = \Delta^s_N.
\]

From this relation, considering the definitions in Equations (2) and (9), we have $\text{SW}(t^i) = a_i$ and, by Definition 3.3, a-basic classes $c_i \in A(N)$ are twice the Poincaré duals of elements of $F$ with nonvanishing invariant $\text{SW}(t^i)$, i.e.,

\[
c_i \in H^2(N; \mathbb{Z})/\text{Tor} \text{ is a-basic } \iff a_i \neq 0 \text{ where } t^{2i} = \text{PD}(c_i).
\]

The use of the relations of Equation (13) allows us to write the Alexander norm in terms of a-basic classes: We can write

\[
\|\phi\|_A = \max_{ij} \phi(t^i \cdot t^{-j}) = \max_{A(N)} (c \cdot \phi).
\]

The latter equality comes as follows: For any fixed couple $i,j$ with nonzero coefficient we have

\[
\phi(t^i \cdot t^{-j}) = \phi(t^i) - \phi(t^j) \leq \max_j (|\phi(t^{2i})|, |\phi(t^{2j})|) \leq \max_k |\phi(t^{2k})|
\]

where $k$ ranges among all indexes with nonzero coefficient. Being $\Delta^s_N$ symmetric, equality in Equation (16) is attained for some choice of index with $j = -i = \pm k$. By Equation (14), we can thus write $\phi(t^{2k}) = (c_k \cdot \phi)$ (and remove the absolute value). This completes the proof of our statement. □
The content of Proposition 4.1 is the good one for our purpose, because of the results of [KM] on the relations between monopole classes and Thurston norm: We have now all we need to prove Theorem 1.1.

First, we observe that we can restrict the proof to the case of an irreducible manifold. In fact, the SW polynomial vanishes for the connected sum of manifolds with $b_1 > 0$, and the set of a-basic classes of a manifold is preserved for connected sum with a rational homology sphere (the SW polynomial gets multiplied by the order of the torsion of the rational homology sphere).

For irreducible manifolds we have now the equality expressed in:

**Theorem 4.3** (Kronheimer-Mrowka). Let $N$ be a manifold as above: Then the Thurston norm of a class $\phi \in H^1(N, \mathbb{Z})$ is given by

$$\|\phi\|_T = \max_{C(N)} (c \cdot \phi). \quad (17)$$

Putting together the inclusion $\mathcal{A}(N) \subset \mathcal{C}(N)^F$, Theorem 4.3 and Proposition 4.1, we deduce the inequality (1) for any closed three manifold (for the sole purpose of proving inequality 1.1, it is sufficient at this point to prove an inequality in Equation (17) for all basic classes, as in [A]).

**Remark.** Equation (15) states that the unit ball of the Alexander norm, up to a factor, is dual to the Newton polyhedron of $\text{SW}(N)$. Equation (17), in light of the content of [KM] (see also [K2]) states that the Thurston unit ball is dual, up to the same factor, to the polyhedron of "SWF($N$)", an element of $\mathbb{Z}[F]$ that can be constructed from the Seiberg-Witten-Floer invariants of $N$ (still lacking a rigorous treatment). This answers a longstanding question of Fried in [F].

5. Symplectic $S^1 \times N$.

In this section we will use the Alexander and Thurston norm to study the following conjecture:

**Conjecture 5.1** (Taubes). Let $N$ be a 3-manifold such that $S^1 \times N$ admits a symplectic structure $\omega$. Then $N$ admits a fibration over $S^1$.

We will assume again that $b_1(N) > 1$. Under this condition manifolds that fiber over $S^1$ are irreducible, and it is known that for any class $\phi \in H^1(N, \mathbb{Z})$ representing a fibration we have $\|\phi\|_A = \|\phi\|_T$ (see e.g., [McM]). Fibered classes are known to satisfy the following condition (see [Th2]): The integral points laying in the cone over a (top dimensional) face of the Thurston unit sphere have the property of being all fibered, or none does. This implies in particular that a fibered face of the unit ball of Thurston norm is contained in a face of the unit ball of the Alexander norm.

We would like to prove that the latter condition holds for any $N$ such that $S^1 \times N$ is symplectic. We will be able to do so under the further assumption
that \( N \) is irreducible; in view of the results of [McC], this is a reasonable assumption. Our proof adapts to the case of \( b_1 > 1 \) the strategy of [K1].

We observe that, as simplicity is an open condition, there is no restriction (see [D]) in assuming that the symplectic form \( \omega \) on \( S^1 \times N \) is the reduction of an integer class of \( H^2(S^1 \times N, \mathbb{Z}) \). There is a cone, in \( H^2(S^1 \times N, \mathbb{R}) \), of cohomology classes that can be represented by symplectic forms, and in this cone the set of classes which are in the image of the cohomology with rational coefficients is dense. We will be interested to have cohomology classes that lie in the image of the cohomology with integral coefficients, and eventually pass to sufficiently high multiples of the symplectic form: We will implicitly assume this whenever necessary.

We want to recall now some general results that we will apply to our case. The first is the Donaldson theorem on the existence of symplectic submanifolds ([D]). This theorem assures that there exist a connected symplectic submanifold \( H \subset S^1 \times N \) such that

\[
[H] = \text{PD}[\omega] = [S^1] \times \gamma + \tau \in H_2(S^1 \times N, \mathbb{Z})
\]

where \( \gamma \in H_1(N, \mathbb{Z}), \tau \in H_2(N, \mathbb{Z}) \) and \( \gamma \cdot \tau > 0 \) (for sake of notation we will denote all products, both on \( N \) and on \( S^1 \times N \), with a dot, the distinction being clear from the context). Denote \( \phi = \text{PD}(\tau) \in H^1(N, \mathbb{Z}); \) as a consequence of the previous discussion, the \( \phi \)'s associated to symplectic forms as in the relation of Equation (18) define a cone in \( H^1(N, \mathbb{Z}) \).

The second result is that the spin\(^c\)-structures on \( S^1 \times N \) with nontrivial SW invariants must be pull backs of spin\(^c\)-structures on \( N \) (to prove this you can use, e.g., the adjunction inequality); moreover there is an identification between the moduli spaces for a spin\(^c\)-structure \( P_N \) on \( N \) and the moduli space for the pull-back structure on \( S^1 \times N \) (that we will usually denote with the same symbol), once a suitable correspondence of the perturbation terms is set (see [OT]). This allows the identification, up to a sign determined by the choice of homology orientations, of the SW invariants associated to these moduli spaces.

The third point concerns spin\(^c\)-structures on a symplectic four manifold \((M, \omega)\) with canonical bundle \( K \). There exist, in that case, a canonical spin\(^c\)-structure that decomposes as \( \mathbb{C} \oplus K^{-1} \) (and has first Chern class equal to \(-K\)). Any other spin\(^c\)-structure can be written as \( E \oplus (K^{-1} \otimes E) \) for an \( E \in H^2(M, \mathbb{Z}) \). There are some constraints on spin\(^c\)-structures with nonvanishing invariants that arises from Taubes’ work (see [T1], [T2]). In the case of \( b_+(M) > 1 \) the canonical spin\(^c\)-structure has SW invariant \( \pm 1 \) and for any other structure \( E_i \oplus (K^{-1} \otimes E_i) \) with nonzero invariants we have \( K \cdot \omega \geq E_i \cdot \omega \geq 0 \). Equality implies, respectively, \( E_i = \mathbb{C} \) or \( E_i = K \).
This inequality translates, for the basic classes $\kappa_i := \det(E_i \oplus (K^{-1} \otimes E_i))$, in the relation
\begin{equation}
K \cdot \omega \geq |\kappa_i \cdot \omega|,
\end{equation}
with equality only for the case $\kappa_i = \pm K$. Let's apply these results to $S^1 \times N$.

First, the canonical class and all other basic classes are pull backs: There exists a preferred line bundle $K \in H^2(N, \mathbb{Z})$ (for sake of simplicity, we use the same notation on $N$) and a preferred spin$^c$-structure on $N$ of the form $C \oplus K - 1$ with SW invariant $\pm 1$ such that any other spin$^c$-structure on $N$ appears as $E_i \oplus (K^{-1} \otimes E_i)$ for $E_i \in H^2(N, \mathbb{Z})$. The structures with nonzero invariants must satisfy $K \cdot \phi \geq E_i \cdot \phi \geq 0$, the equalities implying respectively $E_i = C$ or $E_i = K$. This translates to a constraint, for the basic classes of $N$, which has the form
\begin{equation}
K \cdot \phi \geq |\kappa_i \cdot \phi|,
\end{equation}
with equality only for the case $\kappa_i = \pm K$.

Using this it is straightforward to prove the following:

**Proposition 5.2.** Let $(S^1 \times N, \omega)$ be a symplectic manifold with $b_1(N) > 1$, and denote by $\phi \in H^1(N, \mathbb{Z})$ the Künneth component of $[\omega]$; then $\|\phi\|_A = K \cdot \phi$. Moreover, $\phi$ lies in the cone over a top dimensional face of the unit ball of the Alexander norm, dual to the vertex $K$ of the Newton polyhedron of $\Delta_N$.

**Proof.** The maximum of $\kappa_i \cdot \phi$ for $\kappa_i$ basic is attained for and only for $K$. We want to use this property to evaluate the Alexander norm, in the form expressed in Proposition 4.1. To do this we need only to prove that $K$ (or, more precisely, its image $K^F$ in $H^2(N, \mathbb{Z})/\text{Tor}$) is an a-basic class. But no other basic class $\kappa_i$ can coincide up to torsion with $K$ without violating Equation (20), so that the sum of Equation (10), namely $\sum_{c^F \in \text{Tor}(P_N)} \text{SW}(P_N)$, contains only one nonzero term, that term being equal to 1. This means that $K^F$ is an a-basic class. We can conclude, following Proposition 4.1, that $\|\phi\|_A = K \cdot \phi$. The rest of the proposition is an obvious consequence of what was previously stated. \qed

We will use Proposition 5.2 to write the genus of the symplectic submanifold $H$ of Equation (18), in conjunction with the adjunction inequalities for manifolds of type $S^1 \times N$ that are contained in [K1]. These apply to irreducible manifolds $N$ which do not have a basis of $H_2(N, \mathbb{Z})$ composed of tori. Leaving aside this totally degenerate case, for which the equality of Alexander and Thurston norm is trivial, we have the following:

**Proposition 5.3.** Let $(S^1 \times N, \omega)$ be a symplectic manifold with $N$ irreducible, $b_1(N) > 1$, and denote by $\phi \in H^1(N, \mathbb{Z})$ the Künneth component of $[\omega]$: Then $\|\phi\|_A = \|\phi\|_T$. 
Proof. The adjunction inequality for embedded submanifolds of $S^1 \times N$ of \cite{K1} can be written in the form

\begin{equation}
\chi_-(H) \geq H \cdot H + \|\phi\|_T = 2\gamma \cdot \tau + \|\phi\|_T.
\end{equation}

As $H$ is symplectic, the adjunction formula for symplectic submanifolds gives

\begin{equation}
\chi_-(H) = H \cdot H + K \cdot H = 2\gamma \cdot \tau + \|\phi\|_A.
\end{equation}

These formulae are compatible with the content of Equation (1.1) if and only if $\|\phi\|_A = \|\phi\|_T$. \qed

We can somehow strengthen this result. By Equation (1.1) the unit ball of the Thurston norm is contained in the unit ball of the Alexander norm; it is clear that extending the Alexander norm to real coefficients, and using the denseness of $H^1(N, \mathbb{Q})$, the equality stated in Proposition 5.3 continues to hold in an open cone of $H^1(N, \mathbb{R})$ determined by the cone of classes of $H^2(S^1 \times N, \mathbb{R})$ admitting a symplectic representative (the norm is a continuous function). Therefore a (top dimensional) face $F_T$ of the unit ball of the Thurston norm (containing $\phi/\|\phi\|_T$) intersects a face $F_A$ of the unit ball of the Alexander norm (the face dual to $K$); but this implies that the entire $F_T$ is contained in $F_A$.

![Diagram of Alexander and Thurston unit balls](image.png)

**Figure 1.** Alexander unit ball and Thurston unit ball for a 3-manifold such that $S^1 \times N$ is symplectic.

Figure 1 shows a possible case of the relation between the norms for $N$ as described.

This observation completes the proof of Theorem 1.2.
6. Representability of a cohomology class by a symplectic form.

A classical problem of symplectic topology is to determine necessary and sufficient conditions for a 2-cohomology class \( \alpha \) on an even-dimensional closed, smooth, oriented manifold \( M \) admitting an almost complex structure with canonical class \( K \), to be represented by a symplectic form. This problem merges with the general problem of the existence of any symplectic structure on \( M \). Necessary conditions arise from the very definition of symplectic form; in particular, if \( M \) has dimension 4, we need \( \alpha \in \mathcal{P} \), where

\[
\mathcal{P} = \{ \beta \in H^2(M, \mathbb{R}) | \beta \cdot \beta > 0 \}. \tag{23}
\]

An early conjecture, in [Th1], speculated that every almost complex manifold with nonempty \( \mathcal{P} \) admitted a symplectic structure. Since then, other constraints have been identified. In particular, more refined conditions arise from Taubes’ constraints on SW basic classes: As mentioned, we must have \( \text{SW}(K) = \pm 1 \), and the class \( \alpha \) must satisfy the conditions of Equation (19). We denote by \( \mathcal{T} \) the cone composed of elements of \( \mathcal{P} \) satisfying these constraints, i.e.,

\[
\mathcal{T} := \{ \alpha \in \mathcal{P} | K \cdot \alpha \geq |\kappa_i \cdot \alpha| \}, \tag{24}
\]

with strict inequality when \( \kappa_i \neq \pm K \).

It is well-known that satisfying Taubes’ constraints is not a sufficient condition for \( \alpha \) to be represented by a symplectic form. In fact, as discussed in [KMT], if we consider the manifold \( X \# \Sigma \) where \( X \) is symplectic and \( \Sigma \) is an homology 4-sphere admitting a nontrivial cover, and the cohomology class \( \alpha_\omega \) on \( X \# \Sigma \) induced by a symplectic form \( \omega \) on \( X \) under the natural isomorphism \( H^2(X, \mathbb{R}) = H^2(X \# \Sigma, \mathbb{R}) \), we have identity of the Seiberg-Witten polynomials \( \text{SW}_X = \text{SW}_{X \# \Sigma} \) and \( \alpha_\omega \) lies in \( \mathcal{T}_{X \# \Sigma} \), but cannot be represented by a symplectic form for the simple reason that \( X \# \Sigma \) itself does not admit symplectic structures (it has a cover with trivial SW polynomial). There is another class of potential, more refined, examples of couples \((M, \alpha)\) satisfying these constraints with \( \alpha \) not representable by a symplectic form. These are knot surgery manifolds homotopic to a K3 surface (see [FS] for the definition) obtained from a knot \( K \): It is commonly conjectured that whenever \( K \) is not a fibered knot, \( M \) can not be symplectic, but it is easy to find nonfibered knots such that Taubes’ constraints are satisfied for a class \( \alpha \).

In both the previous cases, the absence of a symplectic form representing \( \alpha \) has to be attributed, in some sense, to the manifold \( M \) (which does not admit \textit{tout court} symplectic structures) and not to the cohomology class itself. We can ask about the situation for manifolds known to be symplectic. In particular, it has been conjectured (see [LL], Section 4) that if we assume that \( X \) is a symplectic manifold, the cone \( \mathcal{T} \) coincides with the “symplectic
cone''

\begin{equation}
\mathcal{W} := \{ \alpha \in \mathcal{P} | \alpha \text{ is represented by a symplectic form} \}.
\end{equation}

The conjecture gives a possible answer to the following problem, outlined in the beginning of this section, namely:

**Question.** Let \( M \) be a symplectic manifold. Determine the cone, in \( H^2(M, \mathbb{R}) \), represented by symplectic forms.

Some partial answer to this question are known. For example, Geiges proved in [G] that for \( T^2\)-bundles over \( T^2 \), all classes in the positive cone are represented by symplectic forms (we remark that all these classes satisfy Taubes' constraints, as the canonical class is trivial); it is interesting also to compare with the result of Gromov for the case of open manifolds, where any form in the positive cone lies in \( \mathcal{W} \).

Concerning this Question, we have the following result:

**Theorem 6.1.** There exist symplectic manifolds, of the form \( S^1 \times N^3 \), on which there are cohomology classes of positive square satisfying Taubes' constraints but which can not be represented by symplectic forms, i.e., the strict inclusion \( \mathcal{W} \subset \mathcal{T} \) holds true. In particular, the conjecture of [LL] is false.

**Proof.** The proof is based on the following assumption, that will be proved in the next section (Theorem 7.5): There exists a family of fibered 3-manifolds, whose generic component is denoted by \( N \), with \( H_1(N, \mathbb{Z}) = \mathbb{Z}^2 \), such that the fibered face \( F_T \) of the Thurston unit ball is strictly contained in the face \( F_A \) of the Alexander unit ball. Assuming this, we proceed as follows. Denote by \( V \) the (nonempty) cone, in \( H^1(N, \mathbb{R}) \), over \( F_A \setminus F_T \). Choose a \( \phi \in V \). We claim that we can define an \( \psi \in H^2(N, \mathbb{R}) \) such that the cohomology class \( \alpha \in H^2(S^1 \times N, \mathbb{R}) \) with Künneth decomposition

\begin{equation}
\alpha = \phi \wedge [dt] + \psi
\end{equation}

has positive square and satisfies Taubes' constraints, i.e., \( \alpha \in \mathcal{T} \). This is achieved in the following way. We have \( \alpha \cdot \alpha = 2\phi \cdot \psi \): Identify \( H^2(N, \mathbb{R}) = \text{Hom}(H^1(N, \mathbb{R}), \mathbb{R}) \); to get a positive square, we can choose \( \psi \) to be any element of the cone \( \text{Hom}(\phi, \mathbb{R}_+) \). We observed in Section 5 that the basic classes on \( S^1 \times N \) are pull-back of basic classes on \( N \); the choice of \( \psi \) is therefore irrelevant for the constraints of Equation (24) and \( \alpha \) belongs to \( \mathcal{T} \) if and only if \( \phi \) satisfies the condition (on \( H^*(N, \mathbb{R}) \)) \( K \cdot \phi \geq |\kappa_i \cdot \phi| \) with equality only for the case \( \kappa_i = \pm K \). But this condition is equivalent to the condition that \( \phi \) lies in the cone over \( F_A \), as \( F_A \) is, by definition, the face dual to \( K \), i.e., the elements lying in the cone over \( F_A \) have maximal pairing, among all basic classes, with and only with \( K \).

To complete the proof, we need to show now that \( \alpha \) can not be represented by a symplectic form. By Proposition 5.3 (and the following comments, if we want to work with cohomology with real coefficients), if \( \alpha \) admits
a symplectic representative, then its Künneth component $\phi$ should have the same Alexander and Thurston norm, something we excluded choosing $\phi \in V$.

Note that proceeding as above we can without difficulty choose the class $\alpha$ to lie in the image of cohomology with integer coefficients.

Remarks.
1. The symplectic manifold discussed in Theorem 6.1 is not simply connected, but we believe that there exist simply connected examples. In particular, we expect that the link surgery manifolds obtained using a link with fibered face strictly contained in a face of the Alexander norm (as the one we will discuss in the next section), are possible examples. The difficulty in proving such result arises from the difficulty of proving the analogue of Proposition 5.3 (see Section 7 of [K2] for a discussion of this interesting problem).

2. The failure of Li-Liu conjecture, as expressed in the examples of Theorem 6.1, is due to the mismatch between the convex hull of basic classes and the convex hull of monopole classes, as the latter determines the extension of the fibered cone of $H^1(N, \mathbb{R})$. It is conceivable to improve the conjecture, at least for symplectic manifolds of the form $S^1 \times N$, by reformulating the definition of the cone $T$ as

\[ T := \{ \alpha \in \mathcal{P} | K \cdot \alpha \geq |\kappa_i \cdot \alpha| \text{ for any monopole class } \kappa_i \}, \]

with strict inequality when $\kappa_i \neq \pm K$. In that case, as follows from the results of [Th2], the conjecture would hold true assuming the validity of a strict version of Conjecture 5.1, which takes the form:

Conjecture 6.2. Let $N$ be a 3-manifold such that $S^1 \times N$ admits a symplectic structure $\omega$; then the Künneth component of $[\omega]$ in $H^1(N, \mathbb{R})$ can be represented by a nondegenerate 1-form (i.e., it lies in a fibered cone).

7. Construction of the three manifolds.

In this section we will justify the assumption made in the proof of Theorem 6.1, namely the existence a family of closed, fibered 3-manifolds with the property that $F_T$ is strictly contained in $F_A$. Our construction will be based on the existence of a noteworthy 2-component link, exhibited by Dunfield in [Du], which has the same property. We will need the following result:

Proposition 7.1 (Dunfield). There exists a 2-component oriented link $D = D_1 \cup D_2 \subset S^3$ with Alexander polynomial

\[ \Delta_D(t_1, t_2) = (t_1 - 1)(t_2 - 1) \]
(written in terms of the homology classes of the meridians to the two components) which has a fibered face $F_T$ strictly contained in a face $F_A$, dual to the vertex $t_1 t_2^{-1}$ of the dual polyhedron.

(In Dunfield’s paper, the Alexander polynomial and the norms are discussed in terms of an homology basis different from ours, but it is easy to rewrite them in terms of the standard homology basis for the link exterior, as above.)

We don’t know the exact shape of the Thurston unit ball, but for our purpose it is enough to know the result contained in the previous Proposition. Denote $N_D = S^3 \setminus \nu(D_1 \cup D_2)$:

![Figure 2. Alexander unit ball and Thurston unit ball for $N_D$ — the dotted regions are qualitative.](image)

Figure 2 represents, in the space $H^1(N_D, \mathbb{R})$ with basis vectors the dual basis $\tau_i$, the unit ball of the Alexander norm, and a part of the unit ball of the Thurston norm.

Let now $K_1, K_2$ be a couple of fibered knots of genus $g(K_i) > 0$ and let $\Delta_{K_i}(t)$ be their Alexander polynomials. Next, define the closed manifold

$$N(K_1, K_2) = N_D \cup \left( \coprod_{i=1}^{2} S^3 \setminus \nu K_i \right),$$

where on the boundary tori the gluing map is defined to be the orientation reversing diffeomorphism which identifies the basis $(\mu(D_i), \lambda(D_i))$ with $(\mu(K_i), -\lambda(K_i))$. To interpret this, notice that each knot exterior is an homology solid torus, so that this operation appears as an homology Dehn filling for $N_D$, with surgery coefficient 0. The reason of the choice of this surgery curve appears evident from the fact (for a proof, see [EN], Section 3) that the minimal genus Seifert surface in $N_D$ representing an homology class Poincaré dual to a class $(m_1, m_2) \in H^1(N_D, \mathbb{Z})$ intersects the boundary
torus $T_i$ in $m_i$ copies of the longitude of each link component (note that $\text{lk}(D_1, D_2) = \Delta_D(1, 1) = 0$). Each minimal genus Seifert surface of $N_D$ has therefore a natural capping in $N(K_1, K_2)$, given by the union of $m_i$ copies of the fiber of $S^3 \setminus \nu K_i$. In particular, if $(m_1, m_2)$ is a fibered class, this fibration extends to a fibration of $N(K_1, K_2)$ through the fibrations

$$S^3 \setminus \nu K_i \longrightarrow S^1 \overset{(\cdot)^{m_i}}{\longrightarrow} S^1$$

of the knots’ exteriors. This proves, in particular, that $N(K_1, K_2)$ is irreducible. As the linking number of class $P$oincaré dual to $(m_1, m_2)$ constructed above is the minimal genus representative for the cohomology $S$ of the fiber of $T$ler norm $\parallel \cdot \parallel_{T}$ on the closed manifold, we have the following:

**Lemma 7.2** (Eisenbud-Neumann, Prop. 3.5). If $M$ is a compact irreducible $3$-manifold and $\mathbf{m} \in H^1(M)$, then the Thurston norm $\parallel \mathbf{m} \parallel_{T}$ is the sum of the norms of the restrictions of $\mathbf{m}$ to the Jaco-Shalen-Johannson components of $M$.

As a consequence of this Lemma, denoting with the symbol $\parallel \cdot \parallel_{\hat{\parallel \cdot \parallel}_{T}}$ the norm on the closed manifold, we have the following:

**Corollary 7.3.** The Thurston norm $\parallel(m_1, m_2)\parallel_{\hat{\parallel \cdot \parallel}_{T}}$ of an element $(m_1, m_2) \in H^1(N(K_1, K_2), \mathbb{Z})$ is given by

$$\parallel(m_1, m_2)\parallel_{\hat{\parallel \cdot \parallel}_{T}} = \parallel(m_1, m_2)\parallel_{T} + |m_1|(2g(K_1) - 1) + |m_2|(2g(K_2) - 1)$$

where $\parallel(m_1, m_2)\parallel_{T}$ is the Thurston norm of the corresponding element of $N_D$.

**Proof.** This follows from Lemma 7.2, together with the observation that the class $(m_1, m_2)$ on the closed manifold restricts to the element with same coordinates in $N_D$ and to the classes $m_i \in H^1(S^3 \setminus \nu K_i, \mathbb{Z})$, which have Thurston norm $\parallel m_i \parallel_{T} = |m_i|(2g(K_i) - 1)$, by definition of genus of a knot and linearity on rays.

We want to study now the Alexander norm of the manifold $N(K_1, K_2)$; in order to do this we need a gluing formula for the Alexander polynomial (or the SW invariant) along tori. We have the following:

**Lemma 7.4** (Gluing formula). Let $N(K_1, K_2) = N_D \cup (\bigsqcup_{i=1}^2 S^3 \setminus \nu K_i)$ be defined as above: Then the Alexander polynomials of the manifolds are related by the formula

$$\Delta_{N(K_1, K_2)}(t_1, t_2) = \Delta_{N_D}(t_1, t_2)\frac{\Delta_{K_1}(t_1)}{t_1 - 1} \frac{\Delta_{K_2}(t_2)}{t_2 - 1} = \Delta_{K_1}(t_1)\Delta_{K_2}(t_2).$$
Therefore, the Alexander norm \( \| \cdot \|_A \) on \( N(K_1, K_2) \) is given by
\[
\|(m_1, m_2)\|_A = |m_1|2g(K_1) + |m_2|2g(K_2)
= \|(m_1, m_2)\|_A + |m_1|(2g(K_1) - 1) + |m_2|(2g(K_2) - 1).
\]

Proof. It is known (see [MT], [Tu]) that the Milnor torsion is multiplicative by gluing along tori, with suitable identification of the variables; this torsion coincides with the Alexander polynomial for manifolds having \( b_1 > 1 \) and it is equal to the Alexander polynomial \( \Delta_K(t) \) divided by \( (t - 1) \) for the case of a knot. Remembering Proposition 7.1, Equation (32) above follows. The relation on the Alexander norm is then an easy corollary of this formula, as the degree of the Alexander polynomial of a fibered knot equals twice its genus.

This Lemma says, in particular, that the unit ball of the Alexander norm for \( N_D \) and \( N(K_1, K_2) \) are conformally equivalent (see Figure 3). We are ready to prove:

**Theorem 7.5.** There exist a family of fibered closed 3-manifolds \( N \) with \( H_1(N, \mathbb{Z}) = \mathbb{Z}^2 \) such that a fibered face of the Thurston unit ball is strictly contained in the corresponding face of the Alexander unit ball.

Proof. Our family is given by \( N(K_1, K_2) \) for any choice of the fibered knots \( K_i \). We observed before that each fibration of \( N_D \) in \( F_T \) extends to a fibration of \( N(K_1, K_2) \), defining a fibered face \( F_T^* \) of the Thurston unit ball for \( N(K_1, K_2) \). This face will be contained in \( F_A^* \), one of the four faces of the Alexander unit ball (having the same cone as \( F_A \)). This face is dual to the vertex \( t_1^{2g(K_1)}t_2^{2g(K_2)} \) (square of a vertex of the Newton polyhedron of the symmetrized Alexander polynomial). If a class \( (m_1, m_2) \in H^1(N_D, \mathbb{Z}) \) lies in the cone over \( F_A \setminus F_T^* \) (in particular \( \|(m_1, m_2)\|_A < \|(m_1, m_2)\|_T \)), then the corresponding class on the closed manifold has Alexander norm strictly smaller than the Thurston norm, from Equations (31) and (33), i.e., \( F_T^* \) is strictly contained in \( F_A^* \). From this the statement follows.

Figure 3 describes the Thurston and Alexander norm for a particular choice of \( g(K_i) \).

We want to outline a second proof of the same statement, based on the fact that a class on a closed three manifold is fibered if and only if all the restrictions to each JSJ component are fibered (see [EN], Theorem 4.2): If the Thurston norm on the closed manifold coincided with the Alexander norm on a larger cone than the one on \( N_D \), then there would be fibered classes on \( N(K_1, K_2) \) which restrict, on \( N_D \), to nonfibered ones (as mentioned above, all integral points laying on a face of the unit ball of the Thurston norm containing at least one fibration are fibered).

We finish this section pointing out that, although the link \( D \) above is the only example worked out in detail, fibered links with the properties of
Figure 3. Alexander unit ball and Thurston unit ball for $N(K_1, K_2)$ with $g(K_1) = 2, g(K_2) = 4$ — the dotted regions are qualitative.

Proposition 7.1 are likely to be “frequent” (compare the discussion in [Du]). From these examples, other closed 3-manifolds can be constructed.

References


[LL] T. Li and A. Liu, Uniqueness of symplectic canonical class, surface cone and symplectic cone of 4-manifolds with $b^+ = 1$, J. Differential Geom., 58(2) (2001), 331-370, CMP 1 913 946.


Received July 12, 2001 and revised January 26, 2002.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
IRVINE, CA 92697

DEPARTMENT OF MATHEMATICS
KANSAS STATE UNIVERSITY
MANHATTAN, KANSAS 66506
E-mail address: vidussi@math.ksu.edu
ZEROS OF EXTREMAL FUNCTIONS IN WEIGHTED BERGMAN SPACES

Rachel J. Weir

For $-1 < \alpha \leq 0$ and $0 < p < \infty$, the solutions of certain extremal problems are known to act as contractive zero-divisors in the weighted Bergman space $A^p_\alpha$. We show that for $0 < \alpha \leq 1$ and $0 < p < \infty$, the analogous extremal functions do not have any extra zeros in the unit disk and, hence, have the potential to act as zero-divisors. As a corollary, we find that certain families of hypergeometric functions either have no zeros in the unit disk or have no zeros in a half-plane.

1. Introduction.

For $-1 < \alpha < \infty$ and $0 < p < \infty$, the weighted Bergman space $A^p_\alpha$ is the space of functions $f$ analytic in the unit disk $D$ for which

$$||f||^p_{p,\alpha} = \int_D |f(z)|^p w_\alpha(z) \, d\sigma(z) < \infty,$$

where $w_\alpha(z) = (\alpha + 1)(1 - |z|^2)\alpha$ and $d\sigma$ is normalized area measure on the disk. When $\alpha = 0$, this space is precisely the unweighted Bergman space $A^p$. Hedenmalm [5] showed that contractive zero-divisors in $A^2$ can be constructed as solutions of extremal problems. This result was generalized to $A^p$ for $0 < p < \infty$ by Duren, Khavinson, Shapiro and Sundberg [2, 3]. More precisely, given an $A^p$ zero-set $\{\zeta_j\}$ with $\zeta_j \neq 0$, the solution $G$ of the extremal problem

$$\sup \{\text{Re} f(0) : f \in A^p, f(\{\zeta_j\}) = 0, ||f||_p = 1\}$$

is a contractive divisor. In other words, if $f \in A^p$ vanishes on $\{\zeta_j\}$, then $||f/G||_p \leq ||f||_p$.

Shimorin [9, 10, 11] showed that analogous results hold in $A^2_\alpha$ for $-1 < \alpha \leq 1$ and in $A^p_\alpha$ for $-1 < \alpha < 0$ and $0 < p < \infty$. On the other hand, it was shown by Hedenmalm and Zhu [7] that contractive divisors do not exist in $A^2_\alpha$ for $\alpha > 1$. For $p \neq 2$, it is an open problem to determine the precise values of $\alpha$ for which the extremal functions will act as contractive divisors.

To act as a divisor, an extremal function must vanish only at the points in the prescribed zero-set. To examine when this is true, we first consider the case when the zero-set consists of a single point. Indeed, showing that the
extremal function corresponding to a singleton zero-set has no extra zeros is a key step in proving the same for a general extremal function. For details see, for instance, [2].

In [13], we showed that for $-1 < \alpha < \infty$ and $1 \leq p < \infty$ this single-point extremal function can be expressed in terms of a hypergeometric function. This representation is valid whenever it is known that the hypergeometric function is nonvanishing in the unit disk, in which case the extremal function will have no extra zeros in $\mathbb{D}$. Equivalently, if the extremal function is known a priori to have no extra zeros, the same technique can be used to establish the formula and, as a consequence, we may conclude that the corresponding hypergeometric function has no zeros in $\mathbb{D}$.

When $\alpha$ is a nonnegative integer, the hypergeometric functions in question are in fact polynomials. Using this fact and a result of Osipenko and Stessin [8], we were able to establish that the single-point extremal function in $A_{p}^{\alpha}$ has no extra zeros when $\alpha = 1, 2, 3$ [13]. For other values of $\alpha$, describing the zeros of these hypergeometric functions directly is difficult.

In this paper, using techniques of Bergman space theory we establish a result about the zeros of hypergeometric functions. Indeed, we show that for $0 < \alpha \leq 1$ and $0 < p < \infty$ extremal functions in $A_{p}^{\alpha}$ have no extra zeros in the unit disk. In particular, the single-point extremal functions can not have any extra zeros in $\mathbb{D}$. From this we may conclude that the corresponding hypergeometric functions can not vanish in the unit disk.

The key step is to show that the single-point extremal functions in $A_{1}^{p}$ act as expansive multipliers on $A_{1}^{p}$ for $0 < \alpha \leq 1$, although they are constructed as functions in $A_{1}^{p}$. Since single-point extremal functions have no extra zeros when $\alpha = 1$, the result follows. This method is a modification of the technique used by Shimorin [11] to prove that extremal functions in $A_{p}^{\alpha}$ have no extra zeros when $-1 < \alpha < 0$ and $0 < p < \infty$. We also use properties of the weighted biharmonic Green function $\Gamma_{1}$ corresponding to the weight $w_{1}$. These properties can be established using the explicit formula for $\Gamma_{1}$ stated by Hedenmalm [6].

2. Background.

A sequence of points $\{\zeta_{j}\}$ is called an $A_{p}^{\alpha}$ zero-set if there is a function in $A_{p}^{\alpha}$ which vanishes precisely on this sequence. In particular, if a point $\alpha$ appears in the sequence $m$ times, then the function must have a zero at $\alpha$ of order exactly $m$. Given such a sequence $\{\zeta_{j}\}$ with $\zeta_{j} \neq 0$, we can pose the extremal problem

$$\sup \{ \text{Re} f(0) : f \in A_{p}^{\alpha}, f(\{\zeta_{j}\}) = 0, \|f\|_{p,\alpha} = 1 \}. $$

A normal families argument shows that for $\alpha > -1$ and $0 < p < \infty$ an extremal function $G$ will exist.
The Gauss hypergeometric function $F(a, b; c; z)$ is defined as

\begin{equation}
F(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!},
\end{equation}

where

$$(x)_n = \frac{\Gamma(x+n)}{\Gamma(x)} = x(x+1)\ldots(x+n-1)$$

is Pochhammer’s symbol [1]. Here, $\Gamma$ denotes the usual Gamma function. In [13], we showed that for each $\beta \in \mathbb{D}$ the extremal function $G_\beta$ corresponding to the zero-set $\{\beta\}$ can be expressed in terms of hypergeometric functions. Indeed, for $-1 < \alpha < \infty$ and $1 \leq p < \infty$,

\begin{equation}
G_\beta(z) = \frac{\beta-z}{|\beta|} \frac{\beta-z}{1-\beta^2} \left[ \frac{F(-\alpha-1, \frac{p}{2}; \frac{p}{2}+1; \beta; \beta-z)}{F(-\alpha-1, \frac{p}{2}; \frac{p}{2}+1; 1; \beta)} \right]^{\frac{1}{p}},
\end{equation}

provided that $F(-\alpha-1, \frac{p}{2}; \frac{p}{2}+1; \zeta) \neq 0$ in the disk $\{|\zeta| < |\beta|\}$. In this case, $G_\beta$ will not vanish in $\mathbb{D}$ except for a simple zero at $\beta$. Conversely, if we assume a priori that $G_\beta$ vanishes only at $\beta$, we can derive the same formula and conclude that the function $F(-\alpha-1, \frac{p}{2}; \frac{p}{2}+1; \zeta)$ is nonvanishing in $\{|\zeta| < |\beta|\}$.

For each fixed $\zeta$ in $\mathbb{D}$ the weighted biharmonic Green function $\Gamma_\alpha(z, \zeta)$ is the solution of the boundary-value problem

\begin{equation}
\begin{cases}
\Delta z \left( \frac{1}{w_\alpha(z)} \Delta z \Gamma_\alpha(z, \zeta) \right) = \delta_\zeta(z) & \text{for } z \in \mathbb{D} \\
\Gamma(z, \zeta) = \frac{\partial \Gamma_\alpha}{\partial n}(z, \zeta) = 0 & \text{for } z \in \partial \mathbb{D}.
\end{cases}
\end{equation}

Here, $\frac{\partial}{\partial n}$ is the outward normal derivative and $\Delta$ denotes one-fourth of the usual Laplacian. When $\alpha = 0$, the function $\Gamma_\alpha$ is precisely the well-known biharmonic Green function $\Gamma$ which is given by the formula

$$\Gamma(z, \zeta) = |z-\zeta|^2 \log \left| \frac{z-\zeta}{1-\overline{\zeta}z} \right|^2 + (1-|\zeta|^2)(1-|z|^2).$$

For details of the derivation of this formula, see [4].

Essential to the proof of our main result is the fact that an explicit formula for the weighted biharmonic Green function is also known when $\alpha = 1$. 
Indeed, Hedenmalm [6] presented the formula

\begin{align}
\Gamma_1(z, \zeta) &= 2 \left( |z - \zeta|^2 - \frac{1}{4} |z^2 - \zeta^2|^2 \right) \log \left| \frac{z - \zeta}{1 - \bar{\zeta}z} \right|^2 \\
&\quad + \frac{1}{4} (1 - |z|^2)(1 - |\zeta|^2) \left( 7 - |z|^2 - |\zeta|^2 - |z\zeta|^2 - 4 \Re(\bar{\zeta}z) \right) \\
&\quad - \frac{1}{2} (1 - |z|^2)^2 (1 - |\zeta|^2)^2 \Re \left\{ \frac{1 + \bar{\zeta}z}{1 - \bar{\zeta}z} \right\}.
\end{align}

The Green function \( \Gamma_1 \) is positive for \( z, \zeta \) in \( \mathbb{D} \) [6]. Moreover, the function \( \Delta_z \Gamma_1(z, \zeta) \) has the form

\begin{align}
\Delta_z \Gamma_1(z, \zeta) &= w_1(z) \left( \log \left| \frac{z - \zeta}{1 - \bar{\zeta}z} \right|^2 + H_1(z, \zeta) \right),
\end{align}

where \( H_1(\cdot, \zeta) \) is harmonic in \( \mathbb{D} \) for fixed \( \zeta \in \mathbb{D} \). Indeed, since

\[ \Delta \left( \log \left| \frac{z - \zeta}{1 - \bar{\zeta}z} \right|^2 \right) = \delta_\zeta(z), \]

the function

\[ \frac{1}{w_1(z)} \Delta_z \Gamma_1(z, \zeta) - \log \left| \frac{z - \zeta}{1 - \bar{\zeta}z} \right|^2 \]

must be harmonic in \( \mathbb{D} \).

When \( \alpha = 1 \) we are also able to establish some important properties of the single-point extremal functions. For the remainder of this paper, we will use the notation \( G_\beta \) to denote the single-point extremal function in the space \( A^p_1 \). In other words, \( G_\beta \) will be the solution of the extremal problem

\[ \sup \{ \Re f(0) : f \in A^p_1, f(\beta) = 0, \|f\|_{p,1} = 1 \}. \]

By Theorem 4.1 in [13], the hypergeometric function

\[ F\left(-2, \frac{p}{2}; \frac{p}{2} + 1; z \right) = \frac{(p + 2)(p + 4) - 2p(p + 4)z + p(p + 2)z^2}{(p + 2)(p + 4)} \]

does not vanish in \( \mathbb{D} \) when \( 1 \leq p < \infty \). In fact, it is easy to see that it does not vanish in \( \overline{\mathbb{D}} \). Hence, using Formula (2), we see that

\begin{align}
G_\beta(z) &= \frac{\beta - z}{|\beta|} \left[ \frac{P \left( \frac{\beta - z}{1 - \beta\bar{z}} \right)^2}{8P(|\beta|^2)} \right]^{1/p},
\end{align}

where \( P(z) = (p + 2)(p + 4) - 2p(p + 4)z + p(p + 2)z^2 \). Consequently, \( G_\beta(z) \neq 0 \) in \( \overline{\mathbb{D}} \) except for a simple zero at \( \beta \).
It is clear that $G_\beta(0)$ must be real and positive. Indeed, if not, multiplying $G_\beta$ by a suitable unimodular constant would produce a competitor function with larger real part. We can say more about $G_\beta(0)$.

**Lemma 2.1.** For $1 \leq p < \infty$, it is true that $0 < G_\beta(0) < 1$.

This result follows from the subharmonicity of $|G_\beta|^p$. The same arguments can be used to conclude that $0 < G(0) < 1$, where $G$ is an extremal function in $A^p_\alpha$ corresponding to an arbitrary $A^p_\alpha$ zero-set for $-1 < \alpha < \infty$ and $0 < p < \infty$.

### 3. The expansive multiplier property.

The first step in the proof of the main result is showing that $G_\beta$ acts as an expansive multiplier on $A^p_\alpha$ for $0 < \alpha \leq 1$, where $G_\beta$ is the single-point extremal function in the space $A^p_1$. The technique used to establish this is a modification of a method of Shimorin [11]. From (6), it is clear that the function $G_\beta$ is analytic in $D$. Hence, the function $(|G_\beta(z)|^p - 1)w_1(z)$ is continuous in $D$. Thus, the boundary value problem

$$
\begin{align*}
\Delta \phi &= (|G_\beta|^p - 1)w_1 \quad \text{in } D \\
\phi &= 0 \quad \text{on } \partial D
\end{align*}
$$

has a unique solution $\phi$ in $C^2(D)$.

**Lemma 3.1.** The solution $\phi$ has the following properties:

(i) $\frac{\partial \phi}{\partial n} = 0$ on $\partial D$

(ii) $\phi(\zeta) = \int_D \Gamma_1(z, \zeta)\Delta \left(\frac{1}{w_1(z)}\Delta \phi(z)\right) d\sigma(z)$ for each $\zeta$ in $D$.

**Proof.** (i) The proof is a modification of the proof of Lemma 3(a) in [2] and will be omitted. It uses the fact that

$$
\int_D (|G_\beta|^p - 1)fw_1 d\sigma = 0
$$

for all $f$ in $h^\infty$, which can be proved by adapting an argument given in [3].

(ii) We would like to show that, since $\phi = \frac{\partial \phi}{\partial n} = 0$ on $\partial D$, we can write

$$
\phi(\zeta) = \int_D \Gamma_1(z, \zeta)\Delta \left(\frac{1}{w_1(z)}\Delta \phi(z)\right) d\sigma(z).
$$

The result would then follow since $\Delta \left(\frac{1}{w_1}\Delta \phi\right) = \Delta(|G_\beta|^p - 1) = \Delta(|G_\beta|^p)$. Identity (8) can be obtained as a consequence of Green’s formula. However, since the integrand is smooth only in $D \setminus \{\beta, \zeta\}$, we must apply Green’s formula to the integral

$$
\int_{D_\epsilon} \Gamma_1(z, \zeta)\Delta \left(\frac{1}{w_1(z)}\Delta \phi(z)\right) d\sigma(z),
$$
where \( \mathbb{D}_\epsilon \) is the domain formed by removing from \( \mathbb{D} \) the balls \( B_\zeta = \{ z : |z - \zeta| < \epsilon \} \) and \( B_\beta = \{ z : |z - \beta| < \epsilon \} \). By letting \( \epsilon \to 0 \), the result follows from several estimates on the growth of \( \phi \) and \( \Gamma_1 \). First, on \( \partial B_\beta \),

\[
\epsilon \frac{\partial}{\partial n} \left( \frac{1}{w_1} \Delta \phi \right) \leq \epsilon p |G_\beta| \left| \frac{G'_\beta}{G_\beta} \right| = O(\epsilon^p),
\]
as \( \epsilon \to 0 \).

Second, direct calculation using Formula \((4)\) shows that

\[
\epsilon \frac{1}{w_1(\zeta + e^{i\theta})} \Delta \Gamma_1(\zeta + e^{i\theta}, \zeta) = O(\epsilon \log \epsilon),
\]
and

\[
\epsilon \frac{\partial}{\partial r} \left( \frac{1}{w_1(\zeta + re^{i\theta})} \Delta \Gamma_1(\zeta + re^{i\theta}, \zeta) \right) \bigg|_{r=\epsilon} = 2 + O(\epsilon)
\]
as \( \epsilon \to 0 \). \( \square \)

Using the properties of \( \phi \) established in Lemma 3.1, we can now prove that \( G_\beta \) is an expansive multiplier.

**Theorem 3.2.** For \( 0 < \alpha \leq 1 \) and \( 1 \leq p < \infty \), if \( f \in A_p^\alpha \), then

\[
\|G_\beta f\|_{p,\alpha} \geq \|f\|_{p,\alpha}.
\]

**Proof.** Let \( g = |f|^p h \), where \( f \) is a polynomial and \( h \in C^2(\mathbb{D}) \). Then

\[
\int_{\mathbb{D}} \left( |G_\beta(z)|^p - 1 \right) g(z) w_1(z) \, d\sigma(z)
\]

\[
= \int_{\mathbb{D}} g(z) \Delta \phi(z) \, d\sigma(z)
\]

\[
= \int_{\mathbb{D}} \phi(z) \Delta g(z) \, d\sigma(z)
\]

\[
= \int_{\mathbb{D}} \int_{\mathbb{D}} \Gamma_1(z, \zeta) \Delta \left( |G_\beta(\zeta)|^p \right) \Delta g(z) \, d\sigma(\zeta) \, d\sigma(z),
\]

where the last step follows from Lemma 3.1 (ii). Although the derivatives of \( g \) are not smooth near the zeros of the polynomial \( f \), we can use arguments similar to those used to show \((8)\) to justify our use of Green’s formula.

For \( n = 0, 1, 2, \ldots \), we now let \( g_n(z) = |f(z)|^p \sum_{k=0}^{n} (-1)^k \binom{\alpha - 1}{k} |z|^{2k} \), where \( f \) is a polynomial and \( 0 < \alpha < 1 \). When \( \alpha = 1 \), we let \( g_n(z) = |f(z)|^p \). Each function \( g_n \) is subharmonic. To see this, we need only observe that for each \( k \in \mathbb{N} \) and for \( p > 0 \), the function \( \log \left( |f|^p |z|^{2k} \right) \) is subharmonic and
hence, by Jensen’s inequality, the function $|f|^p |z|^{2k}$ is subharmonic, since the exponential function is convex. As 

$$
(-1)^k \binom{\alpha - 1}{k} = \frac{1}{k!} (1 - \alpha)(2 - \alpha) \ldots (k - \alpha) > 0
$$

for $0 < \alpha < 1$, each function $g_n$ is a positive linear combination of subharmonic functions and must also be subharmonic. Setting $g = g_n$ in (9), we conclude that 

$$
\int_{D} (|G_\beta(z)|^p - 1) |f(z)|^p \sum_{k=0}^{n} (-1)^k \binom{\alpha - 1}{k} |z|^{2k} w_1(z) d\sigma(z) \geq 0,
$$

since $g_n$ is subharmonic and $\Gamma_1$ is positive. An application of the Dominated Convergence Theorem then shows that 

$$
\int_{D} (|G_\beta(z)|^p - 1) |f(z)|^p (1 - |z|^2)^{\alpha - 1} w_1(z) d\sigma(z) \geq 0,
$$

from which we can conclude that 

$$
\|G_\beta f\|_{p,\alpha} \geq \|f\|_{p,\alpha}
$$

for all polynomials $f$. Since $G_\beta$ is bounded and polynomials are dense in $A^p_\alpha$, the result follows. \[\square\]

4. Boundary values of extremal functions.

Before proving the main theorem, we first show that since $G_\beta$ is an expansive multiplier, its boundary values can not be too small. The ideas in this section generalize an approach of Dragan Vukoti\’c [12].

The weighted Berezin transform $B_\alpha u$ of a function $u \in L^1(D, w_\alpha d\sigma)$ is defined as 

$$
B_\alpha u(a) = \int_{\mathbb{D}} u(\varphi_\alpha(\zeta)) w_\alpha(\zeta) d\sigma(\zeta) = \int_{\mathbb{D}} (1 - |a|^2)^{\alpha + 2}(1 - |z|^2)^{\alpha} \frac{u(z)}{|1 - \overline{a}z|^{2\alpha + 4}} d\sigma(z),
$$

where $\varphi_\alpha(\zeta) = \frac{a - \overline{\zeta}}{1 - \overline{a}\zeta}$.

**Proposition 4.1.** Let $a_0 \in \partial\mathbb{D}$ and $u \in L^1(\mathbb{D}, w_\alpha d\sigma)$. If $\lim_{a \to a_0} u(a) = L$ exists and is finite, then $\lim_{a \to a_0} B_\alpha u(a) = L$.

**Proof.** Fix $\epsilon > 0$. We can find $r > 0$ for which $|u(z) - L| < \epsilon/2$ whenever $z \in D_0 = \mathbb{D} \cap D(a_0, r)$. It is not hard to show that $|\varphi_\alpha'(z)|^{\alpha + 2} \to 0$ uniformly on compact subsets of $\overline{\mathbb{D}} \setminus \{a_0\}$ as $a \to a_0$. Hence, we can choose $a$ close enough to $a_0$ to ensure that 

$$
|\varphi_\alpha'(z)|^{\alpha + 2} < \frac{\epsilon}{2(|u|_{1,\alpha} + |L|)}
$$
for all $z \in \mathbb{D}\setminus D_0$. Noting that
\[
\int_{D_0} |\varphi'_a(z)|^2 w_\alpha(\varphi_a(z)) d\sigma(z) = 1,
\]
we see that
\[
|B_\alpha u(a) - L| \leq \frac{\epsilon}{2} \int_{D_0} |\varphi'_a|^2 w_\alpha(\varphi_a) d\sigma + \int_{\mathbb{D}\setminus D_0} |u - L| |\varphi'_a|^\alpha w_\alpha d\sigma \leq \epsilon.
\]

\[
\text{□}
\]

**Theorem 4.2.** Suppose $-1 < \alpha < \infty$ and $0 < p < \infty$ and let $G$ be an expansive multiplier in $A^p_\alpha$. If $G$ has a continuous extension across the point $a_0 \in \partial \mathbb{D}$, then $|G(a_0)| \geq 1$.

**Proof.** For $a \in \mathbb{D}$ let
\[
f_a(z) = \left(\frac{1 - |a|^2}{(1 - \overline{a}z)^2}\right)^{\alpha+2/p}.
\]
Then $f$ is in $A^p_\alpha$ since $\|f_a\|_{p,\alpha} = 1$. Now by Proposition 4.1, since $|G|^p \in L^1(\mathbb{D}, w_\alpha d\sigma)$,
\[
\int_{\mathbb{D}} |f_a G|^p w_\alpha d\sigma = B_\alpha(|G|^p)(a) \to |G(a_0)|^p
\]
as $a \to a_0$. But the expansive multiplier property implies that for each $a \in \mathbb{D}$,
\[
\int_{\mathbb{D}} |f_a G|^p w_\alpha d\sigma \geq \int_{\mathbb{D}} |f_a|^p w_\alpha d\sigma = 1.
\]
Hence, $|G(a_0)|^p \geq 1$. \[\text{□}\]

Since $G_\beta$ is analytic in $\overline{\mathbb{D}}$ and, by Theorem 3.2, acts as an expansive multiplier on $A^p_\alpha$ for $0 < \alpha \leq 1$, we have the following consequence of Theorem 4.2:

**Corollary 4.3.** $|G_\beta(z)| \geq 1$ for all $z$ in $\partial \mathbb{D}$.

## 5. The main result.

We now show that extremal functions in $A^p_\alpha$ can not have extra zeros in $\mathbb{D}$ when $0 < \alpha \leq 1$ and $1 \leq p < \infty$.

**Theorem 5.1.** For $0 < \alpha \leq 1$ and $1 \leq p < \infty$, let $G$ be an extremal function in $A^p_\alpha$ corresponding to an $A^p_\alpha$ zero-set. Then $G$ has no extraneous zeros in $\mathbb{D}$.

**Proof.** Suppose $G$ has an extra zero, $\beta$, for some $p$ and $\alpha$. Let $G_\beta$ be the corresponding single-point extremal function in $A^p_\alpha$ given by Formula (2) and consider the function $G/G_\beta$. Since $|G_\beta(z)| \geq 1$ for all $z \in \partial \mathbb{D}$ by Corollary 4.3, it is not hard to see that $G/G_\beta \in A^p_\alpha$. Moreover,
\[
1 = \|G\|_{p,\alpha} = \|G_\beta(G/G_\beta)\|_{p,\alpha} \geq \|G/G_\beta\|_{p,\alpha},
\]
by Theorem 3.2. Since the function
\[
\frac{G/G_\beta}{\|G/G_\beta\|_{p,\alpha}}
\]
is a competitor for the extremal property, we see that
\[
\text{Re}\{G(0)\} \geq \text{Re}\left\{\frac{G(0)/G_\beta(0)}{\|G/G_\beta\|_{p,\alpha}}\right\} > \text{Re}\{G(0)\},
\]
where we have also used Lemma 2.1. From this contradiction, we see that
\(G\) can not have any extraneous zeros. \(\square\)

In particular, for \(0 < \alpha \leq 1\) and \(1 \leq p < \infty\) the single-point extremal functions in \(A^p_\alpha\) can not have any extra zeros in \(\mathbb{D}\). Hence, these extremal functions must be given by Formula (2). Shimorin's results [11] allow us to conclude the same for \(-1 < \alpha < 0\). We then have the following result about the corresponding hypergeometric functions:

**Corollary 5.2.** When \(1 \leq p < \infty\) and \(-1 < \alpha \leq 1\), the hypergeometric function \(F(-\alpha - 1, \frac{p}{2}; \frac{p}{2} + 1; z)\) has no zeros in \(\mathbb{D}\).

Using two well-known transformation formulas, we can prove additional corollaries involving hypergeometric functions. The first corollary is a consequence of Pfaff’s transformation

\[
F(a, b; c; z) = (1 - z)^{-a}F\left(a, c - b; c; \frac{z}{z - 1}\right).
\]

A derivation of this transformation can be found in [1].

**Corollary 5.3.** When \(1 \leq p < \infty\) and \(-1 < \alpha \leq 1\), the hypergeometric functions \(F(-\alpha - 1, 1; \frac{p}{2}; \frac{p}{2} + 1; z)\) and \(F(\frac{p}{2} + \alpha + 2, \frac{p}{2}; \frac{p}{2} + 1; z)\) have no zeros in the half-plane \(\text{Re}\{z\} < \frac{1}{2}\).

By exploiting the symmetry of the hypergeometric function in the parameters \(a\) and \(b\), Pfaff’s transformation can be used to establish Euler’s formula

\[
(11) \quad F(a, b; c; z) = (1 - z)^{c-a-b}F(c - a, c - b; c; z).
\]

We can then conclude the following:

**Corollary 5.4.** When \(1 \leq p < \infty\) and \(-1 < \alpha \leq 1\), the hypergeometric function \(F(\frac{p}{2} + \alpha + 2, 1; \frac{p}{2} + 1; z)\) has no zeros in \(\mathbb{D}\).
6. Extremal functions for $0 < p < 1$.

In [13], Formula (2) was established only when $1 \leq p < \infty$, since the proof required the use of certain Banach space results. We now show that when $\alpha = 1$ the formula holds for $0 < p < 1$ also. This allows us to prove that, even for $0 < p < 1$, extremal functions in $A^p_\alpha$ can not have extra zeros when $0 < \alpha \leq 1$.

For $0 < p < 1$, define the function $G_\beta$ as

$$G_\beta(z) = \frac{\beta - z}{|\beta|} \left[ \frac{F \left( -2, \frac{p}{2}; \frac{p}{2} + 1; \frac{\beta z - 1}{1 - \beta z} \right)^2}{F \left( -2, \frac{p}{2}; \frac{p}{2} + 1; 1 \right) F \left( -2, \frac{p}{2}; \frac{p}{2} + 1; |\beta|^2 \right)} \right]^{\frac{1}{p}}. \quad (12)$$

In other words, let Formula (2) be the definition of $G_\beta$ when $0 < p < 1$ and $\alpha = 1$. We first establish an orthogonality property of this function.

**Lemma 6.1.** For $0 < p < 1$, the function $G_\beta$ has the property

$$\int_D |G_\beta|^p f w_1 d\sigma = f(0) \quad (13)$$

for all $f \in h^\infty$.

**Proof.** First, using Equation (11), we can write $G_\beta$ in the form

$$\bar{\beta} \varphi_\beta(z) \left[ \frac{(1 - |\beta|^2)^3 F \left( \frac{p}{2} + 3, 1; \frac{p}{2} + 1; \bar{\beta} \varphi_\beta(z) \right)^2}{2 B \left( \frac{p}{2} + 1, 2 \right) (1 - \bar{\beta} z)^6 F \left( \frac{p}{2} + 3, 1; \frac{p}{2} + 1; |\beta|^2 \right)} \right]^{\frac{1}{p}},$$

where $\varphi_\beta(z) = (\beta - z)/(1 - \bar{\beta} z)$ and $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x + y)$ is the Beta function. If $f$ is a polynomial, setting $\zeta = \varphi_\beta(z)$ we see that,

$$\int_D |G_\beta(z)|^p f(z) w_1(z) d\sigma(z)$$

$$= \left[ B \left( 2, \frac{p}{2} + 1 \right) F \left( \frac{p}{2} + 3, 1; \frac{p}{2} + 1; |\beta|^2 \right) \right]^{-1}$$

$$\cdot \int_D |\zeta|^p F \left( \frac{p}{2} + 3, 1; \frac{p}{2} + 1; |\beta|^2 \right) f(\varphi_\beta(\zeta))(1 - |\zeta|^2) d\sigma(\zeta).$$

Letting $F \left( \frac{p}{2} + 3, 1; \frac{p}{2} + 1; \bar{\beta} \zeta \right) = \sum_{n=0}^\infty b_n \zeta^n$ and

$$F \left( \frac{p}{2} + 3, 1; \frac{p}{2} + 1; |\beta|^2 \right) f(\varphi_\beta(\zeta)) = \sum_{n=0}^\infty c_n \zeta^n,$$
we find that
\[
\int_{D} |G_{\beta}(z)|^{p} f(z) w_{1}(z) \, d\sigma(z)
= \left[ B \left( 2, \frac{p}{2} + 1 \right) F \left( \frac{p}{2} + 3, 1; \frac{p}{2} + 1; |\beta|^{2} \right) \right]^{-1}
\cdot \sum_{n=0}^{\infty} B \left( 2, \frac{p}{2} + n + 1 \right) b_{n} c_{n}.
\]

Now, rewriting the definition of \( F \left( \frac{p}{2} + 3, 1; \frac{p}{2} + 1; z \right) \), we see that
\[
F \left( \frac{p}{2} + 3, 1; \frac{p}{2} + 1; z \right) = \sum_{n=0}^{\infty} B \left( 2, \frac{p}{2} + 1 \right) B \left( 2, \frac{p}{2} + n + 1 \right) z^{n},
\]
and so, \( B(2, \frac{p}{2} + n + 1)b_{n} = B(2, \frac{p}{2} + 1)\beta^{n} \). Thus,
\[
\int_{D} |G_{\beta}(z)|^{p} f(z) w_{1}(z) \, d\sigma(z)
= \left[ B \left( 2, \frac{p}{2} + 1 \right) F \left( \frac{p}{2} + 3, 1; \frac{p}{2} + 1; |\beta|^{2} \right) \right]^{-1}
\cdot \sum_{n=0}^{\infty} B \left( 2, \frac{p}{2} + 1 \right) \beta^{n} c_{n}
\]
\[
= \left[ F \left( \frac{p}{2} + 3, 1; \frac{p}{2} + 1; |\beta|^{2} \right) \right]^{-1} \sum_{n=0}^{\infty} c_{n} \beta^{n} = f(0).
\]

This shows that (13) holds for all polynomials \( f \). It is clear that the result holds for harmonic polynomials also. Since \( |G_{\beta}|^{p} \in L^{1}(D, w_{1} d\sigma) \), we can conclude that the relation holds for all \( f \in h^{\infty} \). □

We note that (13) is equivalent to the condition
\[
\int_{D} (|G_{\beta}|^{p} - 1) f w_{1} d\sigma = 0
\]
for all \( f \) in \( h^{\infty} \), using the mean-value property for harmonic functions.

**Proposition 6.2.** Suppose \( 0 < p < 1 \), \( \beta \in D \setminus \{0\} \) and \( \alpha = 1 \). Then the function \( G_{\beta} \) defined by Formula (12) is the unique extremal function in \( A_{p}^{1} \) corresponding to the zero-set \( \{\beta\} \).

**Proof.** For \( 0 < p < 1 \), the function defined by Formula (12) is analytic in \( \overline{D} \). Thus, the boundary-value problem (7) has a solution \( \phi \in C^{2}(\overline{D}) \) in this case also. In view of Lemma 6.1, we can then establish an analogue of Lemma 3.1 and use the representation for \( \phi \) given in the second part of this lemma to show that \( G_{\beta} \) acts as an expansive multiplier on \( A_{p}^{1} \) for \( 0 < \alpha \leq 1 \). Thus, by Theorem 4.2, we see that \( |G_{\beta}(z)| \geq 1 \) for all \( z \in \partial D \). Hence, if \( f \in A_{p}^{1} \),
and \( f(\beta) = 0 \), then \( f/G_\beta \in A_1^p \) and \( \|f/G_\beta\|_{p,\alpha} \leq \|f\|_{p,\alpha} \). We use this last property to show that \( G_\beta \) is extremal.

Let \( f \in A_1^p \) satisfy \( f(\beta) = 0 \) and \( \|f\|_{p,\alpha} = 1 \), and assume that \( f \) is not a constant multiple of \( G_\beta \). For every \( g \in A_\alpha^1 \), since \( |g|^p \) is subharmonic, \( |g(0)|^p \leq \|g\|_{p,\alpha} \), unless \( g \) is constant. Thus since \( G_\beta(0) > 0 \),

\[
\frac{\text{Re}\{f(0)\}}{\text{Re}\{G_\beta(0)\}} \leq \left| \frac{f(0)}{G_\beta(0)} \right| < \|f/G_\beta\|_{p,\alpha} \leq \|f\|_{p,\alpha} = 1.
\]

In other words, \( \text{Re}\{f(0)\} < \text{Re}\{G_\beta(0)\} \).

Thus, the only remaining competitors for the extremal property are constant multiples of \( G_\beta \). But since we know that the extremal function has norm 1 and is real and positive at the origin (see the remark after Lemma 2.1), \( G_\beta \) is the only possible candidate. \( \square \)

We can now conclude that our main result holds for \( 0 < p < 1 \) also.

**Theorem 6.3.** For \( 0 < \alpha \leq 1 \) and \( 0 < p < 1 \), let \( G \) be an extremal function in \( A_\alpha^p \) corresponding to an \( A_\alpha^p \) zero-set. Then \( G \) has no extraneous zeros in \( \mathbb{D} \).

---

**References**


Received October 23, 2001.

Department of Mathematics
University of Virginia
Charlottesville, VA 22904
E-mail address: rw8t@virginia.edu
**Guidelines for Authors**

Authors may submit manuscripts at [pjm.math.berkeley.edu/about/journal/submissions.html](http://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](mailto: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\LaTeX 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](mailto: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 will 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.
Construction de la tour des 2-corps de classes de Hilbert de certains corps biquadratiques
Abdelmalek Azizi

Connections on principal bundles over curves in positive characteristic
Indranil Biswas and S. Subramanian

Cubic singular moduli, Ramanujan’s class invariants $\lambda_n$ and the explicit Shimura Reciprocity Law
Heng Huat Chan, Victor Tan and Alice Goo

Sums of products of generalized Bernoulli polynomials
Kwang-We Chen

The moduli space of real algebraic curves of genus 2
Francisco-Javier Cirre

Equivariant framings, lens spaces and contact structures
Siddhartha Gadgil

Carlson’s convergence theorem for Dirichlet series
Håkan Hedenmalm and Eero Saksman

An Inverse Problem from sub-Riemannian geometry
Thomas A. Ivey

Small contractions of smooth varieties
Gianluca Occhetta

Orbifold Hodge numbers of Calabi–Yau hypersurfaces
Mainak Pagad

Norms on the cohomology of a 3-manifold and SW theory
Stefano Vidussi

Zeros of extremal functions in weighted Bergman spaces
Rachel J. Weir