MULTIPLIERS OF OPERATOR SPACES, AND THE INJECTIVE ENVELOPE

DAVID P. BLECHER AND VERN I. PAULSEN

We study the injective envelope $I(X)$ of an operator space $X$, showing amongst other things that it is a self-dual $C^*$-module. We describe the diagonal corners of the injective envelope of the canonical operator system associated with $X$. We prove that if $X$ is an operator $A$-$B$-bimodule, then $A$ and $B$ can be represented completely contractively as subalgebras of these corners. Thus, the operator algebras that can act on $X$ are determined by these corners of $I(X)$ and consequently bimodule actions on $X$ extend naturally to actions on $I(X)$. These results give another characterization of the multiplier algebra of an operator space, which was introduced by the first author, and a short proof of a recent characterization of operator modules, and a related result. As another application, we extend Wittstock’s module map extension theorem, by showing that an operator $A$-$B$-bimodule is injective as an operator $A$-$B$-bimodule if and only if it is injective as an operator space.

1. Introduction.

In this paper we investigate some connections between the following topics: Injectivity of operator spaces, self-dual Hilbert $C^*$-modules (in the sense of Paschke); completely contractive actions of one operator space on another; and the notion of a multiplier operator algebra of an operator space which was recently introduced by the first author.

The results and definitions follow a natural logical sequence, so we begin without further delay. We refer to [7] for additional information, and complementary results, and to [29, 15, 32] for background information on operator spaces and completely bounded maps.

Recall that an operator space $X$ is injective, if for any operator spaces $W \subset Z$, and any completely bounded linear $T : W \to X$, there exists a linear $\tilde{T} : Z \to X$ extending $T$, with $\|T\|_{cb} = \|\tilde{T}\|_{cb}$. It has been known for a long time that $B(H)$ is an injective operator space (see [29, 39, 28, 2]) for any Hilbert space $H$. In 1983, Youngson showed [40] that an injective operator space $X$ is a ‘corner’ of a $C^*$-algebra $A$, by which we mean that there exist
projections \( p, q \) in the multiplier algebra \( \mathcal{M}(A) \) of \( A \), such that \( X = p A q \). It is well-known \cite{34} that this last condition is equivalent to saying that such an \( X \) is a Hilbert \( C^* \)-module. Then Hamana in 1985 (see notes in \cite{19, 20}), and Ruan independently \cite{35}, showed that any operator space \( X \) has an operator space injective envelope \( I(X) \). To prove the existence of this envelope, one may follow the classical construction for Banach space injective envelopes. We will sketch the main idea: One begins by choosing any injective object \( B \) containing \( X \). Then one considers the \( X \)-projections on \( B \), by which we mean completely contractive idempotent maps on \( B \) which fix \( X \). An idempotent map of course is one such that \( \Psi \circ \Psi = \Psi \).

There is a natural ordering on such maps, and, with a little work one can show, by a Zorn’s lemma argument, that there is a minimal \( X \)-projection \( \Phi \). The range of \( \Phi \) in \( B \) may be taken to be the injective envelope \( I(X) \) of \( X \), and one has \( X \subset I(X) \subset B \). Thus one sees that \( I(X) \) is the smallest injective space containing \( X \). As in the Banach space case (see \cite{21, Section 11} for details and references), one proves that \( I(X) \) is an ‘essential’ and ‘rigid’ extension of \( X \). The latter term, rigidity, means that the identity map is the only completely contractive map on \( I(X) \) extending the identity map on \( X \).

Note that if \( S \) is a linear subspace of \( B(H) \) containing \( I_H \) (for example, if \( S \) is a unital \( C^* \)-algebra), then one may choose \( B = B(H) \) in the above. Since \( \Phi(I) = I \), it follows that \( \Phi \) is completely positive \cite{29}. A well-known theorem of Choi and Effros \cite{12} states that the range of a completely positive unital idempotent map on a \( C^* \)-algebra, is a \( C^* \)-algebra with respect to a certain multiplication. Hence it follows that \( I(S) \) is a unital \( C^* \)-algebra.\(^1\)

Hamana also gives another construction of \( I(X) \) to the one outlined above, which allows one to prove something a little stronger. Since this construction will be important for us, we will outline some of the ideas. By the method popularized by the second author (see \cite{29} Lemma 7.1), we may embed the operator space \( X \) in a canonical unital operator system\(^2\)

\[
S(X) = \left( \begin{array}{cc} \mathbb{C} & X \\ X^* & \mathbb{C} \end{array} \right).
\]

If one forms the injective envelope \( I(S(X)) \), it will be a unital \( C^* \)-algebra, by the aforementioned argument using the Choi-Effros result. Indeed, since the minimal \( S(X) \)-projection fixes the \( C^* \)-algebra \( \mathbb{C} \oplus \mathbb{C} \) which is the diagonal of \( S(X) \), it follows immediately (for example by Lemma 1.6 below, although this is not necessary), that the following elements of \( S(X) \) are two selfadjoint

\(^1\)In fact, \( I(A) \) is a unital \( C^* \)-algebra even for a non-unital \( C^* \)-algebra \( A \). Since we have no good reference for this no doubt well known fact, we supply a proof later.

\(^2\)An operator system is a selfadjoint linear subspace of \( B(H) \) containing \( I_H \).
projections with sum 1 in the $C^*$-algebra $I(S(X))$:

$$p = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad q = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$ 

Therefore, with respect to $p$ and $q$, we may decompose $I(S(X))$ to write it as consisting of $2 \times 2$ matrices. Hamana shows that $pI(S(X))q$, the 1-2 corner of $I(S(X))$, is the injective envelope of $X$. This recovers and strengthens Youngson’s result. The four corners of $I(S(X))$ we will name:

$$I(S(X)) = \begin{pmatrix} I_{11} & I(X) \\ I(X)^* & I_{22} \end{pmatrix}.$$ 

It is clear that $I_{11}$ and $I_{22}$ are also injective $C^*$-algebras.

We will write $j$ for the canonical inclusion of $X$ inside $I(X)$.

Writing $I = I(X)$ for a moment, we define a subset of $I(S(X))$ by

$$\mathcal{L}(I(X)) = \begin{pmatrix} II^* & I \\ I^* & I^*I \end{pmatrix}$$

where $II^*$ for example, is the closed span in $I_{11}$ of terms $xy^*$ with $x, y \in I$.

Constructions related to this one have been studied by Hamana, Ruan, and C. Zhang (\cite{41}). Henceforth, we will reserve the letters $C(X)$ and $D(X)$ for $II^*$ and $I^*I$ respectively. When $X$ is understood we will simply write $C$ and $D$. Thus $C \subset I_{11}, D \subset I_{22}$ as $C^*$-subalgebras. Notice also that $\mathcal{L}(I(X))$ coincides with the smallest closed 2-sided ideal in $I(S(X))$ containing the copy of $I(X)$ in the 1-2-corner of $I(S(X))$, and this fact will be used below.

Clearly $\mathcal{L}(I(X))$ is a $C^*$-subalgebra of $I(S(X))$. From this, or otherwise, it is easy to see that $I(X)$ is a $C - D$-bimodule which, with respect to the natural $C$- and $D$-valued inner products given by $xy^*$ and $x^*y$, is a ‘strong Morita equivalence $C - D$-bimodule’ in the language of Rieffel. Sometimes this is also referred to as a ‘$C - D$-imprimitivity bimodule’. It follows by basic $C^*$-algebraic Morita theory (see \cite{34, 22} say), that $C \cong \mathbb{K}(I(X))$ as $C^*$-algebras, where $\mathbb{K}(I(X))$ is the so called ‘imprimitivity $C^*$-algebra’ of the right $D$-module $I(X)$. Also, $\mathcal{M}(C) \cong \mathbb{B}_D(I(X))$, the adjointable $D$-module maps on $I(X)$. In fact we shall see that the left multiplier algebra $LM(C)$ (which by a result of Lin \cite{24} may be identified with the space $B_D(I(X))$ of bounded right $D$-module maps on $I(X)$) coincides with $\mathcal{M}(C)$. Also $\mathcal{L}(I(X))$ may be identified with the ‘linking $C^*$-algebra’ \cite{34} of the bimodule $I(X)$. Although this will not be explicitly used below, it is a useful perspective.

We shall write $S_0(X)$ for the subspace of $S(X)$ consisting of those elements with 0’s on the main diagonal.

**Proposition 1.1.** For any operator space $X$, we have that

$$J = I(S(X))S_0(X)I(S(X))$$

is an essential ideal in $I(S(X))$. 
Proof. Suppose that $K$ is an ideal in $I(S(X))$ whose intersection with $J$ is zero. Let $\pi : I(S(X)) \to I(S(X))/K$. Thinking of $I(S(X))$ as $2 \times 2$ matrices, it is clear that $\pi$ maps each of the 4 corners into a matching ‘corner’ of $I(S(X))/K$. Since $\pi$ is 1-1 on $J$, it is completely isometric on $S_0(X)$. Let $\Phi$ be the restriction of $\pi$ to $S(X)$. By Lemma 7.1 in [29], $\Phi$ is a complete order injection.

Extend the map from $\pi(S(X)) \to S(X)$ which is the inverse of $\Phi$, to a map $\gamma : I(S(X))/K \to I(S(X))$. Since $\gamma \circ \Phi = Id_{S(X)}$, it follows by rigidity that $\gamma \circ \pi = Id_{I(S(X))}$. Thus $K = (0)$. □

Corollary 1.2. For any operator space $X$, we have that $\mathcal{L}(I(X))$ is an essential ideal in $I(S(X))$, and that $\mathcal{C} = C(X)$ is an essential ideal in $I_{11}$. Thus we have the following canonical inclusions of $C^*$-algebras

$$\mathcal{C} \subset I_{11} \subset M(\mathcal{C}).$$

If $\mathcal{C}$ is represented faithfully and nondegenerately on a Hilbert space $H$, then this string may be regarded as inclusions of subalgebras of $B(H)$. Similar assertions hold for $\mathcal{D}(X) \subset I_{22}$.

Proof. Clearly $J \subset \mathcal{L}(I(X)) \subset I(S(X))$, where $J$ is as in Proposition 1.1. Thus $\mathcal{L}(I(X))$ is an essential ideal in $I(S(X))$, by that Proposition. To see the second assertion, notice that if $t \in I_{11}$ and $t\mathcal{C} = 0$, then $t\mathcal{C}t^* = 0$, which implies that $tz^*t^* = tz = 0$ for all $z \in I(X)$. It follows immediately that $(t \oplus 0)\mathcal{L}(I(X)) = 0$. Hence by the first assertion, $t = 0$. Thus $\mathcal{C}$ is an essential ideal in $I_{11}$.

That $I_{11} \subset M(\mathcal{C})$ follows from the universal property of the multiplier algebra, namely that $M(\mathcal{C})$ contains a copy of any $C^*$-algebra containing $\mathcal{C}$ as an essential ideal ([22] Chapter 2, say). □

Since $\mathcal{C}$ is essential in $I_{11}$, it follows from [16] Theorem 4.5, that $M(\mathcal{C}) \subset I_{11}$. Thus in fact $I_{11} = M(\mathcal{C}) \cong \mathcal{B}_{\mathcal{D}}(I(X))$. However we will deduce all this in a self-contained way, from some machinery we develop next:

Corollary 1.3. If $t \in I_{11}$, and if $tx = 0$ for all $x \in X$, then $t = 0$.

Proof. Without loss of generality, we may assume that $\|t\| \leq 1$. By replacing $t$ by $t^*t$, we may also suppose that $0 \leq t \leq 1$, where this last ‘1’ is the identity of $I_{11}$. Let $p = 1 - t$. Define $\phi(z) = pz$, for $z \in I(X)$. Since $\phi(x) = x$ for $x \in X$, we obtain $\phi = Id$, by rigidity. Thus $tI(X) = 0$, which (we showed in the proof of Corollary 1.2) implies that $t = 0$. □

Definition 1.4. Let $X$ be an operator space. We define the left multiplier operator algebra of $X$ to be $IM_l(X) = \{T \in I_{11} : TX \subset X\}$. We define the left multiplier $C^*$-algebra of $X$ to be $IM^*_l(X) = \{T \in IM_l(X) : T^* \in IM_l(X)\}$. 
Note that $IM_l^*(X)$ is a C*-algebra, whereas $IM_l(X)$ is a unital nonselfadjoint operator algebra in general. There is a similar definition for right multipliers.

We shall soon see that these multiplier algebras coincide with the ones introduced in $[7]$ §4. These simultaneously generalize the common operator algebras associated with Hilbert C*-modules, the multiplier function algebras of a Banach space introduced by Alfsen and Effros $[1]$ (who only considered the real scalar case - see $[4]$ for the complex case), and the multiplier algebras of a nonselfadjoint operator algebra with c.a.i..

**Definition 1.5.** Let $H, K$ be Hilbert spaces. A multiplication situation on $H \oplus K$ consists of three concrete operator spaces $X, Y, Z$ such that $Y \subset B(H)$, $X \subset B(K, H)$, and $Z \subset B(K)$, and such that $YX \subset X$ and $XZ \subset X$.

We will need the following elementary lemma (c.f. $[29]$ Ex. 4.2-4.5):

**Lemma 1.6 (Choi $[11]$).** Suppose that $\phi : A \to B$ is a completely positive map between C*-algebras, with $\phi(1) = 1$. Suppose that there is a C*-subalgebra $N$ of $A$ with $1_A \in N$, such that $\pi = \phi|_N$ is a *-homomorphism. Then $\phi$ is an ‘$N$-bimodule map’. That is,

$$\phi(an) = \phi(a)\pi(n) \quad \text{and} \quad \phi(na) = \pi(n)\phi(a)$$

for all $a \in A, n \in N$.

We will refer to the following as the ‘multiplication theorem’:

**Theorem 1.7.**

(i) Suppose that $X$ is an operator space, and that $I_{11}, I_{22}$ are the diagonal corners of $I(S(X))$, as usual. If $Y, Z$ are two operator spaces such that $X, Y, Z$ form a multiplication situation on $H \oplus K$, as above, then there exist unique completely contractive linear maps $\theta : Y \to IM_l(X)$ and $\pi : Z \to IM_r(X)$ such that $\theta(y)j(x) = j(yx)$, and $j(x)\pi(z) = j(xz)$, for all $x \in X, y \in Y, z \in Z$.

(ii) If in addition, $Y$ is a subalgebra (resp. *-subalgebra) of $B(H)$, then $\theta$ is also a homomorphism (resp. *-homomorphism into $IM_l^*(X)$). Similarly for $Z$.

**Proof.** By $[29]$ Lemma 7.1, we have a completely order isomorphic copy $G$ of $S(X)$ inside $B(H \oplus K)$. By $[17]$ Corollary 4.2, there exists a surjective *-homomorphism from the C*-subalgebra $C_e^*(G)$ of $B(H \oplus K)$ generated by $G$, onto the C*-envelope $C_e^*(S(X))$, which fixes the copies of $S(X)$. Let $\phi : B(H \oplus K) \to I(S(X))$ be a completely positive map extending the *-homomorphism. Since $\phi$ fixes the diagonal scalars $C \oplus C$, it follows by 1.6 (this is a common argument), that $\phi$ decomposes as a $2 \times 2$ matrix of maps, each corner map defined on the corresponding ‘corner’ of $B(H \oplus K)$. In
particular, we have
\[ \phi\left(\begin{bmatrix} y & x \\ 0 & z \end{bmatrix}\right) = \begin{bmatrix} \theta(y) & j(x) \\ 0 & \pi(z) \end{bmatrix}, \]
for a map \( \theta : Y \to I_{11}, \pi : Z \to I_{22}, \) and for all \( x \in X, y \in Y, z \in Z. \) By the previous lemma, \( \phi \) is a ‘\( C^*(G) \)-bimodule map’. Hence for \( y \in Y, x \in X \) we have:
\[
\begin{bmatrix} 0 & j(yx) \\ 0 & 0 \end{bmatrix} = \phi\left(\begin{bmatrix} y & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & x \\ 0 & 0 \end{bmatrix}\right) = \phi\left(\begin{bmatrix} y & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & jx \\ 0 & 0 \end{bmatrix}\right) = \begin{bmatrix} \theta(y) & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & jx \\ 0 & 0 \end{bmatrix}.
\]
Thus \( j(yx) = \theta(y)j(x). \) The uniqueness of \( \theta \) follows from Corollary 1.3. Similarly for \( \pi. \)

For (ii), note that for \( y_1, y_2 \in Y \) and \( x \in X \) we have \( (y_1y_2)x = \theta(y_1)y_2)x = \theta(y_1)\theta(y_2)x. \) Now use Corollary 1.3 to conclude that \( \theta \) is a homomorphism. The last assertion follows from the fact that a contractive representation of a \( C^* \)-algebra is a *-homomorphism. \( \square \)

We recall that a right \( C^* \)-module \( Z \) over \( D \) is ‘self-dual’, if it satisfies the equivalent of the Riesz representation theorem for Hilbert spaces, namely that every \( f \in B_D(Z, D) \) is given by the inner product with a fixed \( z \in Z. \)

**Corollary 1.8.** If \( X \) is an operator space, then:

(i) \( I_{11} = M(C(X)) = LM(C(X)) = RM(C(X)) = QM(C(X)) = I(C(X)). \)

Thus \( I_{11} \cong B_D(I(X)) = B_D(I(X)). \) Similarly, \( I_{22} = M(D) = LM(D) = I(D). \)

(ii) \( I(X) \) is a self-dual right \( C^* \)-module (over \( D \) or over \( M(D) \)). Similarly, it is a self-dual left \( C^* \)-module.

(iii) \( I(S(X)) \) is the multiplier \( C^* \)-algebra of \( L(I(X)). \) Also, \( I(S(X)) \) is the injective envelope of \( L(I(X)). \)

**Proof.** (i): Represent \( L(I(X)) \) non-degenerately on a Hilbert space \( H \oplus K. \) We obtain a multiplication situation on \( H \oplus K \) given by the actions of \( LM(C) \) and \( RM(D) \) on \( I(X). \). By the multiplication theorem, we get a completely contractive homomorphism \( \theta : LM(C) \to I_{11} \) such that \( \theta(T)x = Tx, \)
for all \( x \in I(X) \) and \( T \in LM(C). \) Hence \( T = \theta(T) \in I_{11}. \) Thus by Corollary 1.2, we have \( I_{11} = M(C(X)) = LM(C(X)). \) By Lin’s theorem [24], the latter \( C^* \)-algebra may be identified with \( B_D(I(X)). \) Taking adjoints gives \( RM(C(X)) = M(C(X)). \) As noted in [25], (ii) below implies that \( M(C(X)) = QM(C(X)). \) Finally, note that we have the following \( C^* \)-subalgebras: \( C \subset I_{11} = M(C) \subset I(C). \) The last inclusion is a fact from [16] which is reproved at the end of our paper. By the injectivity of \( I_{11} \) there
exists a completely contractive projection $\Phi$ from $I(\mathcal{C})$ onto $I_{11}$; since $\Phi$ fixes $\mathcal{C}$ we see by rigidity that $\Phi = Id$ and $I_{11} = I(\mathcal{C})$.

(ii): We apply (i) but with $X$ replaced by the right C*-module sum $I(X) \oplus_c M(D)$. Clearly the latter is a full C*-module over $M(D)$, and it is injective since it is a corner of $I(S(X))$. Using (i) and Corollary A.7 in [7] if necessary, we obtain that $B_{M(D)}(I(X) \oplus_c M(D)) = B_{M(D)}(I(X) \oplus_c M(D))$, from which it is easy to see that $I(X)$ is a self-dual $M(D)$-module. However $B_D(I(X), D) = B_{M(D)}(I(X), M(D))$ by Cohen’s factorization theorem.

(iii): It is well-known that for any full right C*-module $Z$ over a C*-algebra $B$ say, we have that $K_D(Z \oplus_c D)$ is a copy of the linking C*-algebra of $Z$. Also $B_D(Z \oplus_c D) = M(K_D(Z \oplus_c D))$. Thus we have $M(L(I(X))) = B_D(I(X))\oplus D)$. This of course splits into four corners. The 1-1 corner is $B_D(I(X)) = I_{11}$ by (i). The 2-1 corner is $B_D(I(X), D) \cong I(X)^*$ by (ii). The 2-2 corner is $B_D(D, D) = M(D) = I_{22}$ by (i). This gives the first result. The second follows just as $I(\mathcal{C}) = I_{11}$ in (i). □

**Theorem 1.9.** If $X$ is an operator space, then:

(i) $IM_1(X)$ (resp. $IM_1^*(X)$) is completely isometrically isomorphic (resp.

- *-isomorphic) to the left multiplier algebras $M_l(X)$ (resp. $B_l(X) = A_l(X))$ defined in [7, Section 4].

(ii) $IM_1^*(X)$ is isometrically isomorphic to a closed subalgebra of $B(X)$.

Also, $IM_1^*(X)$ is completely isometrically isomorphic to a closed sub-

algebra of $B_l(X)$ or $CB_l(X)$.

Before we prove this, we define $B_l(X)$ and $CB_l(X)$ for an operator space $X$. Namely $B_l(X) = B(X)$ but with matrix norms

\[ \|T_{ij}\|_n = \sup \left\{ \left\| \left[ \sum_{k=1}^{n} T_{ik}(x_k) \right] \right\|_{C_n(X)} : x \in \text{BALL} \left( C_n(X) \right) \right\}. \]

(Here $C_n(X) = X^n$, but with the operator space structure one gets by

identifying $C_n(X)$ with the ‘first column’ of the operator space $M_n(X)$.)

With these norms $B_l(X)$ is not a matrix normed space in the traditional sense. However $M_n(B_l(X))$ is a unital Banach algebra. Similarly one defines $CB_l(X)$, the only difference being that one replaces $x_k$ in the expression above by $x_{(k,p),q}$. Here $[x_{(k,p),q}]$ is a matrix indexed on rows by $(k,p)$ and on

columns by $q$. Again $M_n(CB_l(X))$ is a unital Banach algebra.

**Proof.** (i): By 1.8, any $T \in IM_1(X)$ may be viewed as a bounded, and hence完全可以 bounded, module map on $I(X)$. We obtain a canonical sequence of completely contractive homomorphisms

\[ IM_1(X) \overset{\rho}{\to} M_l(X) \to CB(X) \]

given by restriction of domain. By Corollary 1.3, these homomorphisms are 1-1.
On the other hand, we have a multiplication situation given by the action
of $M_l(X)$ on $X$. Hence, by the multiplication theorem, there exists a
completely contractive homomorphism $\theta : M_l(X) \to I_{11}$ such that $\theta(T)x = Tx$
for all $T \in M_l(X), x \in X$. Thus $\rho \theta = Id$. Thus $\rho$ is onto, and since $\rho$ is 1-1
we obtain that $IM_l(X) \cong M_l(X)$ completely isometrically and as operator
algebras.

(ii) The first statement follows from a result which may be found in [37]
Proposition 1.1 or [5] Corollary 1, which asserts that any contractive homo-
morphism from a C*-algebra into a Banach algebra, is a $*$-homomorphism
onto its range, which is a C*-algebra (with the norm and algebra structure
inherited from the Banach algebra). Hence the canonical homomorphism
$B_l(X) \to CB(X)$ (or $B_l(X) \to B(X)$) is a $*$-homomorphism onto a C*
-algebra. Since the homomorphism is 1-1 it is therefore isometric.

The second statement follows by considering the following canonical iso-
metric inclusions

$$M_n(B_l(X)) \subset M_n(B_D(I(X))) \subset M_n(B_l(I(X))),$$
the last inclusion following from Section 3 of [26]. Thus by restriction
of domain, we get a contractive unital 1-1 homomorphism $M_n(B_l(X)) \to
M_n(B_l(X))$. Now we can apply [5] Corollary 1 to deduce that this last
homomorphism is an isometry.

A similar argument works for $CB_l$. □

It follows from (i) and a result in [7, Section 4], that the subalgebra of
$B(X)$ or $CB(X)$ corresponding to $IM_l^*(X)$ by (ii) above, is the C*-algebra
of (left) adjointable operators $A_l(X)$ on $X$.

We do not know whether $IM_l^*(X)$ is completely isometrically contained
inside $CB(X)$ in general. However it is not hard to find examples showing
that the canonical contraction $M_l(X) \to B(X)$ (or into $CB(X)$) is not
an isometry in general (see [7]). This shows that if $T \in B_D(I(X))$, with
$T(X) \subset X$, then one cannot expect anything like $\|T\|_{cb} = \|T\|_{cb}$.

Finally we remark that if $IM_l(X) = \mathbb{C}$, then it follows from the multi-
plication theorem that for any linear complete isometry $i : X \to B(K,H)$
such that $[i(X)K] = H$, we have that scalar multiples of $I_H$ are the only
operators $T \in B(H)$ such that $Ti(X) \subset i(X)$.

2. Applications.

We recall from [7] that an operation of an operator space $Y$ on an operator
space $X$, is a bilinear map $\circ : Y \times X \to X$, such that

1. $\|y \circ x\|_n \leq \|y\|_n \|x\|_n$, for all $n \in \mathbb{N}, x \in M_n(X), y \in M_n(Y),$
2. there is an element $e \in Y_1$ such that $e \circ x = x$ for all $x \in X$.
In (1), \( y \circ x \) is computed by the usual rule for multiplying matrices.

We will apply the multiplication theorem from §1 to give a proof of the ‘operation theorem’ from [7] §5:

**Theorem 2.1.** Suppose that \( Y, X \) are operator spaces, and suppose that \( \circ : Y \times X \to X \) is an oplication, with ‘identity’ \( e \in Y \). Then there exists a unique completely contractive linear map \( \theta : Y \to M_1(X) \) such that \( y \circ x = \theta(y)x \), for all \( y \in Y, x \in X \). Also \( \theta(e) = 1 \). Moreover, if \( Y \) is, in addition, an algebra with identity \( e \), then \( \theta \) is a homomorphism if and only if \( \circ \) is a module action. On the other hand, if \( Y \) is a \( C^* \)-algebra (or operator system) with identity \( e \), then \( \theta \) has range inside \( \mathbb{B}_1(X) \), and is completely positive and *-linear.

**Proof.** As in [7] the difficult part is to prove the first statement, the existence of \( \theta \). (The uniqueness follows from 1.3.) Indeed it suffices to find Hilbert spaces \( H, K \), a complete isometry \( \Phi : X \to B(K, H) \), and a linear complete contraction \( \theta : Y \to B(H) \), such that \( \theta(e) = I \), and such that \( \Phi(y \circ x) = \theta(y)\Phi(x) \), for all \( y \in Y, x \in X \). For in that case \( \theta(Y), \Phi(X), \mathbb{C} \) would form a multiplication situation on \( H \oplus K \). Now use the multiplication theorem above, together with the fact that \( IM_1(X) \cong M_1(X) \), to obtain the existence of \( \theta \).

The existence of such \( \Phi \) etc., follows easily from Le Merdy’s proof of the ‘BRS’ theorem ([23] 3.3). Namely, first suppose that \( X \subset B(K) \). By the ‘multilinear Stinespring’ theorem of [14, 30], we may write \( y \circ x = \beta_1(y)\alpha_1(x) \), for completely contractive maps \( \alpha_1 : X \to B(K_1, H) \) and \( \beta_1 : Y \to B(H_1, K) \). Similarly, \( \alpha_1(y \circ x) = \beta_2(y)\alpha_2(x) \), where now \( \alpha_2 : X \to B(K, H_2) \) say. Inductively we obtain, \( \alpha_{k-1}(y \circ x) = \beta_k(y)\alpha_k(x) \), where \( \alpha_k : X \to B(K, H_k) \), say. Since \( e \circ x = x \), we see that each \( \alpha_k \) is a complete isometry. Let \( W = X \otimes K \), and define \( f_k : W \to H_k \) by \( f_k(x \otimes \zeta) = \alpha_k(x)(\zeta) \). It is easy to check that \( \|f_k(w)\|_{H_k} \leq \|f_{k+1}(w)\|_{H_{k+1}} \), for \( w \in W \), as in [23]. Hence (by the parallelogram law if necessary) it is clear that \( \lim_k \|f_k(\cdot)\|_{H_k} \) defines a seminorm which gives rise to a Hilbert space norm (on the quotient of \( W \) by the nullspace of the seminorm). Write this resulting Hilbert space as \( H \). There is an obvious map \( \theta : Y \to L(H) \), given by \( \theta(y)([x \otimes \zeta]) = [(y \circ x) \otimes \zeta] \). It is easy to see that this is completely contractive as in [23]. The map \( \Phi : X \to B(K, H) \) given by \( \Phi(x)(\zeta) = [x \otimes \zeta] \) is clearly a complete contraction, too. On the other hand, for any \( \zeta \in \text{Ball}(K) \), we have

\[
\|\Phi(x)\| \geq \|\Phi(x)(\zeta)\| = \lim_k \|\alpha_k(x)(\zeta)\| \geq \|\alpha_1(x)(\zeta)\|.
\]

Thus \( \|\Phi(x)\| \geq \|\alpha_1(x)\| = \|x\| \), showing that \( \Phi \) is an isometry. A similar argument shows that \( \Phi \) is a complete isometry. \( \square \)

As shown in [7], this theorem has very many consequences, containing as special cases, the ‘BRS’ theorem [10], and many other results. The original
proof of the above theorem in [7] was much more difficult. The first author has subsequently replaced that proof there with another, part of which is similar to the proof above, but avoids use of the ‘multiplication theorem’ above. The version in [7] also does not give the fact that $\text{IM}_l(X) = M_l(X)$.

A left operator module $X$ over a unital operator algebra $A$, is an operator space which is also a unitary left $A$-module (unitary means that $1x = x$ for all $x \in X$), such that $\|ax\| \leq \|a\|\|x\|$ for all matrices $a$ with entries in $A$ and $x$ with entries in $X$. A similar definition holds for right modules or bimodules.

One may deduce from the oplication theorem the following refinement of the Christensen-Effros-Sinclair representation theorem [13]. We will in fact give an independent proof (which is the essentially the same as the proof above, except that we use the original Christensen-Effros-Sinclair theorem (which is quite simple) instead of the Le Merdy argument). In any case note that the method shows that the $A$-$B$-action on an operator $A$-$B$-bimodule $X$, may be extended to an action on $I(X)$, making $I(X)$ an operator $A$-$B$-bimodule.

Theorem 2.2. Suppose that $A$ and $B$ are unital operator algebras, and that $X$ is an operator space and a unitary $A$-$B$-bimodule with respect to a bimodule action $m : A \times X \times B \to X$. The following are equivalent:

(i) $X$ is an operator $A$-$B$-bimodule.

(ii) There exist Hilbert spaces $H$ and $K$, and a linear complete isometry $\Phi : X \to B(K,H)$ and completely contractive unital homomorphisms $\theta : A \to B(H)$ and $\pi : B \to B(K)$, such that $\Phi(m(a,x,b)) = \theta(a)\Phi(x)\pi(b)$, for all $a \in A, x \in X$ and $b \in B$.

(iii) There exists unique completely contractive unital homomorphisms $\theta : A \to M_l(X)$ and $\pi : B \to M_r(X)$ such that $\theta(a)x\pi(b) = m(a,x,b)$ for all $a \in A, x \in X$ and $b \in B$.

Proof. That (i) is equivalent to (ii) is a restatement of the Christensen-Effros-Sinclair representation theorem.

(iii) $\Rightarrow$ (ii): Obvious.

(ii) $\Rightarrow$ (iii): $\theta(A), \Phi(X), \pi(B)$ form a multiplication situation on $H \oplus K$.

The result then follows by the multiplication theorem.

We will refer to a triple $(\Phi, \theta, \pi)$ as in (ii) above as a CES representation of the bimodule $X$. We will call it a faithful CES representation if $\theta$ and $\pi$ are also completely isometric. It is always possible, by an obvious direct sum trick, to find a faithful CES representation for an operator bimodule. For any CES representation of $X$ we obtain an ‘upper triangular $2 \times 2$ operator algebra’, namely

$$U(X) = \begin{bmatrix} \theta(A) & \Phi(X) \\ 0 & \pi(B) \end{bmatrix}.$$
We will write $U_e(X)$ for this algebra in the case that $(j, \theta_e, \pi_e)$ is the representation in (iii) above, into the multiplier algebras. Thus $U_e(X) \subset I(S(X))$.

If $A$ and $B$ are $C^*$-algebras, and if we take faithful CES representations and form $U(X)$, then it is clear from [36] that $U(X)$ as an abstract operator algebra (i.e., as an algebra and an operator space), is independent of the particular faithful CES representation. This is probably not true if $A, B$ are non-selfadjoint. However in either case, we can easily see that the triple $(j, \theta_e, \pi_e)$ given in (iii) above, is the ‘smallest’ CES representation of $X$:

Corollary 2.3. Suppose that $A$ and $B$ are unital operator algebras, and that $X$ is an $A$-$B$-operator bimodule. Let $(\Phi, \theta, \pi)$ be a CES representation of $X$, and let $U(X)$ be the corresponding upper triangular $2 \times 2$ operator algebra. Then there is a canonical completely contractive unital homomorphism $\phi : U_e(X) \to U(X)$, which takes each corner of $U(X)$ into the same corner for $U_e(X)$. Indeed $\phi$ induces completely contractive unital homomorphisms $\rho$ and $\sigma$, from $\theta_e(A)$ and $\pi_e(B)$, to $\theta_e(A)$ and $\pi_e(B)$ respectively, such that $\rho \circ \theta = \theta_e$ and $\sigma \circ \pi = \pi_e$.

Proof. This follows immediately from the proof of Theorem 1.7, and of the implication ‘(ii) $\implies$ (iii)’ of 2.2. The $\phi$ is as in the proof of Theorem 1.7, which is easily seen to be a homomorphism on $U(X)$. ∎

Note that if $\theta_e$ is 1-1, then it follows that $\theta$ is also 1-1. Notice also that if the action of $A$ on $X$ is faithful (i.e., if $aX = 0$ implies that $a = 0$), then $\theta_e$ is 1-1. This follows from Corollary 1.3. If the action of $A$ on $X$ is ‘completely 1-faithful’ (that is, the norm of $a \in M_n(A)$ is achieved as the supremum of the norms of the action of $a$ on $X$ in a sensible way), then $\theta_e$ and $\theta$ in Corollary 2.3 are complete isometries.

Proposition 2.4. For an operator bimodule $X$, with the notations above, we have $I(U_e(X)) = I(S(X))$. If $X$ is an operator $A$-$B$-bimodule over $C^*$-algebras $A$ and $B$, which is faithful as a left and as a right module, and if $U(X)$ is the triangular operator algebra associated with a CES representation of $X$, then $I(U(X)) = I(S(X))$.

Proof. Clearly $U_e(X) \subset I(S(X))$. Any minimal $U_e(X)$-projection [17] is the identity on $S(X)$, and is therefore the identity map. The rest is clear. ∎

Definition 2.5. Let $Y$ be an operator $A$-$B$-bimodule. We shall call $Y$ an $A$-$B$-injective bimodule provided for every pair of operator $A$-$B$-bimodules $V$ and $W$, with $V$ a submodule of $W$, each completely contractive $A$-$B$-bimodule map $T : V \to Y$ extends to a completely contractive $A$-$B$-bimodule map from $W$ to $Y$.

We should remark that the above definition corresponds to what was called in [16] a tight $A$-$B$-injective bimodule.
The following result extends Wittstock’s theorem [39] that an injective C*-algebra is an injective operator module over a unital C*-subalgebra. In the following we consider unital C*-algebras, but it is not difficult to remove the ‘unital’ hypothesis.

**Theorem 2.6.** Let $A$ and $B$ be unital C*-algebras.

(i) If $Y$ is an operator space which is also an operator $A$-$B$-bimodule, then $Y$ is injective as an operator space if and only if $Y$ is an $A$-$B$-injective bimodule.

(ii) If $Y$ is an operator $A$-$B$-bimodule, then the operator space injective envelope $I(Y)$, is the operator $A$-$B$-bimodule injective envelope of $Y$. That is, $I(Y)$ is an $A$-$B$-injective bimodule which is rigid and essential, as an operator $A$-$B$-bimodule containing $Y$. Rigidity here, for example, means: Any completely contractive $A$-$B$-bimodule map $I(Y) \to I(Y)$, which is the identity on $Y$, is the identity on $I(Y)$.

*Proof.* One direction of (i) is obvious. Namely, suppose that $Y$ is $A$-$B$-injective. By CES, $Y$ may be realized as an $A$-$B$-submodule of some $B(K,H)$, where $H$ is a Hilbert $A$-module and $K$ is a Hilbert $B$-module. Indeed $B(K,H)$ is an operator $A$-$B$-bimodule. By the $A$-$B$-injectivity of $Y$, there is a completely contractive projection from $B(K,H)$ onto $Y$. Since $B(K,H)$ is injective as an operator space, so is $Y$.

The other direction of (i) is harder. In a previous version of this paper we had a proof which used almost all the results established until now. Instead we shall only use a few results from part 1, and 2.2 above. We will also use the fact, which is a simple consequence of Wittstock’s original result, or Suen’s modification of this result [36], that if $H$ is a Hilbert $A$-module, then for any Hilbert space $K$, we have that $B(K,H)$ is $A$-injective. Indeed their result gives that $B(H)$ is $A$-injective if $A \subset B(H)$ isometrically. However in the contrary case, one may use the following kind of trick: Pick a Hilbert space $H'$ in which $A$ is faithfully represented, then one obtains a faithful representation of $A$ on $H \oplus H'$. Then one may apply the Wittstock or Suen result to conclude that $B(H \oplus H')$ is $A$-injective, from which it is easy to see by compression that $B(H)$ is $A$-injective. See [6] Theorem 4.1 for another proof of this simple consequence.

Suppose that $Y$ is injective. Represent the C*-algebra $I(S(Y))$ faithfully and non-degenerately on a Hilbert space; then the two diagonal projections determine a splitting of the Hilbert space as $H \oplus K$, say. So $I_{11}$ is a unital *-subalgebra of $B(H)$, and so on. Now by injectivity there is a completely positive projection $\phi$ from $B(H \oplus K)$ onto $I(S(Y))$. As in the proof of 1.7, the Choi Lemma implies that this projection is an $I(S(Y))$-module map, and that $\phi$ decomposes as a $2 \times 2$ matrix of maps. Let $\psi$ be the ‘1-2 corner’ of $\phi$. Thus $\psi : B(K,H) \to Y$ is a completely contractive projection onto $Y$, and its easy to see, as in 1.7, that $\psi$ is a left $I_{11}$-module map. However
if $Y$ is an operator $A$-$B$-bimodule, then by 2.2 above there is a unital $*$-

homomorphism $\theta : A \to I_{11} \subset B(H)$ implementing the left module action.

Hence $H$ is a Hilbert $A$-module, via $\theta$. Since $\theta$ maps into $I_{11}$, we see that 

the projection $\psi$ is a left $A$-module map onto $Y$. Similarly, $\psi$ is a right $B$-module map onto $Y$. Since $B(K,H)$ is $A$-$B$-injective, so is $Y$.

(ii) is obvious, given (i) and the fact, observed earlier, that the $A$-$B$-action on $Y$ extends to make $I(Y)$ an operator $A$-$B$-bimodule.

□

The proof in fact shows that any injective operator space $X$ is an $I_{11} - I_{22}$-injective operator bimodule.

At this point we may give another proof of the self-duality of $I(X)$: By 2.6 (i), $I(X)$ is injective in the category of right $C^*$-modules over $I_{22}$. Now

appeal to [25] Prop. 3.10.

The following corollary of 1.8 generalizes a standard fact for Hilbert spaces:

**Corollary 2.7.** Let $Y$ be a right $C^*$-module over a $C^*$-algebra $A$, which (with respect to its canonical operator space structure) is an injective operator space. Then:

(i) $Y$ is a self-dual $C^*$-module over $A$.

(ii) $B_A(Y) = B_Y(Y)$.

(iii) Every bounded module map $Y \to Z$ is adjointable, for any other $C^*$-module $Z$.

(iv) $B_A(Y)$ is an injective $C^*$-algebra.

**Proof.** We may suppose w.l.o.g. that $Y$ is a full $C^*$-module over $A$. Thus we may regard $Y$ as a $\mathbb{K}(Y) - A$-imprimitivity bimodule. Let $Z = I(Y) = Y$ equipped with its $C(Y) - D(Y)$-imprimitivity bimodule structure. Here $C(Y) \subset I_{11}$ as usual. From [19] or A.7 in [7], we know that $Z \cong Y$, as imprimitivity bimodules. Thus $B_A(Y) \cong I_{11}$, giving (iv). Similarly we get (i). It is well-known, and fairly obvious, that (i) implies (ii) and (iii). □

The following result must be well-known. Since we cannot give a precise reference, we prove it:

**Proposition 2.8.** If $A$ is a $C^*$-algebra without identity, then the operator space injective envelope $I(A)$ is a unital $C^*$-algebra. Indeed $I(A) = I(A^1)$, where $A^1$ is the unitization of $A$.

**Proof.** Represent $A$ non-degenerately and faithfully on a Hilbert space $H$. If 

$\{e_i\}$ is a contractive approximate identity for $A$, then $e_i \to I_H$ in the SOT. 

We also have $I_H \subset A^1 \subset B(H)$. Let $\Phi : B(H) \to I(A) \subset B(H)$ be a minimal $A$-projection. So $\Phi$ is a completely contractive idempotent map whose range contains $A$. If $\Phi(I) = I$ we would be done, since in that case $\Phi$ is completely
positive, and then we can deduce that $I(A)$ is a unital C*-algebra as in the introduction (i.e., by the Choi-Effros result quoted there).

In order to see that $\Phi(I) = I$, we choose $\zeta \in H$ with $\|\zeta\| = 1$. Let $\phi(T) = \langle \Phi(T)\zeta, \zeta \rangle$, for $T \in B(H)$. Then $\|\phi\| \leq 1$ and $\phi(e_i) \to 1$. It is no doubt well-known and easy to see that this implies that $\phi(I) = 1$. One way to do this is to write $\phi(T) = \langle \pi(T)\eta, \xi \rangle$, where $\pi$ is a unital *-representation of $B(H)$ on a Hilbert space $K$, and where $\eta, \xi \in \text{BALL}(K)$. Set $K' = [\pi(A)K]^{\prime}$, and let $P$ be the projection onto $K'$. Then it is easy to see that the net $\pi(e_i)$ has $P$ as a WOT limit point. For if $T$ is a WOT limit point of $\pi(e_i)$, then for $x, y \in K$ we have

$$
\langle Tx, y \rangle = \lim_i \langle \pi(e_i)x, y \rangle = \lim_i \langle \pi(e_i)x, Py \rangle = \langle x, Py \rangle = \langle Px, y \rangle.
$$

Thus $\langle P\eta, \xi \rangle = \lim_i \phi(e_i) = 1$. By the converse to Cauchy-Schwarz, $\xi = P\eta = P^2\eta = P\xi$. Thus

$$
\phi(I) = \langle \eta, \xi \rangle = \langle \eta, P\xi \rangle = \langle P\eta, \xi \rangle = 1.
$$

In [16], it is shown that for a C*-algebra $A$, there is a canonical inclusion $LM(A) \subset I(A)$. Indeed, for any essential ideal $K$ of $A$, we have $LM(K) \subset I(A)$. We can offer another proof of these results using our methods. Firstly, we just saw that if $X = A$ considered as an operator space, then $I(X)$ is a unital injective C*-algebra. From this it follows from abstract principles that $I(S(X)) = I(S(I(X))) = M_2(I(A))$ (see for example [7] 4.18 (i)). Thus in this case $I_{11} = I(A)$. Hence

$$
LM(A) = M_1(A) = IM_1(A) \subset I(A).
$$

If $K$ is an essential ideal in $A$, then as we just proved, $M(K) \subset I(K)$. Thus we have

$$
K \subset A \subset M(K) \subset I(K).
$$

Any minimal $A$-projection on $I(K)$, is a $K$-projection, and is therefore equal to the identity map, by rigidity of $I(K)$. Thus $I(K) = I(A)$. By the first part, $LM(K) \subset I(K) = I(A)$. This gives the result we need.

Notice from the above, that since $A$ is an essential ideal in $M(A)$, we have $I(A) = I(M(A)) = I(A^1)$.

The same is true even if $A$ is a nonselfadjoint operator algebra with contractive approximate identity, with some modifications in proof.

If $X$ is a C*-module we see from results above that

$$
I(\mathbb{B}(X)) = I(M(\mathbb{K}(X))) = I(\mathbb{K}(X)),
$$

where $\mathbb{K}(X)$ is the so-called ‘imprimitivity C*-algebra’ of so-called ‘compact’ maps on $X$. Does this equal $I_{11}$ in this case? Certainly $\mathbb{K}(X) \subset I_{11}$ as a *-subalgebra, so that $I(\mathbb{K}(X)) \subset I_{11}$. 

Since this paper was written further advances have been made, for example in the papers [8, 9].

References


Received October 5, 1999 and revised January 31, 2000. Portions of this work were presented at the Canadian Operator Theory and Operator Algebras Symposium, May 20, 1999. The authors were supported by a grant from the NSF.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON
HOUSTON, TX 77204-3476
E-mail address: dblecher@math.uh.edu

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON
HOUSTON, TX 77204-3476
E-mail address: vern@math.uh.edu
EXISTENCE AND NONEXISTENCE OF INTERIOR-PEAKED SOLUTION FOR A NONLINEAR NEUMANN PROBLEM

DAOMIN CAO, E.S. NOUSSAIR, AND SHUSEN YAN

We show that the critical problem
\[
\begin{aligned}
-\Delta u + \lambda u &= u^{2^* - 1} + au^{q-1}, \quad u > 0 \text{ in } \Omega, \\
\frac{\partial u}{\partial \nu} &= 0 \quad \text{on } \partial \Omega, \quad 2 < q < 2^* = \frac{2N}{N-2},
\end{aligned}
\]
has no positive solutions concentrating, as \( \lambda \to \infty \), at interior points of \( \Omega \) if \( a = 0 \), but for a class of symmetric domains \( \Omega \), the problem with \( a > 0 \) has solutions concentrating at interior points of \( \Omega \).

1. Introduction.

In this paper we consider the semilinear Neumann problem
\[
\begin{aligned}
-\Delta u + \lambda u &= |u|^{2^* - 2}u + a|u|^{q-2}u \quad \text{in } \Omega, \\
\frac{\partial u}{\partial \nu} &= 0 \quad \text{on } \partial \Omega,
\end{aligned}
\]
(1.1)
where \( \Omega \subset \mathbb{R}^N (N \geq 3) \) is a bounded domain with smooth boundary having certain symmetries (to be specified below), \( \lambda > 0, \ a \geq 0, \) are constants, \( \nu \) is the unit outer normal to \( \partial \Omega \), \( 2^* = \frac{2N}{N-2} \) and \( 2 < q < 2N/(N-2) \).

Certain physical as well as biological situations have been modeled by such boundary value problems. We refer to [9], for such applications.

Much progress on the existence and qualitative behaviour of solutions for (1.1) with \( a = 0 \) has been made in the last few years. To state the results, we define the energy functional associated with (1.1):
\[
I_\lambda(u) = \frac{1}{2} \left( |\nabla u|^2 + \lambda u^2 \right) - \frac{1}{2^*} \int_\Omega |u|^{2^*} - \frac{a}{q} \int_\Omega |u|^q, \quad u \in H^1(\Omega).
\]
(1.2)

A solution of (1.1) is called a least energy solution if it minimizes the functional \( I_\lambda(u) \) in \( \{ u \in H^1(\Omega) : \langle I'_\lambda(u), u \rangle = 0, \ u \neq 0 \} \). Of particular interest has been the existence of solutions which exhibit a concentration phenomenon when \( \lambda \to \infty \). This was first shown for the problem
\[
\begin{aligned}
-\Delta u + \lambda u &= u^{p-1}, \quad u > 0 \text{ in } \Omega, \\
\frac{\partial u}{\partial \nu} &= 0 \quad \text{on } \partial \Omega,
\end{aligned}
\]
(1.3)
in the subcritical case $2 < p < 2N/(N - 2)$, by Lin and Ni in [10], Lin, Ni and Takagi in [11], Ni and Takagi in [13, 14], where it is shown that a nonconstant least energy solution exists for $\lambda$ sufficiently large, and that such a solution must have only one maximum point on $\Omega$, which must be on the boundary $\partial \Omega$ and tends to the point of maximum mean curvature of the boundary as $\lambda$ goes to $\infty$.

Similar studies have also been carried out for (1.3) in the critical case ($p = 2N/(N - 2)$). In particular, Adimurthi and Mancini [1], Adimurthi, Mancini and Yavada [2], Adimurthi, Pacella and Yavada [4], and Wang [24, 26] have established the existence and multiplicity of solutions. In [3] as well as in [27], the concentration behavior of such solutions is also established.

A solution of (1.1) is said to be a low-energy solution if its energy is less than $\frac{S}{2^{\frac{N}{2}}}$, where $S$ is the Sobolev constant. It has been shown in [4], [15], that all such solutions share a common property, namely that they are single-peaked on $\partial \Omega$ in the sense that each solution attains its maximum on $\Omega$, only at one point on $\partial \Omega$. In [2] and [20], among other results, the existence of solutions of (1.3) concentrating at a nondegenerate critical point of the mean curvature of $\partial \Omega$ as $\lambda$ goes to infinity has been established when $N \geq 6$. It has also been shown in [4], [27] for $N \geq 5$, that if $P_0$ is a strict local maximum point of the mean curvature function of the boundary $\partial \Omega$, then for $\lambda$ sufficiently large, there exists a nonconstant low energy solution of (1.3) with $p = 2N/(N - 2)$, with its peaks $P_\lambda \in \partial \Omega$ converging to $P_0$ as $\lambda \to \infty$. See also [28] for a survey. The same result was proved in [21] for $N = 3$.

In addition to the study of low energy solutions of (1.2) with $2 < p \leq 2N/(N - 2)$, recently much effort has been made to study the existence and qualitative behaviour of higher energy solutions, see for example [20], [25], [26], [27], [29]. For the special case when the domain has certain symmetries, Wang in [25], [29], Maier-Paape, K. Schmitt, Z.-Q. Wang in [16], showed the existence of solutions of (1.3), ($p = 2N/(N - 2)$) with multiple peaks (of related symmetry), located on $\partial \Omega$.

Recently Z.-Q. Wang [25] showed that if $\Omega = -\Omega$, $N = 3, 4$, then for $\lambda$ sufficiently large, there is a solution of (1.3), $p = 2N/(N - 2)$, which concentrates at either two points on $\partial \Omega$ or only at the origin. Wang’s result does not specify which of the two cases occurs.

To the best of the authors knowledge there does not seem to be any result on the existence of solutions of (1.3), $p = 2N/(N - 2)$, with interior peaks. In the sub-critical case there are some results on the existence and location of interior peaks. See for example [8], [30], [31]. The purpose of this paper is to study the existence of positive solutions of (1.1) which concentrate at interior points of $\Omega$. 
One of our main results in this paper, Theorem 1.2 shows that if \( a = 0 \) in (1.1), then there are no positive solutions with interior peaks. On the other hand, if \( a > 0 \), we show, Theorem 1.1, that, under some symmetry conditions on \( \Omega \), (1.1) has, for large \( \lambda \), a positive solution \( u_\lambda \) and \( u_\lambda \) concentrates at an interior point of \( \Omega \) as \( \lambda \) goes to infinity.

To state our main results, let us first introduce some notations.

For the existence of an interior-peaked solution we shall make the following symmetric assumption on \( \Omega \).

\((S)\) There is a subgroup \( G \subset O(N) \), the orthogonal group in \( \mathbb{R}^N \), such that \( \Omega \) is \( G \)-invariant, that is, \( x \in \Omega \) if and only if \( Gx \in \Omega \), \( 0 \in \Omega \), and for any \( x \in \Omega \setminus \{0\} \), \( \#Gx \geq 3 \). Here \( \#Gx \) denotes the cardinal number of \( Gx \), the \( G \)-orbit of \( x \).

We approach the existence problem using variational methods. Let \( H^1_G \) be the subspace of \( H^1(\Omega) \), defined by

\[
H^1_G = \{ u \in H^1(\Omega) \mid gu(x) \triangleq u(g(x)) = u(x) \text{ for all } g \in G \}.
\]

Let \( I_\lambda(u) \) be defined by (1.2) for \( u \in H^1_G \). It is not difficult to check that there exists \( 0 \not\equiv u_0 \in H^1_G(\Omega) \) such that \( u_0 \geq 0 \) and \( I_\lambda(u_0) < 0 \).

Define

\[
\Gamma = \{ \gamma \in C^1([0,1], H^1_G) \mid \gamma(0) = 0, \gamma(1) = u_0 \}
\]

\[
c_\lambda = \inf_{\gamma \in \Gamma} \sup_{t \in [0,1]} I_\lambda(\gamma(t))
\]

\[
S = \inf \left\{ \frac{\int_\Omega |\nabla u|^2}{\left(\int_\Omega |u|^{2^*}\right)^{2/2^*}} \mid u \in H^1(\Omega) \setminus \{0\} \right\}.
\]

It is known [17] that critical points of \( I_\lambda \) in \( H^1_G \) are also critical points of \( I_\lambda \) in \( H^1(\Omega) \), and hence classical solutions of (1.1) by standard regularity arguments. The main difficulty in applying variational methods in this case is the fact that \( I_\lambda \) fails to satisfy the Palais-Smale condition. Our arguments for the existence proof are based on establishing that the value \( c_\lambda \) of \( I_\lambda \) in \( H^1_G \) lies in \( (-\infty, S^{N/2}/N) \). In such an interval \( I_\lambda \) satisfies the Palais-Smale condition. This will establish the existence of a solution \( u_\lambda \), which is a “global” minimizer of \( I_\lambda \) in \( \{ u \in H^1_G \mid \langle I'_\lambda(u), u \rangle = 0, u \neq 0 \} \). Using concentration-compactness arguments and the symmetry condition \((S)\), we show that \( I_\lambda(u_\lambda) \to S^{N/2}/N \) as \( \lambda \to \infty \), and that \( u_\lambda \) concentrates at exactly one interior point of \( \Omega \). More precisely, define for \( x \in \mathbb{R}^N \)

\[
U(x) = \frac{[N(N-2)]^{\frac{N-2}{2}}}{(1 + |x|^2)^{\frac{N-2}{2}}}
\]
and for $\mu > 0$, $y \in \mathbb{R}^N, u \in H^1(\mathbb{R}^N)$ set

$$u_{\mu,y}(x) = \mu \frac{N-2}{2} u(\mu(x-y)).$$

Our first main result is:

**Theorem 1.1.** Suppose $a > 0, \Omega$ satisfies condition (S) and $q \in (2, 2^*)$ for $N \geq 4$, $q \in (4, 6)$ for $N = 3$. Then (1.1) has a positive solution $u_\lambda$ for $\lambda$ sufficiently large. Furthermore,

$$\lim_{\lambda \to \infty} \|u_\lambda(x) - \mu_\lambda^{-\frac{N-2}{2}} U(\mu_\lambda x)\|_{H^1(\Omega)} = 0,$$

where $\mu_\lambda \to \infty$, as $\lambda \to \infty$.

Our second main result is:

**Theorem 1.2.** Suppose $N \geq 5$. Then (1.1) with $a = 0$ has no positive solution $u_\lambda$ such that for some integer $k \geq 1$

$$\lim_{\lambda \to \infty} \left\| u_\lambda - \sum_{i=1}^{k} U_{\mu_i,\lambda,y_i} \right\|_{H^1(\Omega)} = 0,$$

with $\mu_{i,\lambda} \to \infty$, $y_i^\lambda \in \Omega$, $y_i^\lambda \to y_i \in \Omega$, as $\lambda \to \infty$ and $y_i \neq y_j$ for $i, j = 1, \cdots, k, \ i \neq j$.

This paper is organized as follows. In Section 2, by exploiting the symmetry of $\Omega$, we establish the existence of an interior-peaked positive solution of (1.1) with $a > 0$, and prove Theorem 1.1. In Section 3 we prove Theorem 1.2 by a reduction procedure.

2. **Existence of interior-peaked positive solutions.**

In this section we give the proof of Theorem 1.1. This is accomplished by the following Lemmas.

**Lemma 2.1.** Let $u_0, c_\lambda$ be as in (1.5), (1.6). Then

(i) $c_\lambda$ does not depend on the choice of $u_0$.

(ii) $0 < c_\lambda = \inf \{ I_\lambda(u) \mid u \in H^1_G(0), \langle I_\lambda'(u), u \rangle = 0 \}$.

Lemma 2.1 follows by the same argument as in W.M. Ni and I. Takagi [13].

**Lemma 2.2.** Suppose $a > 0$. Let $u_0, c_\lambda$ be as in (1.5), (1.6). Then

$$0 < c_\lambda < \frac{S^{N/2}}{N}$$

for $q \in (2, 2^*)$ if $N \geq 4$ and for $q \in (4, 6)$ if $N = 3$. 

**Proof.** We only need to construct a suitable test function. Let
\[ w_\epsilon(x) = \frac{1}{(\epsilon + |x|^2)^{N/2}}, \quad \epsilon > 0. \]
Then, as in H. Brezis and L. Nirenberg [7], we have
\[ \sup_{t \geq 0} I_\lambda(tw_\epsilon) < \frac{S^N}{N} \]
provided \( \epsilon \) is sufficiently small. The same argument in [7] then implies that \( c_\lambda > 0 \) and \( c_\lambda < \frac{S^N}{N} \). This completes the proof of Lemma 2.2.

The following result can be established using arguments based on the concentration compactness principle of P.L. Lions [12], see also Struwe [22]. Similar results were given by Z-Q. Wang [25] and D. Pierotti and S. Terracini [18].

**Lemma 2.3.** Let \( \lambda > 0 \) be fixed. Let \( (u_n)_n \subset H^1(\Omega) \), be a Palais-Smale sequence for \( I_\lambda \) at level \( c \), i.e., \( I_\lambda(u_n) \to c \) and \( I_\lambda'(u_n) \to 0 \), as \( n \to \infty \). Then there exist integers \( k_1 \geq 0 \), \( k_2 \geq 0 \), a solution \( u^* \) of (1.1), \( x_i \in \Omega \), and sequences \( (\mu_i)_n \), \( \mu_i \to \infty \), as \( n \to \infty \), \( i = 1, 2, \cdots, k_1 + k_2 \), such that, up to a subsequence,
\begin{align}
& u_n \to u^* \quad \text{weakly in } H^1(\Omega), \quad n \to \infty, \\
& \left\| u_n - u^* - \sum_{i=1}^{k_1+k_2} \frac{V_i}{\mu_i} x_i \right\|_{H^1(\Omega)} \to 0, \quad n \to \infty, \\
& c = I_\lambda(u^*) + \sum_{i=1}^{k_1} \left( \frac{1}{2} \int_{\mathbb{R}^N} |\nabla V_i|^2 - \frac{1}{2^*} \int_{\mathbb{R}^N} |V_i|^{2^*} \right) \\
& \quad + \sum_{i=k_1+1}^{k_1+k_2} \left( \frac{1}{2} \int_{\mathbb{R}^N} |\nabla V_i|^2 - \frac{1}{2^*} \int_{\mathbb{R}^N} |V_i|^{2^*} \right).
\end{align}
Furthermore, we have
\begin{align}
& \mu_i d(x_n, \partial \Omega) \leq C < \infty, \quad i = 1, \cdots, k_1, \\
& \quad \text{and} \quad \mu_i d(x_n, \partial \Omega) \to \infty, \quad n \to \infty, \quad i = k_1 + 1, \cdots, k_1 + k_2.
\end{align}
For \( i = 1, \cdots, k_1 \), \( |\nabla V_i| \in L^2(\mathbb{R}^N_+) \), and \( V^i \) is a nonzero solution of
\begin{align}
-\Delta V = |V|^{2^* - 2} V & \quad x \in \mathbb{R}^N_+, \\
\frac{\partial V}{\partial x_N} = 0 & \quad x_N = 0.
\end{align}
For \( i = k_1 + 1, \cdots, k_1 + k_2 \), \( |\nabla V^i| \in L^2(\mathbb{R}^N) \), and \( V^i \) is a nonzero solution of
\begin{align}
-\Delta V = |V|^{2^* - 2} V & \quad x \in \mathbb{R}^N.
\end{align}
Proof. By the assumptions on the sequence \( u_n \), it is easy to show that \( \{u_n\} \) is bounded in \( H^1(\Omega) \). Thus there exist \( u_* \in H_G \) such that by choosing a subsequence we have

\[
u_n \to u_* \quad \text{weakly in} \quad H^1(\Omega).
\]

Thus \( u_* \) is a weak solution of (1.1). Let \( v_n = u_n - u_* \), then \( v_n \) is a (P.S) sequence of \( I^\infty \) defined by

\[
I^\infty(v) = \frac{1}{2} \int_\Omega |\nabla v|^2 - \frac{1}{2^*} \int_\Omega |v|^{2^*}.
\]

Furthermore, \( I^\infty(v_n) \to c - I_1(u_*) \). We can finish our proof by exactly the same argument used in the proof of Lemma 3.3 in [18] (taking \( \delta = 0 \) in [18]). For readers’ convenience we give here a sketch in the following.

Let \( B_r(x) \) denote the ball in \( \mathbb{R}^N \) centered at \( x \) with radius \( r > 0 \). We will extend function \( v \in H^1(\Omega) \) to \( H^1(\mathbb{R}^N) \) by the extension operation \( E \) such that

\[
E(v)|_{\Omega} = v, \quad \|E(v)\|_{H^1(\mathbb{R}^N)} \leq C(\Omega)\|v\|_{H^1(\Omega)}.
\]

We will denote by the same symbol \( v \) and its extension \( E(v) \).

Let \( D^{1,2}(\mathbb{R}^N) \) be the closure of the set of all the functions in \( C_0^\infty(\mathbb{R}^N) \) with the norm \( \|u\| = \|\nabla u\|_{L^2(\mathbb{R}^N)} \).

By applying the concentration compactness principle there exist \( \mu_n > 0, x_n \in \Omega, \tau > 0 \) such that

\[
\widetilde{v}_n = \mu_n^{\frac{2-N}{2}} v_n \left( \frac{x}{\mu_n} + x_n \right),
\]

satisfies

\[
\sup \left\{ \int_{B_1(x)} |\nabla \widetilde{v}_n|^2 \left| \frac{x}{\mu_n} + x_n \in \Omega \right. \right\} = \int_{B_1(0)} |\nabla \widetilde{v}_n|^2 = \tau.
\]

Let \( \Omega_n = \{ x \in \mathbb{R}^N \mid \frac{x}{\mu_n} + x_n \in \Omega \} \). We then show that \( \widetilde{v}_n \) converges to \( V \) weakly in \( D^{1,2}(\mathbb{R}^N) \) and in \( H^1_{loc}(\mathbb{R}^N) \). As in [22] and [18] we can prove that the convergence actually holds in the strong \( H^1_{loc}(\mathbb{R}^N) \) sense. Thus, \( V \) is not identically zero. From the fact that \( v_n \) converges weakly to zero we get \( \mu_n \to \infty \). Hence, we have the following two cases:

1. \( \mu_n \text{dist}(x_n, \partial \Omega) \) goes to \( \infty \),
2. \( \mu_n \text{dist}(x_n, \partial \Omega) \) is bounded.

In the first case, \( \Omega_n \) tends to \( \mathbb{R}^N \) and \( V \) solves (2.6). In case (2), \( \Omega_n \) converges to a half space in the following sense: Up to a subsequence we can assume that the \( x_n \)'s converge to \( x_0 \in \partial \Omega \); moreover if \( y_n \) realizes the minimal distance of \( x_n \) from \( \partial \Omega \), we can assume that \( \mu_n (y_n - x_n) \to z_0 \). Let \( W \) be a neighborhood of \( x_0 \) and \( \Phi : U \subset \mathbb{R}^N \to W \) be a diffeomorphism such that \( \Phi(U \cap \mathbb{R}^N_+) = W \cap \Omega \) and \( \Phi(U \cap \mathbb{R}^N_-) = W \cap \partial \Omega \), and let \( \xi_n = \Phi^{-1}(y_n) \to \xi_0 = \Phi^{-1}(x_0) \). Define \( \Phi_n(\xi) = \mu_n \{ \Phi(\frac{\xi}{\mu_n} + \xi_n) - x_n \} \). The
Φ_n’s are charts for $\tilde{\Omega}_n$ and converges $C^1$ on compact subsets of $\mathbb{R}^N$ to linear differmorohism $z_0 + d\Phi(\xi_0)$. In this sense the boundaries $\partial\tilde{\Omega}_n$ converge to the hyperplane $(z - z_0) \cdot \nu(x_0) = 0$. Of course we can change our coordinate system over $\mathbb{R}^N$ in such a way that $z_0 = 0$ and $\nu(x_0) = -e_N$. With this convention we conclude that $V$ solves (2.5). We can finish our proof by iterating the above procedure. □

**Remark 2.4.** Notice that,

$$\frac{1}{2} \int_{\mathbb{R}^N} |\nabla V_i|^2 - \frac{1}{2^*} \int_{\mathbb{R}^N} |V_i|^{2^*} \geq \frac{1}{2N} S^{N/2}$$

for $i = 1, \ldots, k_1$

$$\frac{1}{2} \int_{\mathbb{R}^N} |\nabla V_i|^2 - \frac{1}{2^*} \int_{\mathbb{R}^N} |V_i|^{2^*} \geq \frac{1}{N} S^{N/2}$$

for $i = k_1 + 1, \ldots, k_1 + k_2$.

**Remark 2.5.** Notice that, by (2.2), $u_n - u_*$ concentrates, as $n \to \infty$, at the points $x_i^n$, $i = 1, \ldots, k_1$. By (2.4) the points $x_i^n$, $i = 1, \ldots, k_1$, approach $\partial\Omega$. It then follows from $u_n - u_* \in H^1_G$, and condition (S), that either $k_1 = 0$ or $k_1 \geq 3$.

**Lemma 2.6.** $I_\lambda$ satisfies the (P.S)$_c$ condition in $H^1_G(\Omega)$ for $c \in (-\infty, S^{N/2}/N)$.

**Proof.** Suppose $\{u_n\} \subset H^1_G(\Omega)$, $u_n \geq 0$, is a (P.S)$_c$ sequence, with $c < S^{N/2}/N$. It follows from Lemma 2.3 that there are integers $k_1, k_2 \geq 0$, a solution $u_*$ of (1.1), and sequences $x_i^n \in \Omega$ and $\mu_i^n \to \infty$, such that (2.2)-(2.4) hold.

We first notice that $k_2 = 0$ since, otherwise, $c \geq S^{N/2}/N$, contradicting our hypothesis on $c$. Notice also (see Remark 2.4) that $k_1 = 0$ since, otherwise $k_1 \geq 3$, which implies that $c \geq \frac{3S^{N/2}}{2N}$ contradicting our hypothesis on $c$. It then follows from (2.2) that $u_n \to u_*$ strongly in $H^1(\Omega)$, as $n \to \infty$. This completes the proof of Lemma 2.6. □

**Lemma 2.7.** Assume $\Omega$ satisfies condition (S) and $q \in (2, 2^*)$ for $N \geq 4$, $q \in (4, 6)$ for $N = 3$. Then $c_\lambda \to S^{N/2}/N$, as $\lambda \to \infty$.

**Proof.** To prove Lemma 2.7, we only need to show, by Lemma 2.2, that

$$\lim_{\lambda \to \infty} c_\lambda \geq \frac{S^{N/2}}{N}.$$  \hspace{1cm} (2.7)

We argue by contradiction. Suppose that we have a sequence $\lambda_n \to \infty$, as $n \to \infty$, such that

$$c_{\lambda_n} \to c < \frac{S^{N/2}}{N}.$$  \hspace{1cm} (2.8)

It follows from Lemma 2.2 and Lemma 2.6 that for each $\lambda_n$ ($n$ large enough) we have a sequence $\{u_n\} \subset H^1_G(\Omega)$ satisfying

$$I_{\lambda_n}(u_n) = c_{\lambda_n} \to c < \frac{S^{N/2}}{N}.$$
On the other hand, we have $I_{\lambda_n}(|u_n|) = I_{\lambda_n}(u_n)$, which means that $|u_n|$ achieves $\inf\{I_{\lambda_n}(u) : u \in H^1_G \backslash \{0\}, \langle I'_{\lambda_n}(u), u \rangle = 0\}$. By Langrange principle we have $I'(u_n) = 0$. Thus, without loss of generality we can assume that $u_n \geq 0, u_n \neq 0$ and satisfies

$$
\begin{cases}
-\Delta u_n + \lambda_n u_n = u_n^{2^* - 1} + au_n^{q - 1} & \text{in } \Omega, \\
\frac{\partial u_n}{\partial \nu} = 0 & \text{on } \partial \Omega.
\end{cases}
$$

By the standard regularity procedure and maximum principle we have $u_n > 0$ in $\Omega$. It follows from (2.8), (2.9), and the Hölder’s inequality, that

$$0 < \ell_1 < \int_{\Omega} |\nabla u_n|^2 + \lambda_n u_n^2 < \ell_2,$$

uniformly for $n$, where $\ell_1$ and $\ell_2$ are some positive constants, independent of $n$. Hence by passing to a subsequence if necessary, we may assume that $u_n \rightharpoonup 0$ weakly in $H^1(\Omega)$, and by the compact embedding of $H^1(\Omega) \hookrightarrow L^2(\Omega)$, $\int_{\Omega} u_n^2 \to 0$ as $\lambda_n \to \infty$. Set

$$v_n(x) = \lambda_n^{\frac{N-2}{4}} u_n \left( \frac{x}{\sqrt[2]{\lambda_n}} \right), \quad x \in \Omega_n \triangleq \left\{ x \mid \frac{x}{\sqrt[2]{\lambda_n}} \in \Omega \right\}.$$

Then $v_n(x)$ satisfies

$$
\begin{cases}
-\Delta v_n + v_n = v_n^{2^* - 1} + a \frac{v_n^{q - 1}}{\lambda_n} & \text{in } \Omega_n, \\
v_n > 0 & \text{in } \Omega_n, \\
\frac{\partial v_n}{\partial \nu} = 0 & \text{on } \partial \Omega_n,
\end{cases}
$$

where

$$\sigma = \frac{2^* - q}{4} (N - 2) > 0.$$

We also have

$$c_{\lambda_n} = \frac{1}{2} \int_{\Omega_n} |\nabla v_n|^2 + v_n^2 - \frac{1}{2^*} \int_{\Omega_n} |v_n|^{2^*} - \frac{a}{\frac{q}{4} \lambda_n} \int_{\Omega_n} |v_n|^q.$$

Set

$$\psi_n(x) = \chi_{\Omega_n} |v_n|^{2^*},$$

where $\chi_{\Omega_n}$ is the characteristic function of $\Omega_n$. Then $\int_{\mathbb{R}^N} \psi_n(x) dx = \int_{\Omega_n} |v_n|^{2^*}$. Without loss of generality, we may assume that $\int_{\Omega_n} |v_n|^{2^*} \to \ell > 0$, as $n \to \infty$, by (2.10). We now apply the concentration-compactness principle of P.L. Lions \cite{12} to $\psi_n(x)$. We then have that one of the following three cases must occur:

(i) vanishing
(ii) compactness
(iii) dichotomy.
If (i) occurs, then, by a result of Z.-Q. Wang, Lemma 4.6 in [26], we have \( \int_{\Omega_n} |v_n|^{2^*} \to 0 \), as \( n \to \infty \), which leads to \( \int_{\Omega_n} |\nabla v_n|^2 + v_n^2 \to 0 \), as \( n \to \infty \), contradicting (2.10). Thus (i) is excluded.

As in Z.-Q. Wang [27] (iii) implies \( c_{\lambda_n} \geq \frac{3S^{N/2}}{2^N} + o(1) \), since \( v_{\lambda_n} \in H^1_G(\Omega_n) \). This contradicts (2.8).

We show next that (ii) cannot occur. Suppose (ii) occurs, that is, there is a sequence \( \{y_n\} \), \( y_n \in \Omega_n \) such that for all \( \epsilon > 0 \), there is \( R_{\epsilon} > 0 \), satisfying

\[
\int_{B_{R}(y_n) \cap \Omega_n} |v_n|^{2^*} \geq \ell - \epsilon \quad \text{for all } R \geq R_{\epsilon}.
\]

We claim that \( \{y_n\} \) is bounded. In fact, since \( v_n \in H^1_G(\Omega_n) \), if \( |y_n| \to \infty \), as \( n \to \infty \), we get a contradiction to (2.13). Hence suppose \( y_n \to y_0 \).

Set

\[
\mu_n = \left( \max_{x \in \Omega} v_n(x) \right)^{4/(N-2)}.
\]

We claim that

\[
\lim_{n \to \infty} \mu_n = \infty.
\]

To prove (2.14), suppose, to the contrary, that \( \{\mu_n\} \) is bounded. Since \( \int_{\Omega_n} |\nabla v_n|^2 + |v_n|^2 = \int_{\Omega} |\nabla u_n|^2 + \lambda_n u_n^2 \), is bounded by (2.10), we may assume that

\[
v_n \to v_0 \quad \text{weakly in } H^1(\mathbb{R}^N).
\]

Hence

\[
\int_{B_{R}(y_n) \cap \Omega_n} v_n^2 \geq \frac{1}{\mu_n^{(2^*-2)/(2^* - 2)}} \int_{B_{R}(y_n) \cap \Omega_n} v_n^{2^*} = \frac{1}{\mu_n} \int_{B_{\ell_{\epsilon}}(y_n) \cap \Omega_n} v_n^{2^*} \geq \frac{1}{\mu_n} (\ell - \epsilon).
\]

On the other hand,

\[
\int_{B_{2R}(y_0)} v_0^2 \geq \lim_{n \to \infty} \int_{B_{R}(y_n) \cap \Omega_n} v_n^2 \geq \alpha > 0
\]

by (2.16), for some \( \alpha > 0 \), which implies that \( v_0 \geq 0 \), \( v_0 \not\equiv 0 \). But \( v_0 \) solves

\[
\begin{cases}
-\Delta v_0 + v_0 = v_0^{2^*-1} & \text{in } \mathbb{R}^N \\
v_0 > 0 & \text{in } \mathbb{R}^N,
\end{cases}
\]
by (2.11), which is impossible, since the above equation has no solution. This proves (2.14). Set

\[ w_n(x) = \frac{\mu_n^{-(N-2)}}{4} v_n \left( \frac{x}{\sqrt{\mu_n}} \right), \quad x \in \sqrt{\mu_n} \Omega_n, \]

\[ \tilde{\Omega}_n \triangleq \sqrt{\mu_n} \Omega_n = \sqrt{\mu_n} \lambda_n \Omega. \]

Then \( w_n \) satisfies

\[
\begin{cases}
-\Delta w_n + \frac{1}{\mu_n} w_n = w_n^{2^*-1} + \frac{a_{\lambda_n}}{\lambda_n} v_n^{\theta - 1} & \text{in } \tilde{\Omega}_n, \\
w_n > 0 & \text{in } \tilde{\Omega}_n, \\
\frac{\partial w_n}{\partial \nu} = 0 & \text{on } \partial \tilde{\Omega}_n,
\end{cases}
\]

and

\[
\int_{\tilde{\Omega}_n} |\nabla w_n|^2 = \int_{\Omega_n} |\nabla v_n|^2 \\
\int_{\tilde{\Omega}_n} |w_n|^{2^*} = \int_{\Omega_n} |v_n|^{2^*}.
\]

By standard elliptic regularity arguments, we have

\[ w_n \in C^{2, \gamma}(D), \]

\[ w_n \rightharpoonup w_0 \quad \text{in } C^{2}_{\text{loc}}(\mathbb{R}^N), \]

for any compact subset \( D \subset \mathbb{R}^N \). Hence \( w_0 \) satisfies

\[
\begin{cases}
-\Delta w_0 = w_0^{2^*-1} & \text{in } \mathbb{R}^N, \\
|\nabla w_0| \in L^2(\mathbb{R}^N), \\
w_0 \in L^{2^*}(\mathbb{R}^N), \\
\max_{x \in \mathbb{R}^N} w_0(x) = 1.
\end{cases}
\]

By the uniqueness of solutions of the above problem, we conclude that \( w_0(x) = \mu_0^{\frac{2}{N-2}} U(\sqrt{\mu_0}(x - y)) \) for some \( \mu_0 > 0, y \in \mathbb{R}^N \), such that \( w_0(y) = \max_{x \in \mathbb{R}^N} w_0(x) \). Furthermore, we have

\[
\int_{\mathbb{R}^N} |\nabla w_0|^2 = S^{N/2},
\]

\[
\int_{\mathbb{R}^N} w_0^{2^*} = S^{N/2}.
\]

Therefore, from \( \int_{\mathbb{R}^N} |\nabla w_0|^2 \leq \lim_{n \to \infty} \int_{\tilde{\Omega}_n} |\nabla w_n|^2 \), we obtain

\[
\lim_{n \to \infty} \int_{\tilde{\Omega}_n} |\nabla w_n|^2 \geq S^{N/2},
\]
and
\begin{equation}
\lim_{n \to \infty} \left( \frac{1}{2} \int_{\tilde{\Omega}_n} |\nabla w_n|^2 + \frac{1}{\mu_n} w_n^2 - \frac{1}{2} \int_{\tilde{\Omega}_n} w_n^{2^*} - \frac{a}{q \lambda_n \mu_n} \int_{\tilde{\Omega}_n} w_n^q \right) \geq \frac{S^{N/2}}{N},
\end{equation}

since
\begin{equation}
\frac{1}{\lambda_n \mu_n^{\sigma}} \int_{\Omega_n} w_n^q \leq \frac{1}{\lambda_n^{\sigma}} \left( \frac{1}{2 \mu_n} \int_{\Omega_n} w_n^2 + \frac{1}{2} \int_{\Omega_n} w_n^{2^*} \right)
\end{equation}

by the Hölder’s inequality.

But (2.19) contradicts (2.8), and we conclude that compactness cannot occur. Therefore (2.7) follows. This completes proof of Lemma 2.7.

\section*{Proof of Theorem 1.1.}

Using Lemmas 2.2 and 2.6 we can obtain the existence of positive solution $u_\lambda \in H_0^1$. To establish (1.9), we follow the argument in Lemma 2.7 to show that only compactness can occur. By arguments very similar to that in the proof of Lemma 2.7 we show that there exists a sequence $\{\tilde{\mu}_\lambda\}$, $\tilde{\mu}_\lambda \to \infty$, such that
\begin{equation}
\tilde{\mu}_\lambda^{-\frac{N-2}{4}} u_\lambda \left( \frac{x}{\sqrt{\tilde{\mu}_\lambda}} \right) \to w_0 = \mu_0^{N-2} U(\sqrt{\mu_0}(x - y)), \quad \text{as } \lambda \to \infty,
\end{equation}
in $C^{2}_{\text{loc}}(\mathbb{R}^N)$, for some constant $\mu_0 > 0$, $y \in \mathbb{R}^N$, and
\begin{equation}
\left\| \tilde{\mu}_\lambda^{-\frac{(N-2)}{4}} u_\lambda \left( \frac{x}{\sqrt{\tilde{\mu}_\lambda}} \right) - \mu_0^{\frac{N-2}{4}} U(\sqrt{\mu_0}(x - y)) \right\|_{H^1(\Omega_\lambda)} \to 0,
\end{equation}
as $\lambda \to \infty$, where $\Omega_\lambda = \frac{1}{\sqrt{\tilde{\mu}_\lambda}} \Omega$.

By the symmetry assumption, we have $y = 0$, and
\begin{equation}
\left\| u_\lambda(x) - (\mu_0 \tilde{\mu}_\lambda)^{\frac{N-2}{4}} U(\sqrt{\mu_0 \tilde{\mu}_\lambda} x) \right\|_{H^1(\Omega)} \to 0,
\end{equation}
as $\lambda \to \infty$. This completes the proof of Theorem 1.1.

\section{Nonexistence results.}

In this section we show that (1.1) with $a = 0$ does not have solutions concentrating, as $\lambda \to \infty$, at interior points of $\Omega$.

Consider the problem
\begin{equation}
\begin{cases}
-\Delta u + \lambda u = u^{2^*-1}, & u > 0 \text{ in } \Omega, \\
\frac{\partial u}{\partial \nu} = 0 & \text{in } \partial \Omega.
\end{cases}
\end{equation}

Let $k = 1, 2, \ldots$, be fixed.
Define for $\mu = (\mu_1, \cdots, \mu_k) \in \mathbb{R}^k_+, y = (y^1, \cdots, y^k) \in \Omega^k$,

$$E^k_{\mu,y} = \left\{ w \in H^1(\Omega) \mid \left\langle U_{\mu_i,y^i}, w \right\rangle_\lambda = 0, \right. \\
\left. \left\langle \frac{\partial U_{\mu_i,y^i}}{\partial \mu_i}, w \right\rangle_\lambda = 0, \right. \\
\left. \left\langle \frac{\partial U_{\mu_i,y^i}}{\partial y^i_\ell}, w \right\rangle_\lambda = 0, \right. \\
i = 1, 2, \cdots, k; \quad \ell = 1, 2, \cdots, N \right\},$$

where $\langle u, v \rangle_\lambda = \int_{\Omega} \nabla u \nabla v + \lambda uv$. For $\delta > 0$, define

$$M^k_\delta = \left\{ (\alpha, \mu, y, \omega) : \alpha = (\alpha_1, \cdots, \alpha_k) \in \mathbb{R}^k_+, \mu = (\mu_1, \cdots, \mu_k) \in \mathbb{R}^k_+, \right. \\
y = (y^1, \cdots, y^k) \in \Omega^k, \omega \in E^k_{\mu,y}, \right. \\
|\alpha_i - 1| < \delta, \mu_i > \frac{1}{\delta}, \text{dist}(y^i, \partial \Omega) > \delta \right. \\
\left. \text{for } i = 1, \cdots, k, \|\omega\| < \delta \right\}.$$

In this section we will denote

$$\|u\|^2 = \int_{\Omega} |\nabla u|^2 + u^2 \quad \text{for } u \in H^1(\Omega),$$

$$\|u\|^q_q = \int_{\Omega} |u|^q \quad \text{for } q \geq 1, u \in L^q(\Omega).$$

Let for $(\alpha, \mu, y, \omega) \in M^k_\delta$, $L_\lambda$ be defined by

$$L_\lambda(\alpha, \mu, y, \omega) = I_\lambda \left( \sum_{i=1}^k \alpha_i U_{\mu_i,y^i} + \omega \right).$$

It is now well-known (see Bahri [5]) that if $\delta > 0$ is small enough, then for $(\alpha, \mu, y, \omega) \in M^k_\delta$, $u = \sum_{i=1}^k \alpha_i U_{\mu_i,y^i} + \omega$ is a critical point of $I_\lambda(u)$ in $H^1(\Omega)$ if and only if $(\alpha, \mu, y, \omega)$ is a critical point of $L_\lambda(\alpha, \mu, y, \omega)$ in $M^k_\delta$, which is equivalent to the existence of $a^i \in \mathbb{R}, b^i \in \mathbb{R}, c^i_\ell \in \mathbb{R}$ for $i = 1, \cdots, k.$
\[\ell = 1, \ldots, N,\] such that the following equations hold:

\[
\frac{\partial I_\lambda}{\partial \alpha_i} \left( \sum_{i=1}^{k} \alpha_i U_{\mu, y^i} + \omega \right) = 0, \quad i = 1, \ldots, k, \tag{3.6}
\]

\[
\frac{\partial I_\lambda}{\partial \omega} \left( \sum_{i=1}^{k} \alpha_i U_{\mu, y^i} + \omega \right) = \sum_{i=1}^{k} \left\{ a_i \left( U_{\mu, y^i} , \varphi \right)_\lambda + b_i \left( \frac{\partial U_{\mu, y^i}}{\partial \mu} , \varphi \right)_\lambda \right\} + \sum_{i=1}^{N} c_i \left( \frac{\partial U_{\mu, y^i}}{\partial y_i^\ell} , \varphi \right)_\lambda, \quad \forall \varphi \in H^1(\Omega), \tag{3.7}
\]

\[
\frac{\partial I_\lambda}{\partial y_j} \left( \sum_{i=1}^{k} \alpha_i U_{\mu, y^i} + \omega \right) = b_j \left( \frac{\partial^2 U_{\mu, y^j}}{\partial \mu^2} , \omega \right)_\lambda + \sum_{\ell=1}^{N} c_j^\ell \left( \frac{\partial^2 U_{\mu, y^j}}{\partial y_j^\ell \partial y_i^\ell} , \varphi \right)_\lambda, \quad j = 1, 2, \ldots, k. \tag{3.8}
\]

Before giving the proof of Theorem 1.2, we give the few inequalities that will be needed for our proofs.

By a result of Z.-Q. Wang [27, Lemma 2.3], we have for \( N \geq 5, m \geq 1, \) \( \mu \in \mathbb{R}^+ \), and \( \sigma \in (0,1) \) close to 1, a constant \( C(\sigma) > 0 \) which is bounded if \( \frac{1}{1-\sigma} \) is bounded, such that

\[
\left| \int_\Omega U_{\mu, y} \omega \right| \leq \left( 1 - \frac{\sigma}{2} \right) C(\sigma) \frac{m \pi^2}{\mu} \left\| \omega \right\|_2^2 + \frac{\sigma}{2 \pi^2} \left\| \omega \right\|_2^2 \tag{3.10}
\]

for any \( \omega \in H^1(\Omega) \), where \( \hat{\sigma} = \frac{2(1-\sigma)}{2-\sigma} \).

Direct computation shows that for any \( \omega \in E^k_{\mu, y} \) and \( \mu \in \mathbb{R}^+ \)

\[
\left| \int_\Omega U_{\mu, y} \omega \right| \leq \left( 1 - \frac{\sigma}{2} \right) C(\sigma) \frac{m \pi^2}{\mu} \left\| \omega \right\|_2^2 + \frac{\sigma}{2 \pi^2} \left\| \omega \right\|_2^2 \tag{3.11}
\]

From

\[
\left| \frac{\partial U_{\mu, y}}{\partial \mu} \right| = O \left( \frac{1}{\mu} \right) U_{\mu, y}, \quad \left| \frac{\partial U_{\mu, y}}{\partial y_i} \right| = O \left( \mu U_{\mu, y} \right),
\]

and
(3.10) and (3.11), we then have

\[
\left| \int_{\Omega} \frac{\partial U_{\mu,y}}{\partial \mu} \omega \right| = O \left( \frac{1}{\mu^3} \right) \| \omega \|_{2^*}^2 + O \left( \frac{1}{\mu} \right) \| \omega \|_2^2
\]

(3.12)

\[
\left| \int_{\Omega} \frac{\partial U_{\mu,y}}{\partial y_i} \omega \right| = O \left( \frac{1}{\mu} \right) \| \omega \|_{2^*}^2 + O(\mu) \| \omega \|_2^2,
\]

where \( \hat{\sigma} \) is as in (3.10).

\[
\left| \int_{\Omega} U_{\mu,y}^{2^*-2} \frac{\partial U_{\mu,y}}{\partial \mu} \omega \right| \leq O \left( \frac{1}{\mu \frac{N+2}{2}} \right) \| \omega \| + \lambda \left| \int_{\Omega} \frac{\partial U_{\mu,y}}{\partial \mu} \omega \right|,
\]

(3.14)

\[
\left| \int_{\Omega} U_{\mu,y}^{2^*-2} \frac{\partial U_{\mu,y}}{\partial y_i} \omega \right| \leq O \left( \frac{1}{\mu \frac{N+2}{2}} \right) \| \omega \| + \lambda \left| \int_{\Omega} \frac{\partial U_{\mu,y}}{\partial y_i} \omega \right|,
\]

(3.15)

for any \( \omega \in E_k^{\mu,y} \). (3.14) and (3.15) can be established as (3.11).

**Proof of Theorem 1.2.**

We argue by contradiction. Suppose (3.1) has a solution \( u_\lambda \) such that (1.10) holds. It follows from Bahri [5], Rey [19], that for \( \lambda \) large

\[
u_\lambda = \sum_{i=1}^{k} \alpha_{i,\lambda} U_{\mu_{i,\lambda},y_{\lambda}^i} + \omega_{\lambda},
\]

with \( \alpha_{i,\lambda} > 0, \mu_{i,\lambda} > 0, y_{\lambda}^i \in \Omega \) and \( \omega_{\lambda} \in E_k^{\mu_{\lambda},y_{\lambda}} \), for \( \mu_{\lambda} = (\mu_{1,\lambda}, \ldots, \mu_{k,\lambda}) \), \( y_{\lambda} = (y_{\lambda}^1, \ldots, y_{\lambda}^k) \), satisfying

\[
\alpha_{i,\lambda} \to 1 \quad \text{as} \quad \lambda \to \infty, \quad i = 1, \ldots, k,
\]

(3.17)

\[
\| \omega_{\lambda} \| \to 0 \quad \text{as} \quad \lambda \to \infty,
\]

(3.18)

\[
y_{\lambda}^i \to y^i \quad \text{as} \quad \lambda \to \infty \quad \text{for} \quad i = 1, \ldots, k.
\]

(3.19)

To simplify our notations, we write \( \alpha_{i,\lambda}, y_{\lambda}^i \) and \( \omega_{\lambda} \) simply as \( \alpha_i, \mu_i, y^i \) and \( \omega \), respectively, for \( i = 1, \ldots, k \). By taking \( \varphi = \omega \) in (3.7), we have

\[
\int_{\Omega} \nabla \left( \sum_{i=1}^{k} \alpha_i U_{\mu_i,y^i} + \omega \right) \nabla \omega + \lambda \left( \sum_{i=1}^{k} \alpha_i U_{\mu_i,y^i} + \omega \right) \omega
\]

\[
= \int_{\Omega} \left| \sum_{i=1}^{k} \alpha_i U_{\mu_i,y^i} + \omega \right|^{2^* - 2} \left( \sum_{i=1}^{k} \alpha_i U_{\mu_i,y^i} + \omega \right) \omega,
\]
which gives

\[(3.20) \quad \int_\Omega |\nabla \omega|^2 + \lambda \omega^2 = \sum_{i=1}^k \alpha_i^{2^* - 1} \int_\Omega U_{\mu_i, y_i}^{2^* - 1} \omega + (2^* - 1) \sum_{i=1}^k \alpha_i^{2^* - 2} \int_\Omega U_{\mu_i, y_i}^{2^* - 2} \omega^2 + O(\|\omega\|_2^{2+\delta}),\]

for some \( \delta \in (0, \min\{2^* - 2, 1\}) \). From (3.10), (3.11), and (3.20), we obtain

\[(3.21) \quad \int_\Omega |\nabla \omega|^2 + \lambda \omega^2 - \sum_{i=1}^k \frac{\lambda \sigma}{2m_{2/\sigma}} \alpha_i \int_\Omega \omega^2 - (2^* - 1) \sum_{i=1}^k \alpha_i^{2^* - 2} \int_\Omega U_{\mu_i, y_i}^{2^* - 2} \omega^2 = O\left(\sum_{i=1}^k \frac{\lambda}{\mu_i^2}\right) ||\omega||_2^{\delta} + O\left(\sum_{i=1}^k \frac{1}{\mu_i^{N/2}}\right) ||\omega||.\]

And from a result of Bahri [5], we have, for \( \lambda \) large,

\[(3.22) \quad \int_\Omega |\nabla \omega|^2 + \frac{1}{2} \lambda \int_\Omega \omega^2 - (2^* - 1) \sum_{i=1}^k \alpha_i^{2^* - 2} \int_\Omega U_{\mu_i, y_i}^{2^* - 2} \omega^2 \geq \rho_0 \int_\Omega |\nabla \omega|^2,\]

where \( \rho_0 > 0 \) is a constant, independent of \( \lambda, \mu_i, y_i, i = 1, \ldots, k.\)

Taking \( m \) suitably large, we obtain from (3.21) and (3.22) that

\[(3.23) \quad \int_\Omega |\nabla \omega|^2 + \lambda \omega^2 = O\left(\sum_{i=1}^k \frac{\lambda}{\mu_i^2}\right) ||\omega||_2^{\delta} + O\left(\sum_{i=1}^k \frac{1}{\mu_i^{N/2}}\right) ||\omega||,\]

which gives

\[(3.24) \quad \int_\Omega |\nabla \omega|^2 + \lambda \omega^2 = O\left(\sum_{i=1}^k \left(\frac{\lambda}{\mu_i^2}\right)^{\frac{2}{2-\sigma}}\right) + O\left(\sum_{i=1}^k \frac{1}{\mu_i^{N/2}}\right).\]
On the other hand we obtain from (3.6), for each \( j = 1, \ldots, k \),

\[
\int_{\Omega} \nabla \left( \sum_{i=1}^{k} \alpha_i U_{\mu_i, y^i} + \omega \right) \nabla U_{\mu_j, y^j} + \lambda \left( \sum_{i=1}^{k} \alpha_i U_{\mu_i, y^i} + \omega \right) U_{\mu_j, y^j} = \int_{\Omega} \left| \nabla U_{\mu_j, y^j} \right|^2 + O \left( \sum_{i=1}^{k} \int_{\Omega} U_{\mu_i, y^i} U_{\mu_j, y^j} + \sum_{i=1}^{k} \int_{\Omega} U_{\mu_i, y^i} \left| \omega \right| \right) + O \left( \sum_{i=1}^{k} \int_{\Omega} U_{\mu_i, y^i} \left| \omega \right| \right)
\]

for some \( \theta \in (0, 2^* - 2) \).

Using estimates in Bahri [5] and (3.10), (3.11), we obtain

\[
\alpha_j^{2^* - 1} \int_{\Omega} U_{\mu_j, y^j}^{2^*} - \alpha_j \int_{\Omega} \left| \nabla U_{\mu_j, y^j} \right|^2 = \lambda \int_{\Omega} U_{\mu_j, y^j} U_{\mu_j, y^j} + \lambda \sum_{i=1}^{k} \int_{\Omega} U_{\mu_i, y^i} \left| \omega \right| + O \left( \sum_{i=1}^{k} \frac{1}{\mu_i^2} \right) \left| \omega \right| + O \left( \sum_{i=1}^{k} \frac{1}{\mu_i^{N/2}} \right) \left( \frac{\lambda}{\mu_i^2} \right)^{\frac{N}{2}} + O \left( \sum_{i=1}^{k} \frac{1}{\mu_i^{N-2}} \right)
\]

Notice that, from the definition of \( U_{\mu,y} \), there exist two positive constants
\( D_1, D_2 \), depending on \( N \) only, such that

\[
\frac{D_1}{\mu^2} \leq \int_{\Omega} U_{\mu_i, y}^2 \leq \frac{D_2}{\mu^2}
\]
(3.28) \[ \int_{\Omega} U_{\mu_1,y_1} U_{\mu_2,y_2} = O\left( \frac{1}{\mu_1^{\frac{N}{2}} \mu_2^{\frac{N}{2}}} \right) \] if \( y_1 \neq y_2 \),

and from (3.26) we then have

(3.29) \[ \alpha_j - 1 = \frac{\bar{D}_j \lambda}{\mu_j^2} + O\left( \sum_{i=1}^{k} \frac{1}{\mu_i^{N-2}} \right) \] for \( j = 1, \ldots, k \),

where \( \bar{D}_j \in R^1 \) satisfies \( \frac{1}{2} D_1 \leq \bar{D}_j \leq 2D_2 \). Therefore \( \frac{\lambda}{\mu_j^2} \rightarrow 0 \) as \( \lambda \rightarrow \infty \), since \( \alpha_j \rightarrow 1 \), \( \mu_i \rightarrow \infty \) as \( \lambda \rightarrow \infty \).

Next we establish a contradiction. Let us consider the left-hand side of (3.9), which we denote by \( L_j \), for \( j = 1, \ldots, k \). We have

(3.30) \[ L_j = \int_{\Omega} \nabla \left( \sum_{i=1}^{k} \alpha_i U_{\mu_i,y_i} + \omega \right) \nabla \frac{\partial U_{\mu_j,y_j}}{\partial \mu_j} \]

\[ + \lambda \left( \sum_{i=1}^{k} \alpha_i U_{\mu_i,y_i} + \omega \right) \frac{\partial U_{\mu_j,y_j}}{\partial \mu_j} \]

\[ - \int_{\Omega} \left| \sum_{i=1}^{k} \alpha_i U_{\mu_i,y_i} + \omega \right|^{2^* - 2} \left( \sum_{i=1}^{k} \alpha_i U_{\mu_i,y_i} + \omega \right) \frac{\partial U_{\mu_j,y_j}}{\partial \mu_j} \]

\[ = \sum_{i=1}^{k} \alpha_i \int_{\Omega} \nabla U_{\mu_i,y_i} \nabla \frac{\partial U_{\mu_j,y_j}}{\partial \mu_j} + \lambda \sum_{i=1}^{k} \alpha_i \int_{\Omega} U_{\mu_i,y_i} \frac{\partial U_{\mu_j,y_j}}{\partial \mu_j} \]

\[ - \alpha_j^{2^* - 1} \int_{\Omega} U_{\mu_j,y_i}^{2^* - 1} \frac{\partial U_{\mu_j,y_j}}{\partial \mu_j} \]

\[ - (2^* - 1) \alpha_j^{2^* - 2} \sum_{i=1}^{k} \alpha_i \int_{\Omega} U_{\mu_j,y_j}^{2^* - 2} \frac{\partial U_{\mu_j,y_j}}{\partial \mu_j} U_{\mu_i,y_i} \]

\[ - (2^* - 1) \alpha_j^{2^* - 2} \int_{\Omega} U_{\mu_j,y_j}^{2^* - 2} \frac{\partial U_{\mu_j,y_j}}{\partial \mu_j} \omega \]

\[ + O\left( \frac{1}{\mu_j} \right) \| \omega \|_{2^*}^{1+\theta} + O\left( \sum_{i=1}^{k} \int_{\Omega} U_{\mu_i,y_i}^{2^* - 1 - \theta} \right) \frac{1}{\mu_j}, \]

for some \( \theta \in (0, 2^* - 2) \).
Using (3.12), (3.14), the estimate in (3.24) for $\omega$, and the estimates in Bahri [5], we obtain

\begin{align}
(3.31) \quad L_j &= -\frac{\lambda}{\mu_j^3} D_3 + O \left( \frac{\lambda}{\mu_j^2} \sum_{i=1}^{k} \frac{1}{\mu_i^{\frac{N}{2}}} \right) + O \left( \frac{\lambda}{\mu_j^2} \right)^{\frac{2}{2-\sigma}} \frac{1}{\mu_j'}
\end{align}

\[+ O \left( \frac{1}{\mu_j^{\frac{2}{2+1}}} \right) ||\omega|| + O \left( \frac{1}{\mu_j} \right) ||\omega||^{2^* - 1}_2 \]

\[+ O \left( \sum_{i=1}^{k} \frac{1}{\mu_i^{\frac{N}{2}}} \right) \]

\[= -\frac{\lambda}{\mu_j^3} D_3 + O \left( \frac{\lambda}{\mu_j^2} \sum_{i=1}^{k} \frac{1}{\mu_i^{\frac{N}{2}}} \right) + O \left( \frac{\lambda}{\mu_j^2} \right)^{\frac{2}{2-\sigma}} \frac{1}{\mu_j'}
\]

\[+ O \left( \frac{1}{\mu_j^{\frac{2}{2+1}}} \right) + O \left( \frac{1}{\mu_j} \sum_{i=1}^{k} \left( \frac{\lambda}{\mu_i^2} \right)^{(2^* - 1)} \frac{2}{2-\sigma} \right), \]

where $D_3$ is a positive constant.

To estimate the right-hand side of (3.9) for $u_\lambda = \sum_{i=1}^{k} \alpha_i U_{\mu_i, y_i} + \omega$, we need first to obtain estimates of $b^j, c^j_\ell$ for $i = 1, \ldots, k, \ell = 1, \ldots, N$. To this end, we take $\phi = U_{\mu_j, y_j}$, $\frac{\partial U_{\mu_j, y_j}}{\partial \mu_j}$, and $\frac{\partial U_{\mu_j, y_j}}{\partial y_s}$ for $j = 1, \ldots, k, s = 1, \ldots, N$, respectively, in (3.7). We obtain a system of linear equations satisfied by $a^i, b^i, c^j_\ell, i = 1, \ldots, k, \ell = 1, \ldots, N$, as follows:

\begin{align}
(3.32) \quad \sum_{i=1}^{k} a^i \int_{\Omega} \nabla U_{\mu_i, y_i} \cdot \nabla U_{y_j, y_j} + \sum_{i=1}^{k} b^i \int_{\Omega} \frac{\partial U_{\mu_i, y_i}}{\partial \mu_i} \nabla U_{y_j, y_j}
\end{align}

\[+ \sum_{i=1}^{k} \sum_{\ell=1}^{N} c^i_\ell \int_{\Omega} \nabla \frac{\partial U_{\mu_i, y_i}}{\partial y_{\ell}} \nabla U_{y_j, y_j}
\]

\[\partial I_\lambda \left( \sum_{i=1}^{k} \alpha_i U_{\mu_i, y_i} + \omega \right) \left( U_{\mu_j, y_j} \right) \]

\[= \frac{\partial I_\lambda \left( \sum_{i=1}^{k} \alpha_i U_{\mu_i, y_i} + \omega \right)}{\partial \omega} = 0, \quad j = 1, \ldots, k,
\]
\[
\sum_{i=1}^{k} a_i \int_{\Omega} \nabla U_{\mu_i, y^i} \nabla \frac{\partial U_{\mu_j, y^j}}{\partial \mu_j} + \sum_{i=1}^{k} b_i \int_{\Omega} \nabla U_{\mu_i, y^i} \nabla \frac{\partial U_{\mu_j, y^j}}{\partial \mu_j} = \sum_{i=1}^{k} \sum_{\ell=1}^{N} c_{i,\ell} \int_{\omega} \nabla U_{\mu_i, y^i} \nabla \frac{\partial U_{\mu_j, y^j}}{\partial y^j_{\ell}} \\
+ \sum_{i=1}^{k} \sum_{\ell=1}^{N} a_{i,\ell} \int_{\Omega} \nabla U_{\mu_i, y^i} \nabla \frac{\partial U_{\mu_j, y^j}}{\partial \mu_j} = \frac{1}{\alpha_j} L_{j,\ell}, \quad \text{for } j = 1, \ldots, k,
\]

(3.34)
\[-(2^* - 1)\alpha_j^{2^*-2} \sum_{i=1 \atop i \neq j}^k \int_\Omega U_{\mu_j, y^j}^{2^*-2} \frac{\partial U_{\mu_j, y^j}}{\partial y_i^j} U_{\mu_j, y^j} \]

\[-(2^* - 1)\alpha_j^{2^*-2} \int_\Omega U_{\mu_j, y^j} \frac{\partial U_{\mu_j, y^j}}{\partial y_j^j} \omega \]

\[O(\mu_j) ||\omega||^{1+\theta} + O \left( \sum_{i=1 \atop i \neq j}^k \int_\Omega U_{\mu_j, y^j}^{2^*-1-\theta} U_{\mu_i, y^i}^{1+\theta} \right) \mu_j \]

(for some $\theta \in (0, 2^* - 2)$)

\[= O \left( \frac{\lambda}{\mu_j} \right) + O \left( \frac{\lambda}{y_j^j} \right) \sum_{i=1 \atop i \neq j}^k \frac{1}{\mu_i^{N-2}} + O \left( \frac{\lambda}{\mu_j^2} \right)^{\frac{2}{2^*-\theta}} \mu_j \]

\[+ O \left( \frac{1}{\mu_j} \right) + O(\mu_j) \sum_{i=1}^k \left( \frac{\lambda}{\mu_j^2} \right)^{\frac{2}{2^*-\theta}(2^*-1)} \mu_j = 1, \ldots, k; \quad s = 1, \ldots, N. \]

From (3.32)-(3.34), and the estimates in Bahri [5], we deduce

\[b^j = O \left( \frac{\lambda}{\mu_j} \right) + O \left( \sum_{i=1}^k \frac{1}{\mu_i^{N-2}} \right) \frac{\lambda}{\mu_j^{N-2}} \]

\[+ O \left( \frac{1}{\mu_j} \right) + o \left( \sum_{i=1}^k \frac{\lambda}{\mu_i^{N-2}} \right), \quad \text{for } j = 1, \ldots, k \]

\[c^j = O \left( \frac{\lambda}{\mu_j} \right) + O \left( \sum_{i=1}^k \frac{1}{\mu_i^{N-2}} \right) \frac{\lambda}{\mu_j^{N-2}} \]

\[+ o \left( \sum_{i=1}^k \frac{\lambda}{\mu_i^{N-2}} \right) \frac{1}{\mu_j} + O \left( \frac{1}{\mu_j^2} \right) \]

for $j = 1, \ldots, k, \ell = 1, \ldots, N.$
Substituting (3.35), (3.36) into (3.9), and using (3.31), we obtain

\[
-\frac{\lambda}{\mu_j^3} D_3 = O \left( \frac{\lambda}{\mu_j^{5/2}} \right) \sum_{i=1}^{k} \frac{1}{\mu_i^{3/2}} + O \left( \frac{\lambda}{\mu_j^2} \right)^{\frac{2}{2-\sigma}} \frac{1}{\mu_j} + O \left( \frac{1}{\mu_j^3} \right)
\]

\[+ O \left( \frac{1}{\mu_j} \right) \sum_{i=1}^{k} \left( \frac{\lambda}{\mu_j^2} \right)^{(2^*-1)\frac{2}{2-\sigma}}
\]

\[= O \left( \frac{\lambda}{\mu_j^{5/2}} \right) \sum_{i=1}^{k} \frac{1}{\mu_i^{3/2}} + o \left( \frac{\lambda}{\mu_j^3} \right), \quad j = 1, \cdots, k,
\]

where \( D_3 \) is a positive constant.

For \( N \geq 6 \) (3.37) is clearly impossible. For \( N = 5 \), notice that \( \frac{\lambda}{\mu_j^3} \to 0 \) as \( \lambda \to \infty \). Applying the Hölder inequality to the term \( \frac{\lambda}{\mu_j^{5/2}} \frac{1}{\mu_i^{3/2}} \), we then obtain

\[
-\frac{\lambda}{\mu_j^3} D_3 = \sum_{i=1}^{k} o \left( \frac{\lambda}{\mu_j^3} \right), \quad j = 1, \cdots, k.
\]

A contradiction. This completes the proof of Theorem 1.2.

Acknowledgments. After this work was completed, C. Gui told us that he, cooperated with others, has obtained a nonexistence result for \( k = 1 \) (one peaked solutions) when \( N = 3 \) similar to Theorem 1.2 by using a different method.

References


Received September 27, 1999 and revised February 28, 2000. Part of this work was done while Daomin Cao was visiting the School of Mathematics, University of New South Wales as a visiting professor. The first author was supported by 973 project of the Ministry of Sciences and Technology, NSFC and innovation funds of CAS in China. The second and third authors were supported by ARC in Australia.

INSTITUTE OF APPLIED MATHEMATICS
ACADEMY OF MATHEMATICS AND SYSTEM SCIENCES
CHINESE ACADEMY OF SCIENCES
BEIJING 100080
P.R.CHINA
E-mail address: cao@amath6.amt.ac.cn

SCHOOL OF MATHEMATICS
UNIVERSITY OF NEW SOUTH WALES
SYDNEY NSW 2052
AUSTRALIA
E-mail address: noussair@maths.unsw.edu.au

School of Mathematics and Statistics
University of Sydney
NSW 2006
Australia
E-mail address: shusen@maths.usyd.edu.au
ALEXANDER AND THURSTON NORMS OF FIBERED 3-MANIFOLDS

Nathan M. Dunfield

For a 3-manifold $M$, McMullen derived from the Alexander polynomial of $M$ a norm on $H^1(M, \mathbb{R})$ called the Alexander norm. He showed that the Thurston norm is an upper bound for the Alexander norm. He asked if these two norms were the same when $M$ fibers over the circle. Here, I give examples that show this is not the case. This question relates to the faithfulness of the Gassner representations of the braid groups. The key tool used is the Bieri-Neumann-Strebel invariant, and I show a connection between this invariant and the Alexander polynomial.

1. Introduction.

1.1. Statement of results. For a 3-manifold $M$, McMullen derived from the Alexander polynomial of $M$ a norm on $H^1(M, \mathbb{R})$ called the Alexander norm. He showed that the Thurston norm on $H^1(M, \mathbb{R})$, which measures the complexity of a dual surface, is an upper bound for the Alexander norm. He asked (Question A below) if these two norms were equal on all of $H^1(M, \mathbb{R})$ when $M$ fibers over the circle. Here, I will give examples which show that the answer to Question A is emphatically no. As explained below, Question A is related to the faithfulness of the Gassner representations of the braid groups. The key tool used to understand Question A is the Bieri-Neumann-Strebel invariant from combinatorial group theory. Theorem 1.7 below, which is of independent interest, connects the Alexander polynomial with a certain Bieri-Neumann-Strebel invariant.

I will begin by reviewing the definitions of the Alexander and Thurston norms, and Theorem 1.2 which relates them. Then I'll discuss Question A and the connection to the braid groups. After that, I'll state Question B, a much weaker version of Question A, to which the answer is also no. A brief description of the examples which answer these two questions concludes Section 1.1. In Section 1.5, I'll connect these questions with the Bieri-Neumann-Strebel invariants, and explain why, morally speaking, the answer to both questions must be no. Section 1.8 outlines the rest of the paper.

The Alexander norm is defined in [McM] as follows. Let $M$ be a 3-manifold (all 3-manifolds in this paper will be assumed to be connected).
Let $G$ be the fundamental group of $M$. Let $\text{ab}(G)$ denote the maximal free abelian quotient of $G$, which is isomorphic to $\mathbb{Z}^{b_1(M)}$ where $b_1(M)$ is the first Betti number of $M$. The Alexander polynomial $\Delta_M$ of $M$ is an element of the group ring $\mathbb{Z}[\text{ab}(G)]$. It is an invariant of the homology of the cover of $M$ with covering group $\text{ab}(G)$ (for details see Section 3.1). The Alexander norm on $H^1(M, \mathbb{R})$ is the norm dual to the Newton polytope of $\Delta_M$. That is, if $\Delta_M = \sum_{i=1}^{n} a_i g_i$ with $a_i \in \mathbb{Z}\{0\}$ and $g_i \in \text{ab}(G)$ then the norm of a class $\phi \in H^1(M, \mathbb{R})$ is defined to be

$$\|\phi\|_A = \sup_{i,j} \phi(g_i - g_j).$$

The unit ball $B_A$ of this norm is, up to scaling, the polytope dual to the Newton polytope of $\Delta_M$.

The Thurston norm is defined as follows. For a compact connected surface $S$, let $\chi_-(S) = |\chi(S)|$ if $\chi(S) \leq 0$ and 0 otherwise. For a surface with multiple connected components $S_1, S_2, \ldots, S_n$, let $\chi_-(S)$ be sum of the $\chi_-(S_i)$. Then the Thurston norm of an integer class $\phi \in H^1(M, \mathbb{Z}) \cong H_2(M, \partial M; \mathbb{Z})$ is

$$\|\phi\|_T = \inf \{ \chi_-(S) \mid S \text{ is a properly embedded oriented surface} \text{ that is dual to } \phi \}.$$

As described in [Thu], this norm extends continuously to all of $H^1(M, \mathbb{R})$. The unit ball $B_T$ in this norm is a finite-sided convex polytope.

It should be noted that both of these “norms” are really semi-norms—they can be zero on non-zero vectors of $H^1(M, \mathbb{R})$.

McMullen proved the following theorem which connects the two norms; here $b_i(M) = \text{rank} H_i(M, \mathbb{R})$ denotes the $i^{th}$ Betti number of $M$.

**Theorem 1.2 ([McM]).** Let $M$ be a compact, orientable 3-manifold whose boundary, if any, is a union of tori. Then for all $\phi$ in $H^1(M, \mathbb{R})$, the Alexander and Thurston norms satisfy

$$\|\phi\|_A \leq \|\phi\|_T \text{ if } b_1(M) \geq 2,$$

or

$$\|\phi\|_A \leq \|\phi\|_T + 1 + b_3(M) \text{ if } b_1(M) = 1 \text{ and } \phi \text{ generates } H^1(M, \mathbb{Z}).$$

Moreover, equality holds when $\phi: \pi_1(M) \to \mathbb{Z}$ and $\phi$ can be represented by a fibration $M \to S^1$, where the fibers have non-positive Euler characteristic.

This theorem generalizes the fact that the degree of the Alexander polynomial of a knot is bounded by twice the genus of any Seifert surface. In many simple cases, e.g. almost all the exteriors of the links with 9 or fewer crossings, the Alexander and Thurston norms agree on all of $H^1(M, \mathbb{R})$ (see [McM]). In such cases, this theorem explains D. Fried’s observation from
the 80’s that frequently the shape of the Newton polytope of the Alexander
polynomial is dual to that of the Thurston norm ball.

Before stating Question A, I need to discuss the relationship between
the Thurston norm and cohomology classes \( \phi: \pi_1(M) \to \mathbb{Z} \) which can be
represented by fibrations \( M \to S^1 \). There are top-dimensional faces, called
the fibered faces, of \( B_T \) such that a class \( \phi \in H^1(M, \mathbb{Z}) \) can be represented
by a fibration over the circle if and only if \( \phi \) lies in the cone over the interior
of one of the fibered faces [Thu, §3]. In this context, the last sentence of
Theorem 1.2 is equivalent to “Moreover, the two norms agree on classes that
lie in the cone over the fibered faces of \( B_T \).” The point of this paper is to
answer:

**Question A** (McMullen [McM]). Let \( M \) be a compact, orientable 3-manif-
fold whose boundary, if any, is a union of tori. Suppose that \( M \) fibers over
the circle and that \( b_1(M) \geq 2 \). Do the Alexander and Thurston norms agree
on all of \( H^1(M, \mathbb{R}) \)?

My motivation for studying this question is McMullen’s result that a
yes answer would imply that the Gassner representations of the pure braid
groups are all faithful [McM, §8]. This would answer in the affirmative
the important question: Are the braid groups linear, that is, do they have
faithful, finite-dimensional, linear representations? Sadly, I will show that
the answer to Question A is no in a strong sense. (Note: Since I wrote this
paper, Bigelow and Krammer have independently shown that braid groups
are linear [Big2, Kra1, Kra2]. Their proofs use a different representation,
and it remains unknown whether the Gassner representation is faithful.)

To explain why the answer to Question A is no, let me formulate a weaker
version of Question A which will help make clear some of the issues involved.
Henceforth, I will assume that \( b_1(M) \geq 2 \). A typical example of \( B_T \) is given
in Figure 1.3.

There is a pair of fibered faces and the rest of the faces are not fibered.
Theorem 1.2 tells us that \( \| \cdot \|_A \leq \| \cdot \|_T \) hence that \( B_A \supseteq B_T \). Since the
two norms agree on a fibered face \( F_T \) of \( B_T \), there is a face \( F_A \) of \( B_A \) which
contains \( F_T \). Now, it seems a bit much to expect that if \( M \) fibers over the
circle then the two norms agree on classes that are far from any fibered face.
So it’s reasonable to consider:

**Question B.** Let \( M \) be a compact, orientable 3-manifold whose boundary,
if any, is a union of tori. Suppose that \( M \) fibers over the circle and that
\( b_1(M) \geq 2 \). Let \( F_T \) be a fibered face of \( B_T \) and \( F_A \) the face of \( B_A \) which
contains it. Are \( F_T \) and \( F_A \) always equal?

Figure 1.4 shows the two possibilities. Note that a yes answer to Quest-
ion A implies a yes answer to Question B. I will give examples which show that
Answer. The answer to Question B, and therefore Question A, is no.

I will give two kinds of examples. In Section 2, I will construct examples using the fact that the Burau representation of the braid group on 5 strands is not faithful. Section 2 is independent of the rest of the paper. Section 6 contains an example which is the exterior of a specific 17 crossing link in \( S^3 \).

McMullen’s formulation of Question A restricted attention to those manifolds which are the exteriors of links in \( S^3 \). All my examples are such manifolds, but I felt the more general statement was appropriate here.

1.5. Connection to the BNS invariants. In this section I will describe the connection between Question B and the Bieri-Neumann-Strebel (BNS) invariants. In light of this connection, I will explain why the answer to Question B must be, morally speaking, no. The BNS invariants will also be used in constructing and verifying the example in Section 6.

I’ll begin with the definition of the BNS invariants (for details see [BNS], and from a different point of view, [Bro]). Let \( G \) be a finitely-generated group. Set

\[
S(G) = \left( H^1(G, \mathbb{R}) \setminus \{0\} \right) / \mathbb{R}^+,
\]

where \( \mathbb{R}^+ \) acts by scalar multiplication and \( S(G) \) is given the quotient topology. A point \( [\chi] \) in \( S(G) \) will be thought of as an equivalence class of homomorphisms \( \chi: G \to \mathbb{R} \). For \( [\chi] \in S(G) \) define \( G_\chi = \chi^{-1}([0, \infty)) = \{ g \in G \mid \chi(g) \geq 0 \} \), which is a sub-monoid of \( G \).
Let $H$ be a group acted on by $G$ where $G'$ acts by inner automorphisms (e.g. $H = G'$ where $G$ acts by conjugation). Then the BNS invariant of $G$ and $H$ is:

$$\Sigma_H = \{ [\chi] \in S(G) \mid H \text{ is finitely generated over some finitely generated sub-monoid of } G \chi \}.$$ 

It turns out that $\Sigma_H$ is always an open subset of the sphere $S(G)$.

Let $M$ be a 3-manifold, and $G = \pi_1(M)$. Set $\Sigma = \Sigma_{G'}$. Bieri, Neumann, and Strebel proved the following with the help of Stallings' fibration theorem:

**Theorem 1.6** ([BNS, Thm. E]). Let $M$ be a compact, orientable, irreducible 3-manifold. Then $\Sigma$ is exactly the projection to $S(G)$ of the interiors of the fibered faces of the Thurston norm ball $B_T$.

For convenience, in the rest of this section I will assume that $H_1(M, \mathbb{Z})$ is free. This is not essential, and the theory will be developed without this assumption in Sections 3-5. The commutator subgroup $G'$ is the fundamental group of the universal abelian cover of $M$. So $A = G'/G''$ is the first homology of that cover. Thought of as a module over $\mathbb{Z}[\text{ab}(G)]$, $A$ is the Alexander invariant of $M$, from which the Alexander polynomial is derived. Thus it is not too surprising that the BNS invariant $\Sigma_A$ is connected to the Alexander polynomial:

**Theorem 1.7.** Let $M$ be a compact, orientable 3-manifold. There are top-dimensional faces $F_1, F_2, \ldots, F_n$ of the Alexander ball $B_A$ such that the projection of the interiors of the $F_i$ into $S(G)$ is exactly $\Sigma_A$. Moreover, the $F_i$ are completely determined by the Alexander polynomial of $M$.

Theorem 5.1 below is an expanded version of Theorem 1.7 which explains how the $F_i$ are determined. Now since $A$ is a quotient of $G'$, it follows immediately from the definitions that $\Sigma_A \supset \Sigma$. Combining this with Theorem 1.7, it follows that Question B is equivalent to:

**Question B'.** Let $M$ be a compact, orientable 3-manifold whose boundary, if any, is a union of tori. Suppose that $M$ fibers over the circle and that $b_1(M) \geq 2$. Let $C$ be a connected component of $\Sigma$. If $D$ is the connected component of $\Sigma_A$ which contains $C$, is $D$ always equal to $C$?

Put this way it begins to become clear that the answer to Question B should be no. For many groups $G$, $\Sigma_{G'}$ is strictly contained in $\Sigma_{G'/G''}$. It remains only to produce examples of 3-manifolds whose fundamental groups have this property.

**1.8. Outline of rest of paper.** Section 2 describes how to construct examples using the Burau representation. Section 3 defines the Alexander polynomial and proves a fact about the Alexander invariant that’s needed
to prove Theorem 1.7. Section 4 discusses the BNS invariants and records the properties that will be needed later. Section 5 proves the full version of Theorem 1.7. Finally, Section 6 gives an example of a specific link exterior in $S^3$ for which the answer to Question B is no.

**Acknowledgments.** I wrote this paper while a graduate student at the University of Chicago supported by a Sloan Dissertation Fellowship. I would like to thank Curt McMullen for useful correspondence. I got interested in the connection between the Alexander polynomial and Thurston norm at a problem session at KirbyFest (MSRI, June 1998), where Fried’s observation was related by Joe Christy. I would like to thank the organizers, MSRI, and the NSF for support to attend that conference. I would also like to thank the referee for help clarifying the exposition.

2. Connection with braid groups.

Let $B_n$ denote the $n$-strand braid group. McMullen showed that if the answer to Question A is yes, then the Gassner representation of $B_n$ is faithful \[ \text{[McM]} \]. In this section, I’ll give a very similar argument to show:

**Proposition 2.1.** If the answer to Question B is yes, then the Burau representation of $B_n$ is faithful for all $n$.

Since the Burau representation of $B_n$ is not faithful for $n \geq 5$ \[ \text{[Big1, LP, Moo]} \], the proposition implies that the answer to Question B, and hence Question A, is no.

Before proving the proposition, let me define the braid groups and the Burau representation (see \[ \text{[Bir]} \] for more). Let $D_n$ be the disc with $n$ punctures. Consider the group of homeomorphisms $\text{Hom}^+(D_n, \partial D_n)$ of $D_n$ which are orientation preserving and fix $\partial D_n$ pointwise. The braid group $B_n$ is $\text{Hom}^+(D_n, \partial D_n)$ modulo isotopies which pointwise fix $\partial D_n$.

To define the Burau representation, consider the homomorphism

$$\phi: H_1(D_n) \to \mathbb{Z} = \langle t \rangle$$

which takes any clockwise oriented loop about a single puncture to $t$. Let $\widetilde{D_n}$ be the cover of $D_n$ corresponding to $\phi$. The homology of $\widetilde{D_n}$ is a module over the group ring $\mathbb{Z}[[t]]$ of the group of covering transformations. The module $H_1(\widetilde{D_n}, \mathbb{Z})$ is free of rank $n - 1$. The Burau representation is a homomorphism $\text{Burau}: B_n \to \text{Aut}(H_1(\widetilde{D_n}))$. By $\text{Aut}(H_1(\widetilde{D_n}))$, I mean automorphisms of $H_1(\widetilde{D_n})$ as a $\mathbb{Z}[[t]]$-module. Choosing a $\mathbb{Z}[[t]]$ basis of $H_1(\widetilde{D_n})$ allows one to view the Burau representation as having image in $\text{GL}(n - 1, \mathbb{Z}[[t]])$. Given $\beta$ in $B_n$, $\text{Burau}(\beta)$ is constructed as follows. Let $f: D_n \to D_n$ be a representative of $\beta$. Choose a lift $\widetilde{f}: \widetilde{D_n} \to \widetilde{D_n}$ of $f$. Since the action of $f$ on $H_1(D_n)$ commutes with $\phi$, the lift $\widetilde{f}$ is equivariant. Thus
there is a unique lift of \( f \) which leaves the inverse image of \( \partial D_n \) pointwise fixed. Let \( \tilde{f} \) be that lift and set \( \text{Burau}(\beta) = \tilde{f}_*: H_1(D_n) \to H_1(D_n) \).

I'll need the following property of the Burau representation (see also \([\text{Mor}])\). Suppose \( \beta \) is a braid whose action on the set of punctures is an \( n \)-cycle. Let \( \tilde{M}_\beta \) be the 3-manifold which is the mapping torus of \( \beta \). The manifold \( \tilde{M}_\beta \) has two boundary components, and \( H_1(\tilde{M}_\beta) = \mathbb{Z} \oplus \mathbb{Z} \). Take as a basis of \( H_1(\tilde{M}_\beta) \) the pair \((t', w)\) where \( t' \) is a counter-clockwise loop about a puncture in \( D_n \) and \( w \) is a point in \( \partial D_n \) cross \( S^1 \). It's not hard to see that the universal abelian cover of \( M_\beta \) is \( \tilde{D}_n \times \mathbb{R} \). The covering transformation corresponding to \( t' \) is \((\tilde{d}, r) \mapsto (t(\tilde{d}), r)\), and the covering transformation corresponding to \( w \) is \((\tilde{d}, r) \mapsto (\tilde{f}(\tilde{d}), r + 1)\). If we replace \( t \) by \( t' \) in \( \text{Burau}(\beta) \), the matrix \((wI - \text{Burau}(\beta))\) is a presentation matrix for the homology of the universal abelian cover of \( M_\beta \) as a \( \mathbb{Z}[H_1(M_\beta)] \)-module. Thus

\[ \Delta_{M_\beta} = \det (wI - \text{Burau}(\beta)). \]

I will now prove the proposition.

**Proof of Proposition 2.1.** Suppose the answer to Question B is yes and the Burau representation of \( B_n \) has kernel for some \( n \). As the Burau representation is known to be faithful for \( n = 2 \), assume \( n \) is at least 3. Then there is a pseudo-Anosov element \( \delta \) in the kernel \([\text{Lon}, \text{Iva}]\). Replacing \( \delta \) with a power of \( \delta \) if necessary, we can assume \( \delta \) is a pure braid, that is, fixes each puncture. Let \( \gamma \) be the braid \( \sigma_1 \sigma_2 \ldots \sigma_{n-1} \) where \( \sigma_i \) is the \( i^{th} \) standard generator of \( B_n \) (see Figure 2.2).

![Figure 2.2.](image)

**Figure 2.2.** The braid \( \gamma \) when \( n = 5 \).

Taking a power of \( \delta \) if necessary, we can assume that \( \beta = \delta \gamma \) is pseudo-Anosov. Now \( \beta \) induces an \( n \)-cycle on the punctures because \( \delta \) was a pure braid and \( \gamma \) induces an \( n \)-cycle. Since \( \text{Burau}(\beta) = \text{Burau}(\gamma) \), the Alexander polynomials of \( M_\beta \) and \( M_\gamma \) are the same. The manifold \( M_\gamma \) is Seifert fibered,
and it’s easy to see that the Thurston norm ball is as shown in Figure 2.3, where the two infinite faces are fibered faces. Thus by Theorem 1.2, the Alexander norm ball of $M_\gamma$ has exactly the same shape as the Thurston norm ball. Since $M_\gamma$ and $M_\beta$ have the same Alexander polynomials, the Alexander norm ball of $M_\beta$ is as shown. But $M_\beta$ is hyperbolic, and hence the Thurston norm is non-degenerate. So any face of the Thurston norm ball is bounded. Thus a fibered face of the Thurston norm ball of $M_\beta$ is properly contained in the corresponding face of the Alexander norm ball. This contradicts the assumption that the answer to Question B is yes. □


3.1. Definitions. I’ll begin by reviewing the definition of the Alexander polynomial and related invariants (for more see [Hil, Rol, McM]). Let $X$ be a finite CW-complex with fundamental group $G$. Let $\widetilde{X}$ be the universal free abelian cover of $X$, that is, the cover induced by the homomorphism from $G$ to its free abelianization $\text{ab}(G)$. Let $p$ be a point of $X$, and $\tilde{p}$ its inverse image in $\widetilde{X}$. The Alexander module of $X$ is

$$A_X = H_1(\widetilde{X}, \tilde{p}; \mathbb{Z})$$

thought of as a module over the group ring $\mathbb{Z}[\text{ab}(G)]$. The reason one uses the free abelianization is so that the ring $\mathbb{Z}[\text{ab}(G)]$ has no zero divisors.

For a finitely generated module $M$ over $\mathbb{Z}[\text{ab}(G)]$, the $i$th elementary ideal $E_i(M) \subset \mathbb{Z}[\text{ab}(G)]$ is defined as follows. Take any presentation

$$0 \to (\mathbb{Z}[\text{ab}(G)])^r \xrightarrow{P} (\mathbb{Z}[\text{ab}(G)])^s \to M \to 0$$

and set $E_i(M)$ to be the ideal generated by the $(s - i, s - i)$ minors of the matrix $P$. The Alexander ideal of $X$ is $E_1(A_X)$. The Alexander polynomial of $X$, denoted $\Delta_X$, is the greatest common divisor of the elements of the Alexander ideal. The polynomial $\Delta_X$ is defined up to multiplication by a unit $g \in \text{ab}(G)$ of $\mathbb{Z}[\text{ab}(G)]$. Equivalently, $\Delta_X$ is a generator of the smallest principle ideal containing the Alexander ideal.

I should mention that the Alexander module, and hence Alexander polynomial, depends only on the fundamental group of $X$; it can be thought of as an invariant of a finitely generated group.

I will need to consider $B_X = H_1(\widetilde{X}; \mathbb{Z})$, the Alexander invariant of $X$. When $H_1(X; \mathbb{Z})$ is free, $B_X = G'/G''$. As with $A_X$, the Alexander invariant $B_X$ is to be thought of as a module over $\mathbb{Z}[\text{ab}(G)]$. The two modules are related as follows. Let $m \subset \mathbb{Z}[\text{ab}(G)]$ be the augmentation ideal, that is $m = \langle 1 - g \mid g \in \text{ab}(G) \rangle$. The homology long exact sequence for the pair $(\widetilde{X}, \tilde{p})$ gives rise to the short exact sequence

$$0 \to B_X \to A_X \to m \to 0.$$
The Alexander polynomial of $X$ could just have well been defined as the gcd of $E_0(B_X)$ (for the equivalence of these two definitions see, e.g. [Tra]).

### 3.2. Structure of the Alexander invariant of a 3-manifold.

The following fact about the structure of the Alexander ideal of a 3-manifold was crucial in McMullen’s proof of Theorem 1.2.

**Theorem 3.3** ([McM, 5.1]). Let $M$ be a compact, orientable 3-manifold whose boundary, if any, is a union of tori. Let $G = \pi_1(M)$. Then $E_1(A_M) = m^p \cdot (\Delta_M)$ where

$$p = \begin{cases} 0 & \text{if } b_1(M) \leq 1, \\ 1 + b_3(M) & \text{otherwise}, \end{cases}$$

and $m$ is the augmentation ideal of $\mathbb{Z}[\text{ab}(G)]$.

The corresponding fact about $E_0(B_M)$ will be key to the proof of Theorem 1.7. For a manifold with non-empty torus boundary, Crowell and Strauss [CS] showed that $E_0(B_M) = (\Delta_M) \cdot m^q$ for an explicit value of $q$. The following proposition is weaker than [CS], but it also applies to closed 3-manifolds. It will suffice for my purposes and follows easily from known results.

**Proposition 3.4.** Let $M$ be a compact, orientable 3-manifold whose boundary, if any, is a union of tori. Then

$$\sqrt{E_0(B_M)} \cap m = \sqrt{(\Delta_M)} \cap m.$$

**Proof.** By Theorem 1.1 of [Tra] the short exact sequence

$$0 \to B_M \to A_M \to m \to 0$$

implies that there are integers $r, s \geq 0$ such that

$$E_1(A_M) \cdot m^r \subset E_0(B_M) \quad \text{and} \quad E_0(B_M) \cdot m^s \subset E_1(A_M).$$

Combining and multiplying by $m$ gives

$$E_1(A_M) \cdot m^{r+s+1} \subset E_0(B_M) \cdot m^{s+1} \subset E_1(A_M) \cdot m.$$

Taking radicals of the above and using that $\sqrt{I \cdot J} = \sqrt{I} \cap \sqrt{J}$ gives

$$\sqrt{E_1(A_M)} \cap \sqrt{m} = \sqrt{E_0(B_M)} \cap \sqrt{m}.$$ 

Now $m$ is radical since it is the kernel of the ring homomorphism $\mathbb{Z}[\text{ab}(G)] \to \mathbb{Z}$ which sends every $g \in \text{ab}(G)$ to 1. By Theorem 3.3 we have $E_1(A_M) = (\Delta_M) \cdot m^p$. Combining, we get $\sqrt{E_0(B_M)} \cap m = \sqrt{(\Delta_M)} \cap m$ as desired. □

Recall the definition of the BNS invariant from Section 1.5. Let $G$ be a finitely-generated group. Let $S(G) = \{H^1(G, \mathbb{R}) \setminus \{0\} \}/\mathbb{R}^+$. For $[\chi] \in S(G)$ we have the sub-monoid $G_\chi = \{g \in G \mid \chi(g) \geq 0\}$. Let $H$ be a group acted on by $G$ where $G'$ acts by inner automorphisms. Then the BNS invariant of $G$ and $H$ is:

$$\Sigma_H = \{[\chi] \in S(G) \mid H \text{ is finitely generated over some finitely generated sub-monoid of } G_\chi\}.$$

We can also consider the larger invariant

$$\Sigma'_H = \{[\chi] \in S(G) \mid H \text{ is finitely generated over } G_\chi\}.$$

When $H$ is abelian $\Sigma'_H = \Sigma_H$ [BNS, Theorem 2.4]. The special case of $\Sigma'_H$ when both $G$ and $H$ are abelian was studied by Bieri and Strebel [BS] prior to the development of the full BNS invariant. The rest of this section will be devoted to that special case.

Let $Q$ be a finitely generated free abelian group and $A$ a finitely generated $\mathbb{Z}[Q]$-module. Since $A$ has an action of $Q$, we can form the BNS invariant $\Sigma_A = \Sigma'_A$. To reduce clutter, I'll denote $\mathbb{Z}[Q]$ by $\mathbb{Z}_Q$. A basic property shown in [BS, §1.3] is that $\Sigma_A = \Sigma_{\mathbb{Z}_Q}/\text{Ann}(A)$ where $\text{Ann}(A)$ is the annihilator ideal of $A$. Thus $\Sigma$ can be seen as an invariant of an ideal $I \subset \mathbb{Z}_Q$. The following basic identities hold for any ideals $I, J$ in $\mathbb{Z}_Q$ [BS, §1.3]:

$$\Sigma_{\mathbb{Z}_Q/I} = \Sigma_{\mathbb{Z}_Q/\sqrt{I}} \quad \text{and} \quad \Sigma_{\mathbb{Z}_Q/(I \cdot J)} = \Sigma_{\mathbb{Z}_Q/I} \cap \Sigma_{\mathbb{Z}_Q/J}.$$

For principle ideals $I$, the invariant $\Sigma_{\mathbb{Z}_Q/I}$ can be easily calculated, as I will now describe. Let $p \in \mathbb{Z}_Q$, the Newton polytope $\text{Newt}(p)$ is defined as follows. Consider the vector space $V = Q \otimes \mathbb{R}$ which contains $Q$ as a lattice. The Newton polytope of $p$ is the convex hull in $V$ of those $q \in Q$ which have non-zero coefficient in $p$. The vertices of $\text{Newt}(p)$ lie in $Q$, and I’ll define the coefficient of a vertex of $\text{Newt}(p)$ to be the non-zero coefficient of the corresponding term of $p$. Given a $q$ in $Q$, define the open hemisphere $H_q$ of $S(Q)$ to be

$$\{[\chi] \in S(Q) \mid \chi(q) > 0\}.$$

The following theorem allows us to calculate $\Sigma_{\mathbb{Z}_Q/I}$ for a principle ideal $I$.

**Theorem 4.1** ([BS, 5.2]). Let $Q$ be a finitely generated free abelian group and $p$ an element of $\mathbb{Z}_Q$. The connected components of $\Sigma_{\mathbb{Z}_Q/(p)}$ are in one-to-one correspondence with the vertices of $\text{Newt}(p)$ whose coefficients are $\pm 1$, where such a vertex $v$ corresponds to

$$C_v = \bigcap \{H_{vw^{-1}} \mid w \text{ is a vertex of } \text{Newt}(p) \text{ distinct from } v\}.$$
5. BNS invariants and Alexander polynomial of a 3-manifold.

Let $M$ be a compact, orientable 3-manifold whose boundary, if any, is a union of tori. Let $B_M = H_1(M, \mathbb{Z})$ be the Alexander invariant of $M$. Regarding $B_M$ as a $\mathbb{Z}[\text{ab}(\pi_1 M)]$ module, we can form the BNS-invariant $\Sigma_{B_M}$ which I will denote by $\Sigma_A$. In Section 1.5, I defined $\Sigma_A$ in case where $H_1(M, \mathbb{Z})$ is torsion free, and that definition was slightly different. In the torsion free case, $B_M = G'/G''$ where $G = \pi_1(M)$. Thus only difference between the two definitions is that one is the BNS invariant with respect to $\text{ab}(G)$ and the other $G$. Since $B_M$ is abelian and $G'$ acts trivially on it, the two definitions agree.

In this section I will prove Theorem 5.1 which computes $\Sigma_A$ from the Alexander polynomial $\Delta_M$. Before stating Theorem 5.1, I need to discuss the unit ball $B_A$ in the Alexander norm.

Consider the Newton polytope $\text{Newt}(\Delta_M)$ in $H_1(M, \mathbb{R})$. The Alexander norm on $H_1(M, \mathbb{R})$ can be defined as

$$\|\phi\|_A = \sup\{\phi(x - y) \mid x, y \in \text{Newt}(\Delta_M)\}.$$ 

A polytope $P$ is balanced about 0 if it is invariant under $v \mapsto -v$. More generally, $P$ is balanced about a point $p$ if the translate of $P$ by $-p$ is balanced about 0. Since $M$ is a 3-manifold, $\Delta_M$ is symmetric [Bla, Tur, 4.5], and hence $\text{Newt}(p)$ is balanced about some point $z_0$. Then

$$\|\phi\|_A = \sup\{2\phi(x - z_0) \mid x \in \text{Newt}(\Delta_M)\}$$

and the unit ball in $\|\cdot\|_A$ is

$$B_A = \{\phi \mid \phi(x - z_0) \leq 1/2 \text{ for all } x \in \text{Newt}(\Delta_M)\}.$$ 

Fix a basis of $H_1(M, \mathbb{R})$ and identify $H^1(M, \mathbb{R})$ with $H_1(M, \mathbb{R})$ via the dual basis. Then $B_A$ is, after scaling by a factor of 2, the classical polytope dual of $\text{Newt}(\Delta_M)$ about $z_0$.

Duality of polytopes in an $n$-dimensional vector space exchanges faces of dimension $i$ with faces of dimension $n - i - 1$ (for more on polytope duals, see [Bro]). A vertex $v$ of $\text{Newt}(\Delta_M)$ becomes the top-dimensional face

$$F_v = \{\phi \mid \phi(x - z_0) \leq 1/2 \text{ for all } x \in \text{Newt}(\Delta_M) \text{ and } \phi(v - z_0) = 1/2\}.$$

I can now state the theorem that relates $\Sigma_A$ and $B_A$.

**Theorem 5.1.** Let $M$ be a compact, orientable 3-manifold whose boundary, if any, is a union of tori. Let $F_1, \ldots, F_n$ be the top-dimensional faces of $B_A$ whose corresponding vertices of $\text{Newt}(\Delta_M)$ have coefficient $\pm 1$. Then $\Sigma_A$ is exactly the projection to $S(\text{ab}(\pi_1 M))$ of the interiors of the $F_i$.

**Proof.** Let $Q = \text{ab}(\pi_1(M))$. I will show:

**Lemma 5.2.** Let $M$ be as above. Then $\Sigma_A = \Sigma_{\mathbb{Z}Q/(\Delta_M)}$. 

Let me now deduce the theorem assuming the lemma. By Theorem 4.1, the components of $\Sigma_{\mathbb{Z}Q/(\Delta_M)}$ correspond to the vertices of $\text{Newt}(\Delta_M)$ whose coefficients are $\pm 1$. Such a vertex $v$ corresponds to:

$$C_v = \bigcap \{ H_{qv}w^{-1} \mid w \text{ is a vertex of } \text{Newt}(p) \text{ distinct from } v \},$$

where $H_q$ is the hemisphere $\{ [\chi] \in S(Q) \mid \chi(q) > 0 \}$. To prove the theorem it suffices to show $C_v$ is the same as the projection into $S(Q)$ of the interior of the face $F_v$ of $B_A$ corresponding to $v$. Translate $\text{Newt}(\Delta_M)$ so it is balanced about 0—this doesn’t change $C_v$ or $\| \cdot \|_A$. Now note that the cone over the interior of $F_v$ is

$$\{ \phi \mid \phi(v) > \phi(w) \text{ for all vertices } w \text{ of } \text{Newt}(p) \text{ distinct from } v \}.$$  

It’s easy to see that this cone projects to $C_v$ in $S(Q)$. This proves the theorem modulo the lemma. Let’s go back and prove the lemma.

**Proof of Lemma 5.2.** The idea of the proof is that Proposition 3.4 says that $B_M$ is close, in some sense, to $\mathbb{Z}Q/(\Delta_M)$. Using the properties in Section 4, we have (notation changed for clarity):

$$\Sigma_A = \Sigma(B_M) = \Sigma(\mathbb{Z}Q / \text{Ann}(B_M)) = \Sigma(\mathbb{Z}Q / \sqrt{\text{Ann}(B_M)}).$$

For any finitely generated module $B$ we have $\sqrt{\text{Ann}(B)} = \sqrt{E_0(B)}$, and so

$$\Sigma_A = \Sigma(\mathbb{Z}Q / \sqrt{E_0(B_M)}).$$

Let $m$ be the augmentation ideal of $\mathbb{Z}Q$. Since $\mathbb{Z}Q/m = \mathbb{Z}$, the invariant $\Sigma_{\mathbb{Z}Q/m}$ is all of $S(Q)$. So for any ideal $I$, we have $\Sigma(\mathbb{Z}Q/(I \cap m)) = \Sigma(\mathbb{Z}Q/I) \cap \Sigma(\mathbb{Z}Q/m) = \Sigma(\mathbb{Z}Q/I)$. Thus

$$\Sigma_A = \Sigma(\mathbb{Z}Q / (\sqrt{E_0(B_M)} \cap m)).$$

By Proposition 3.4, $\sqrt{E_0(B_M)} \cap m = \sqrt{(\Delta_M)} \cap m$, so

$$\Sigma_A = \Sigma(\mathbb{Z}Q / (\sqrt{(\Delta_M)} \cap m)) = \Sigma(\mathbb{Z}Q / (\sqrt{(\Delta_M)})) = \Sigma(\mathbb{Z}Q/(\Delta_M)),$$

as required. This completes the proof of the lemma and thus the theorem. 

☐

**5.3. Comparison of $\Sigma_{G'}$ and $\Sigma_A$ when the homology is not free.** Let $M$ be a 3-manifold and $G$ its fundamental group. In Section 1.5, I discussed the connection between $\Sigma_{G'}$ and cohomology classes representing fibrations of $M$ over the circle. This is true independent of whether $H_1(M, \mathbb{Z})$ has torsion. In Section 1.5, $\Sigma_{G'}$ and $\Sigma_A$ were compared under the assumption that $H_1(M, \mathbb{Z})$ is free. In this case, it is easy to see $\Sigma_A \supset \Sigma_{G'}$, because $\Sigma_A = \Sigma_{B_M}$, and $B_M = G'/G''$ is a quotient of $G'$. When $H_1(M, \mathbb{Z})$ is not free, the relation $\Sigma_A \supset \Sigma_{G'}$ is still true, but not immediate since $B_M$ is
a quotient of the kernel of the map $G \to ab(G)$, but that kernel properly contains $G'$.

The purpose of this subsection is simply to prove the that $\Sigma_A \supset \Sigma_{G'}$ for any $M$, and so show that the motivation given in Section 1.5 makes sense regardless of whether $H_1(M, \mathbb{Z})$ is free.

**Proposition 5.4.** Let $M$ be a 3-manifold. Then $\Sigma_A \supset \Sigma_{G'}$.

*Proof.* Let $N$ be the kernel of the map from $G$ to its free abelianization. It is clear that $\Sigma_A \supset \Sigma_N$ as the Alexander invariant $B_M$ is a quotient of $N$. By Proposition 3.4 of [BNS], $\Sigma_N = \Sigma_{G'}$ and we are done. \qed

6. Example of a specific link exterior in $S^3$.

Let $L$ be the link in Figure 6.1.

![Figure 6.1. The link L in S^3.](image)

Let $M = S^3 \setminus N(L)$ be the exterior of $L$. In this section, I’ll show that $M$ is a fibered 3-manifold where the answer to Question B is no. I will do this by explicitly computing the BNS invariants $\Sigma$ and $\Sigma_A$, showing that $\Sigma$ is non-empty and that each component of $\Sigma$ is properly contained in the corresponding component of $\Sigma_A$. The manifold $M$ is hyperbolic with volume $8.997\ldots$, as can be checked with the program SnapPea [W], or better, Snap [G], but I won’t use this fact. I found this example by a brute force search—the program SnapPea was used to find many links whose fundamental groups have a presentation with two generators and one relator. For such groups, it is easy to calculate $\Sigma$ and $\Sigma_A$ directly, as I will do below, and eventually I came across this example.
According to SnapPea, $\pi_1 M$ has a presentation with two generators $a$ and $b$ and defining relation
\[
a^2 b a^{-1} b^2 a b^{-1} b^{-3} a^{-1} b a^2 b a^{-1} b a b^{-1} a^{-2}.
\]

A meridian for the unknotted component is $b^{-1} a^{-1} b a^2 b a^{-1} b a b^{-1} b^{-3}$ and a meridian for the other component is $a^{-1} b^{-1}$.

Let $\tilde{X}$ be the 2-dimensional CW-complex corresponding to the above presentation. Let $G = \pi_1(X)$. The abelian group $ab(G)$ is freely generated by images of $a$ and $b$, and so $\mathbb{Z}[ab(G)] = \mathbb{Z}[(a, b)]$.

Let $\tilde{X}$ be the universal abelian cover of $X$. It is natural to think of the 1-skeleton of $\tilde{X}$ as the integer grid in $H_1(X, \mathbb{R})$. Let $\delta$ be the lift of $aba^{-1} b^{-1}$ starting at 0, which freely generates the 1-chains of $\tilde{X}$ as a $\mathbb{Z}[ab(G)]$ module. The 2-chains of $\tilde{X}$ are generated by any lift of the 2-cell of $X$. Let $\gamma$ be the lift of the relator to 1-skeleton of $\tilde{X}$ starting at 0, which is homologous in the 1-skeleton to $(a^2 b - ab - a + 1)\delta$. Thus

\[
B_X = H_1(\tilde{X}, \mathbb{Z}) = \mathbb{Z}[(a, b)]/(a^2 b - ab - a + 1).
\]

So $\Delta_M = \Delta_X = a^2 b - ab - a + 1$. By Theorem 5.1, or, since $B_M$ cyclic, Theorem 4.1 directly, we find that $\Sigma_A$ is all of $S(ab(G))$ except the four points $\{\pm [b^*], \pm [a^* - b^*]\}$, where $\{a^*, b^*\}$ is the dual basis to $\{a, b\}$.

To compute $\Sigma$, I’ll use Brown’s procedure for computing $\Sigma$ for any group with a 2-generator, 1-relator presentation [Bro, §4]. Think of the 1-skeleton of $\tilde{X}$ as the integer grid in $H_1(X, \mathbb{R})$. Let $C$ be the convex hull of $\gamma$, the lift of the relator. A vertex $v$ of $C$ is called simple if $\gamma$ passes through $v$ only once. Figure 6.2 shows $C$ with the 2 simple vertices $v_1$ and $v_2$ marked.

Theorem 4.4 of [Bro] shows that in our case $\Sigma$ consists of two components $C_i$, for $i = 1, 2$, where

\[
C_i = \bigcap \{H_{v_i w^{-1}} \mid w \text{ is a vertex of } C \text{ distinct from } v_i\}.
\]

Thus $\Sigma$ is the union of the two open intervals pictured in Figure 6.3, and each component of $\Sigma$ is properly contained in the corresponding component of $\Sigma_A$. So $M$ shows that the answer to Question B is no.
Figure 6.2. The region $C$. The two dots are the simple vertices $v_1$ and $v_2$.

Figure 6.3. $\Sigma \subset S(ab(G))$ consists of the two open intervals shown. $\Sigma_A$ is the complement of the four grey dots.

References


[Big2] , Braid groups are linear, preprint, 2000, math/0005038.


Received October 22, 1999 and revised July 17, 2000. This work was partially supported by a Sloan Dissertation Fellowship.

DEPARTMENT OF MATHEMATICS
HARVARD UNIVERSITY
CAMBRIDGE, MA 02138
E-mail address: nathand@math.harvard.edu
SUR LES OPÉRATEURS NILPOTENTS À IMAGES DES ITÉRÉS FERMÉES DANS UN ESPACE DE BANACH

Abdelkhalek Faouzi

This work is dedicated to the memory of Professor M. El Oufir.

Nous montrons que, pour un opérateur linéaire $A$ nilpotent à images des itérés fermées dans un espace de Banach $E$, tout sous-espace de $E$ de codimension finie contient un sous-espace réduisant pour $A$ de codimension finie. D’autre part, par le biais de l’étude des sous-espaces réduisants minimaux contenant un sous-espace donné, nous prouvons que toute extension continue d’un opérateur nilpotent à images des itérés fermées par un opérateur nilpotent défini en dimension finie est aussi à images des itérés fermées. D’autres résultats sur les opérateurs nilpotents à images des itérés fermées sont établis.

1. Introduction.

Soient $E$ un espace de Banach complexe et $B(E)$ l’algèbre des opérateurs linéaires continus sur $E$. Par sous-espace de $E$ nous entendons sous-espace vectoriel fermé de $E$. Pour $A \in B(E)$, nous désignons par $\text{Lat} A$ le treillis des sous-espaces de $E$ invariants pour $A$. L’opérateur $A \in B(E)$ est dit à images des itérés fermées si $\text{Im} A^k = A^k(E)$ est fermée pour tout entier positif $k$. Un résultat assez remarquable de M. El Oufir ([4], Chapitre IV, Théorème 3) affirme qu’un opérateur nilpotent est à images des itérés fermées si et seulement si tout sous-espace de $E$ de dimension finie est contenu dans un sous-espace réduisant pour $A$ de dimension finie. Ce théorème, ainsi que l’étude en dimension finie des extensions d’opérateurs linéaires (voir [3] et [5]) ont motivé notre intérêt pour l’étude des extensions d’opérateurs nilpotents à images des itérés fermées par des opérateurs nilpotents “de dimension finie” (i.e., définis sur des espaces de dimension finie). De ce fait, il s’est avéré nécessaire pour nous d’établir certaines propriétés sur les opérateurs nilpotents à images des itérés fermées. Ainsi, comme conséquence non triviale du théorème d’El Oufir mentionné ci-dessus, nous montrons que si $A \in B(E)$ est nilpotent à images des itérés fermées, alors tout sous-espace invariant pour $A$ de codimension finie contient un sous-espace réduisant pour $A$ de codimension finie. La réciproque reste à étudier. D’autre part, nous démontrons, toujours pour $A \in B(E)$ nilpotent
à images des itérés fermées, que s’il existe $E_1, E_2, F_1, F_2$, dans $\text{Lat} A$ tels que $(E, A) = (E_1, A_1) \oplus (F_1, B_1) = (E_2, A_2) \oplus (F_2, B_2)$ où $A_i = A|E_i$ et $B_i = B|F_i$ (pour $i = 1, 2$) et si $(E_1, A_1)$ et $(E_2, A_2)$ sont isomorphes et de dimension finie, alors $(F_1, B_1)$ et $(F_2, B_2)$ sont isomorphes.

Dans une autre partie, nous montrons que toute extension continue d’un opérateur nilpotent à images des itérés fermées par un opérateur nilpotent “de dimension finie” est aussi à images des itérés fermées. Un contre-exemple montre qu’en général une extension continue d’un opérateur nilpotent à images des itérés fermées n’est pas toujours à images des itérés fermées. Enfin, nous indiquons comment l’étude des extensions des opérateurs nilpotents à images des itérés fermées par des opérateurs nilpotents “de dimension finie” peut être ramenée au cas de la dimension finie. Nous traitons pour cela deux cas particuliers (assez intéressants) mais le principe reste le même à chaque fois qu’une condition d’unicité à isomorphie près est vérifiée.

2. Notions préliminaires.

Soient $E$ un espace de Banach sur le corps des complexes $\mathbb{C}$ et $A \in B(E)$. On peut munir $E$ d’une structure de $\mathbb{C}[X]$-module en posant $P(x) = P(A)x$ pour tout $x \in E$ et pour tout $P \in \mathbb{C}[X]$. Nous notons $(E, A)$ le module ainsi défini.

Nos méthodes sont, pour la plupart, inspirées de la théorie des groupes abéliens (voir [6] et [7]) laquelle, interprétée en termes de modules sur un anneau principal, peut être appliquée aux opérateurs linéaires.

Si $M$ est un sous-ensemble de $E$, on notera $\text{Vect}(M)$ le plus petit sous-espace de $E$ contenant $M$ et $\text{Vect}_A(M)$ le plus petit sous-espace invariant pour $A$ contenant $M$.

Si $A$ est nilpotent, on définit l’exposant $e(x)$ de $x \in E$ comme étant le plus petit entier positif tel que $A^k x = 0$. On définit la hauteur de $x \in E$, notée $h(x)$, comme étant l’entier $k \in \mathbb{N}$ tel que $x \in A^k E$ et $x \notin A^{k+1} E$ avec la convention $h(0) = \infty$. On vérifie facilement les propriétés suivantes de la hauteur:

(i) $h(x + y) \geq \inf(h(x), h(y))$ (avec égalité si $h(x) \neq h(y)$).
(ii) $h(\alpha x) = h(x)$, pour tout $\alpha \in \mathbb{C}, \alpha \neq 0$.
(iii) $h(Ax) \geq h(x) + 1$.

Soit $F$ un sous-espace de $E$. On dit que $F$ est pur dans le $\mathbb{C}[X]$-module $(E, A)$ si:

$$\forall P \in \mathbb{C}[X], \quad P(A)F = F \cap P(A)E.$$ 

On a les propriétés suivantes:

**Lemme 2.1.** Si $F$ est pur dans $(E, A)$, alors $F$ est pur dans $(G, A)$ pour tout sous-espace $G$ invariant pour $A$ et contenant $F$.

**Lemme 2.2.** Tout sous-espace réduisant pour $A$ est pur dans $(E, A)$. 

Lemme 2.3. Si \( E \) est de dimension finie et \( A \in \mathcal{B}(E) \) est nilpotent, alors il y a équivalence entre sous-espace pur dans \((E, A)\) et sous-espace réduisant pour \( A \).

Notation 2.4. Nous réservons le signe \( \oplus \) à la somme directe topologique, alors que nous désignerons une somme directe algébrique par le signe \( \dot{+} \). D’autre part, le signe \( \sim \) désignera l’isomorphie de modules, d’espaces vectoriels ou d’espaces de Banach.

3. Quelques propriétés des opérateurs nilpotents à images des itérés fermées.

Lemme 3.1. Soient \( E \) et \( H \) deux espaces de Banach, \( A \) un opérateur de \( E \) dans \( H \) et \( F \) un sous-espace de codimension finie dans \( E \). Si \( \text{Im} A = A(E) \) est fermée dans \( H \), alors \( \text{Im}(A|F) = A(F) \) est fermée aussi.

Démonstration. Soit le morphisme \( T \) défini de \( F/(F \cap \text{Ker}A) \) dans \( E/\text{Ker}A \) par \( T(y + F \cap \text{Ker}A) = y + \text{Ker}A \), pour tout \( y \in F \). \( T \) est injectif d’image \( \text{Im}T = (F + \text{Ker}A)/\text{Ker}A \). Montrons que \( \text{Im}T \) est fermée dans \( E/\text{Ker}A \). Pour cela, il suffit que l’on montre que \( F + \text{Ker}A \) est fermé dans \( E \). On considère le morphisme \( S \) de \( \text{Ker}A/(F \cap \text{Ker}A) \) dans \( E/F \) défini par \( S(z + F \cap \text{Ker}A) = z + F \), pour tout \( z \in \text{Ker}A \). \( S \) est injectif, ce qui entraîne \( \dim(\text{Ker}A/(F \cap \text{Ker}A)) \leq \dim(E/F) = \text{codim} F < +\infty \). Ainsi, \( F \cap \text{Ker}A \) est de codimension finie dans \( \text{Ker}A \), donc admet un supplémentaire algébrique \( G \) dans \( \text{Ker}A \). Par suite, \( F + \text{Ker}A = F + G \) avec \( G \) de dimension finie, donc \( F + \text{Ker}A \) est fermé, et la somme \( F + G \) est topologique. On en déduit que \( \text{Im}T = (F + \text{Ker}A)/\text{Ker}A \) est fermée dans \( E/\text{Ker}A \), ce qui équivaut à l’existence d’une constante \( C_T \) telle que (voir par exemple [8], Théorème 5.1, Page 70):

\[
\|y + F \cap \text{Ker}A\| \leq C_T\|Ty\|, \quad \text{pour tout } y \in F.
\]

De même, comme \( \text{Im}A \) est fermée, il existe une constante \( C_A \) telle que:

\[
\|x + \text{Ker}A\| \leq C_A\|Ax\|, \quad \text{pour tout } x \in E.
\]

Donc, pour tout \( y \in F \), on a:

\[
\|y + \text{Ker}(A|F)\| = \|y + F \cap \text{Ker}A\| \leq C_T\|Ty\| = C_T\|y + \text{Ker}A\| \leq C_TC_A\|Ay\| = C_TC_A\|(A|F)y\|.
\]

Ainsi il existe une constante \( C_{A|F} = C_TC_A \) telle que:

\[
\|y + \text{Ker}(A|F)\| \leq C_{A|F}\|(A|F)y\|, \quad \text{pour tout } y \in F.
\]

Ceci traduit le fait que \( \text{Im}(A|F) \) est fermée dans \( H \).

Le lemme suivant est facile à établir:
Lemme 3.2. Soient \( E \) un espace de Banach complexe et \( A \in \mathcal{B}(E) \) tels que \( E = E_1 \oplus E_2 \) avec \( E_1, E_2 \in \operatorname{Lat} A \). Alors, on a:
\[
\operatorname{Im} A = \operatorname{Im} A_1 + \operatorname{Im} A_2 \quad \text{et} \quad \overline{\operatorname{Im} A} = \overline{\operatorname{Im} A_1} \oplus \overline{\operatorname{Im} A_2},
\]
où \( A_1 = A|_{E_1}, A_2 = A|_{E_2} \) et \( \overline{\operatorname{Im} A} \) désigne la fermeture de \( \operatorname{Im} A \) pour la topologie de la norme.

Définition 3.3. Soit \( E \) un espace de Banach complexe. Un opérateur \( A \in \mathcal{B}(E) \) est dit à images des itérés fermées si \( \operatorname{Im} A^k = A^k E \) est fermée, pour tout \( k \in \mathbb{N} \).

Dans [4], M. El Oufir avait caractérisé les opérateurs nilpotents à images des itérés fermées dans un espace de Banach à l’aide des sous-espaces réducteurs de dimension finie. Nous reprenons ce résultat en en donnant une démonstration plus rapide.

Théorème 3.4 (El Oufir). Soient \( E \) un espace de Banach complexe et \( A \in \mathcal{B}(E) \) un opérateur nilpotent. Alors, les propriétés suivantes sont équivalentes:

(i) \( A \) est à images des itérés fermées.
(ii) Pour tout élément \( x \) de \( E \), il existe un sous-espace réductant pour \( A \) de dimension finie contenant \( x \).
(iii) Pour tout sous-espace \( F \) de \( E \) de dimension finie, il existe un sous-espace réductant pour \( A \) de dimension finie contenant \( F \).

Démonstration. (i) \(\Rightarrow\) (ii): Si \( x \) est nul, le résultat est trivial. Supposons \( x \) non nul et posons \( m = 1 + h(A^{k-1}x) \) où \( k = e(x) \). Deux cas sont à distinguer:

1er cas. Pour tout \( j \leq k-1, h(A^j x) = j + h(x) \).
Ceci équivaut à \( h(x) = m - k \). Soit \( x_1 \in E \) tel que \( x = A^{m-k} x_1 \). Alors, \( e(x_1) = m \) et \( A^{m-k} x_1 \notin A^m E \). Comme \( A^m E \) est fermée, alors, d’après le théorème de Hahn-Banach, il existe une forme linéaire \( u \) continue sur \( E \) telle que:
\[
u(A^m E) = 0 \quad \text{et} \quad u(A^{i} x_1) = \delta_{m-1}^i \quad \text{pour} \quad 0 \leq i \leq m - 1,
\]
où \( \delta \) désigne le symbole de Kronecker. On vérifie alors aisément que l’opérateur:
\[
B = \sum_{s=1}^{k} A^{s-1}(x_1 \otimes u)A^{k-s}
\]
est continu sur \( E \) d’image \( \operatorname{Vect}_A(x_1) \) et qu’il commute avec \( A \). Par suite, \( \operatorname{Vect}_A(x_1) \) est réductant pour \( A \) et contient \( \operatorname{Vect}_A(x) \).

2ème cas. Il existe \( j \leq k - 1 \) tel que \( h(A^j x) > j + h(x) \).
Soit \( j_1 = \min\{j \leq k - 1; h(A^j x) > j + h(x)\} \). Pour \( x' = A^{k-j_1} x \), on est dans le premier cas \( e(x') = j_1 \) et \( h(x') = h(A^{j_1-1} x') + 1 - j_1 \), donc il
exist $x_1 \in E$ tel que $A^{m-j_1}x_1 = A^{k-j_1}x_1$ et $\text{Vect}_A(x_1)$ est réduisant pour $A$ contenant $\text{Vect}_A(A^{k-j_1}x_1)$. Quitte à changer de générateur, on peut supposer que

$$\text{Vect}_A(x_1, x_1) = \text{Vect}_A(x_1) \oplus \text{Vect}_A(y)$$

avec $y = x - A^{m-k}x_1$. Le vecteur $y$ est d’exposant $e(y) = k - j_1$. On passe alors à $y$. En poursuivant jusqu’au bout, on obtient un sous-espace $H = \bigoplus_{i=1}^n \text{Vect}_A(x_i)$ réduisant pour $A$ et contenant $\text{Vect}_A(x)$.

(ii) $\implies$ (iii): Le résultat étant vrai pour $F = \{0\}$, supposons $\dim F > 0$ et raisonnons par récurrence sur $\dim F$. Soit $x$ un élément non nul de $F$, alors d’après (ii), il existe un sous-espace $E_1$ réduisant pour $A$, de dimension finie et contenant $x$. Soit $E_2 \in \text{Lat}_A$ tel que $E = E_1 \oplus E_2$, alors $F \subseteq (E_1 + F) = E_1 \oplus ((E_1 + F) \cap E_2)$. Comme $E_1 \cap F$ n’est pas réduit à $\{0\}$, on a $\dim(E_1 + F) < \dim E_1 + \dim F$, ce qui entraîne $\dim((E_1 + F) \cap E_2) < \dim F$. $A|E_2$ étant à images des itérés fermées, alors, d’après l’hypothèse de récurrence, il existe $H_2$ et $G_2$ dans $\text{Lat}(A|E_2)$ tels que:

$$E_2 = H_2 \oplus G_2, \quad \dim H_2 < +\infty \quad \text{et} \quad (E_1 + F) \cap E_2 \subseteq H_2.$$

De plus, il est clair que $E = (E_1 \oplus H_2) \oplus G_2$. Ainsi, le sous-espace $E_1 \oplus H_2$ est réduisant pour $A$, de dimension finie et contient $F$.

(iii) $\implies$ (i): Montrons d’abord que $\text{Im}A = A(E)$ est fermée. Soit $x \in \overline{\text{Im}A}$, alors il existe un sous-espace $H$ réduisant pour $A$, de dimension finie et contenant $x$. Soit $G \in \text{Lat}_A$ tel que $E = H \oplus G$. $H$ étant de dimension finie, $\text{Im}(A|H)$ est aussi de dimension finie, donc est fermée. D’après le Lemme 3.2, on a $\overline{\text{Im}A} = \text{Im}(A|H) \oplus \overline{\text{Im}A|G}$. Dans cette décomposition le vecteur $x$ s’écrit $x = y + z$ où $y \in \text{Im}(A|H)$ et $z \in \overline{\text{Im}(A|G)}$. L’unicité de la décomposition entraîne $x = y$ et $z = 0$. Par conséquent, $\text{Im}A$ est fermée.

Comme $\text{Lat}_A \subseteq \text{Lat}_A^i$, pour tout $j \in \mathcal{N}$, on montre de la même manière que les images $\text{Im}A^j = A^jE$ sont fermées.

**Théorème 3.5.** Soient $E$ un espace de Banach complexe et $A \in \mathcal{B}(E)$ un opérateur nilpotent à images des itérés fermées. Alors, tout sous-espace invariant pour $A$ de codimension finie contient un sous-espace réduisant pour $A$ de codimension finie.

**Démonstration.** Soient $F \in \text{Lat}A$ de codimension finie dans $E$ et $G$ un supplémentaire algébrique quelconque de $F$ dans $E$. Notons que $G$ n’est pas nécessairement invariant pour $A$. Comme $G$ est de dimension finie, il résulte du caractère nilpotent de $A$ que $\text{Vect}_A(G)$ est de dimension finie, et par suite $\dim(\text{Vect}_A(G) \cap F) < +\infty$. D’après le Lemme 3.1, l’opérateur $A|F$ est à images des itérés fermées, donc, suite au théorème d’El Oufir (Théorème 3.4), il existe un sous-espace $H$ réduisant pour $A|F$, de dimension finie et contenant $\text{Vect}_A(G) \cap F$. Soit $K \in \text{Lat}(A|F)$ tel que $F = H \oplus K$. 
Posons $H' = \text{Vect}_A(G) + H$, alors:

$$H' \in \text{Lat}A, \quad \dim H' < +\infty \quad \text{et} \quad E = H' \oplus K.$$ 

Ainsi, $K$ est un sous-espace réduisant pour $A$, de codimension finie et contenu dans $F$.

**Notation 3.6.** Soit $A \in \mathcal{B}(E)$ un opérateur nilpotent. Pour $m \in \mathcal{N}$, on pose:

$$(1) \quad f_m(E, A) = \dim(\text{Ker}A \cap A^m E/\text{Ker}A \cap A^{m+1} E).$$

En dimension finie, $f_m(E, A)$ est le nombre de sous-espaces cycliques de $E$ de dimension $m + 1$ qui interviennent dans une décomposition de Jordan de $A$ dans $E$. Il est connu que si $E_1$ et $E_2$ sont deux $C$-espaces vectoriels de dimension finie et si $A_1$ et $A_2$ sont deux opérateurs nilpotents respectivement sur $E_1$ et $E_2$, alors:

$$(2) \quad (E_1, A_1) \cong (E_2, A_2) \iff \forall m \in \mathcal{N}, f_m(E_1, A_1) = f_m(E_2, A_2).$$

**Proposition 3.7.** Soient $E$ un espace vectoriel de dimension finie sur $\mathcal{C}$, $A \in \mathcal{B}(E)$ un opérateur nilpotent et $E_1, E_2, F_1, F_2$ des sous-espaces invariants pour $A$ tels que:

$$\quad (E, A) = (E_1, A_1) \oplus (F_1, B_1) = (E_2, A_2) \oplus (F_2, B_2)$$

où $A_i = A|E_i$ et $B_i = A|F_i$ ($i = 1, 2$).

Si $(E_1, A_1) \cong (E_2, A_2)$, alors $(F_1, B_1) \cong (F_2, B_2)$.

**Démonstration.** D’après $(2)$, il suffit de vérifier que, pour tout $m \in \mathcal{N}$, on a:

$$(3) \quad f_m(E, A) = f_m(E_1, A_1) + f_m(F_1, B_1).$$

Or $\text{Ker}A \cap A^m E = (\text{Ker}A_1 \cap A_1^m E_1) \oplus (\text{Ker}B_1 \cap B_1^m F_1)$, donc en posant:

$$U_m(E, A) = (\text{Ker}A \cap A^m E) / (\text{Ker}A \cap A^{m+1} E),$$

on obtient l’isomorphisme de $C$-espaces vectoriels suivant:

$$U_m(E, A) \cong U_m(E_1, A_1) \oplus U_m(F_1, B_1).$$

L’égalité $(3)$ s’en déduit facilement.

Le résultat de la Proposition 3.7 se généralise au cas des opérateurs nilpotents à images des itérés fermées dans un espace de Banach quelconque, et ce de la façon suivante:

**Théorème 3.8.** Soient $E$ un espace de Banach complexe, $A \in \mathcal{B}(E)$ nilpotent à images des itérés fermées et $E_1, E_2, F_1, F_2$ des sous-espaces invariants pour $A$ tels que:

$$\quad (E, A) = (E_1, A_1) \oplus (F_1, B_1) = (E_2, A_2) \oplus (F_2, B_2)$$

où $A_i = A|E_i$ et $B_i = A|F_i$ ($i = 1, 2$).
Si \((E_1, A_1)\) et \((E_2, A_2)\) sont isomorphes et de dimension finie, alors \((F_1, B_1)\) et \((F_2, B_2)\) sont isomorphes.

Démonstration.

1ère étape. On va commencer par montrer que \(E_1 + E_2\) est contenu dans un sous-espace \(G\) invariant pour \(A\) de dimension finie, que \(F_1 \cap F_2\) contient un sous-espace \(K\) invariant pour \(A\) et que l’on a \(E = G \oplus K\).

D’après le Lemme 2.3 entraîne que \(E\). Il s’ensuit du Lemme 2.1 que

\[ F_1 = H_1 \oplus K_1, \quad \dim H_1 < +\infty \quad \text{et} \quad \operatorname{Vect}_A(E_1 + E_2) \cap F_1 \subseteq H_1. \]

Posons \(H'_1 = H_1 + \operatorname{Vect}_A(E_1 + E_2)\), alors:

\[ H'_1 \in \operatorname{Lat} A, \quad \dim H'_1 < +\infty \quad \text{et} \quad E = H'_1 \oplus K_1. \]

D’après le Lemme 3.1, l’opérateur \(A|K_1\) est à images des itérés fermées. Comme \(\dim(\operatorname{Vect}_A(E_1 + E_2) \cap F_1) < +\infty\), il résulte du théorème d’El Oufr qu’il existe \(H_1\) et \(K_1\) dans \(\operatorname{Lat} B_1\) tels que:

\[ F_1 = H_1 \oplus K_1, \quad \dim H_1 < +\infty \quad \text{et} \quad \operatorname{Vect}_A(E_1 + E_2) \cap F_1 \subseteq H_1. \]

2ème étape. \(E_1\) est réduit pour \(A\), donc pur dans \((E, A)\) d’après le Lemme 2.2 Il s’ensuit du Lemme 2.1 que \(E_1\) est pur dans \((G, A|G)\). Comme \(G\) est de dimension finie, le Lemme 2.3 entraîne que \(E_1\) est réduit pour \(A|G\). On montre de même que \(E_2\) est réduit pour \(A|G\). Soient donc \(G_1\) et \(G_2\) dans \(\operatorname{Lat} A\) tels que \(G = E_1 \oplus G_1 = E_2 \oplus G_2\), alors:

\[ (G, A|G) = (E_1, A_1) \oplus (G_1, C_1) = (E_2, A_2) \oplus (G_2, C_2) \]

où \(C_i = A|G_i\) \(i = 1, 2\).

Comme \((E_1, A_1) \cong (E_2, A_2)\), alors \((G_1, C_1) \cong (G_2, C_2)\) (d’après la Proposition 3.7) et il s’ensuit que:

\[ (G_1, C_1) \oplus (K, T) \cong (G_2, C_2) \oplus (K, T) \]

où \(T = A|K\).

D’autre part, des égalités (4) on déduit:

\[ (E, A) = (E_1, A_1) \oplus (G_1, C_1) \oplus (K, T) = (E_2, A_2) \oplus (G_2, C_2) \oplus (K, T). \]

Aussi a-t-on:

\[ (G_1, C_1) \oplus (K, T) \cong (E, A)/(E_1, A_1) \cong (F_1, B_1). \]

\[ (G_2, C_2) \oplus (K, T) \cong (E, A)/(E_2, A_2) \cong (F_2, B_2) \] .

La conclusion \((F_1, B_1) \cong (F_2, B_2)\) découle immédiatement de (5) et (6).
**Remarque 3.9.** Dans le théorème précédent, supposer uniquement \((E_1, A_1)\) isomorphe à \((E_2, A_2)\) sans que la dimension ne soit finie n’entraîne pas nécessairement l’existence d’un isomorphisme entre \((F_1, B_1)\) et \((F_2, B_2)\). En effet, soient \(T \in \mathcal{B}(E)\) un opérateur nilpotent cyclique d’ordre \(k \in \mathbb{N}^*\) et \((E, A) = \bigoplus_{i \in \mathbb{N}} (H_i, T_i)\) où \((H_i, T_i) = (H, T)\), pour tout \(i \in \mathbb{N}\). Si on pose

\[(E_1, A_1) = (E, A) \quad \text{et} \quad (E_2, A_2) = \bigoplus_{i=2}^{+\infty} (H_i, T_i)\]

alors il est évident que \((E_1, A_1) \cong (E_2, A_2)\). Cependant,

\[(E, A)/(E_1, A_1) \cong \{(0), 0\} \quad \text{et} \quad (E, A)/(E_2, A_2) \cong (H, T).\]

4. **Extensions continues d’opérateurs nilpotents à images des itérés fermées par des opérateurs nilpotents “de dimension finie”**.

Soient \(E’\) et \(E''\) deux espaces de Banach complexes. *Une extension continue* de \(A’ \in \mathcal{B}(E’)\) par \(A'' \in \mathcal{B}(E'')\) est un opérateur \(A \in \mathcal{B}(E)\) tel que le module \((E, A)\) soit une extension du module \((E’, A’)\) par le module \((E'', A'')\). Ceci s’exprime aussi par l’existence d’une suite exacte courte:

\[
0 \longrightarrow (E’, A’) \xrightarrow{\chi} (E, A) \xrightarrow{\sigma} (E'', A'') \longrightarrow 0.
\]  

(7)

Nous allons nous restreindre aux extensions \((E, A)\) telles que \(E\) soit de la forme \(E' \oplus E''\), ce qui permet d’utiliser une représentation matricielle des opérateurs de \(\mathcal{B}(E)\) (une telle restriction est assez naturelle car dans le cas des espaces hilbertiens on a toujours \(E = E' \oplus E''\) -existence du complément orthogonal-). Un opérateur \(A \in \mathcal{B}(E' \oplus E'')\) est une extension de \(A' \in \mathcal{B}(E')\) par \(A'' \in \mathcal{B}(E'')\) si et seulement s’il existe un opérateur linéaire continu \(B\) de \(E''\) dans \(E'\) tel que la matrice de \(A\) soit de la forme:

\[
A = \begin{bmatrix}
A' & B \\
0 & A''
\end{bmatrix}.
\]  

(8)

Dans la suite, on suppose que \(A'' \in \mathcal{B}(E'')\) est “*de dimension finie*” (i.e., \(E''\) *est de dimension finie*). Si \(A \in \mathcal{B}(E)\) est une extension de \(A'\) par \(A''\), alors \(E'\) est un sous-espace de \(E\) de codimension finie, donc admet un supplémentaire topologique dans \(E\). Ceci vient de l’isomorphisme \(E'' \cong E/E'\) et du fait que \(E''\) soit de dimension finie. En identifiant \(E''\) à un supplémentaire de \(E'\) dans \(E\), on peut écrire \(E = E' \oplus E''\). D’autre part, de l’exactitude de (7) on déduit qu’un changement de supplémentaire de \(E'\) dans \(E = E' \oplus E''\) laisse invariante la forme de Jordan de \(A''\). Supposons en plus que \(A'' \in \mathcal{B}(E'')\) est nilpotent, alors quitte à changer, si besoin est, de supplémentaire de \(E'\)
dans $E' \oplus E''$, on peut supposer que l’extension $A$ vérifie:

$$
\begin{cases}
E'' = \bigoplus_{j=1}^{q} \operatorname{Vect}_{A^n}(e_j) & \text{avec } e_j \text{ d’exposant } n_j,
A^s e_j = A'' e_j & \text{pour } s < n_j,
A^n e_j \in E' & \text{pour } 1 \leq j \leq q.
\end{cases}
$$

(9)

On voit donc que l’étude des extensions de $A' \in \mathcal{B}(E')$ par un opérateur nilpotent “de dimension finie” est liée à l’étude des sous-espaces réduisants pour $A'$ minimaux contenant un sous-espace de dimension finie. Un cas qui serait intéressant à étudier est celui où $A' \in \mathcal{B}(E')$ est nilpotent à images des itérés fermées.

**Théorème 4.1.** Soient $A' \in \mathcal{B}(E')$ nilpotent à images des itérés fermées et $A'' \in \mathcal{B}(E'')$ nilpotent “de dimension finie”. Alors toute extension continue de $A'$ par $A''$ est à images des itérés fermées.

**Démonstration.** Soit $A \in \mathcal{B}(E)$ une extension de $A'$ par $A''$. D’après ce qui précède, on peut supposer $E = E' \oplus E''$ et que $A$ vérifie (9). Comme $A'$ est à images des itérés fermées, alors d’après le théorème d’El Ouﬁr (Théorème 3.4), il existe un sous-espace $F'$ réduisant pour $A'$, de dimension finie et contenant $\operatorname{Vect}(A^n e_1, \ldots, A^n e_q)$. Soit $G' \subseteq \operatorname{Lat} A'$ tel que $E' = G' \oplus F'$, alors $E' \oplus E'' = G' \oplus (F' \oplus E'')$. De plus, $G'$ et $F' \oplus E''$ étant dans $\operatorname{Lat} A$, on a $A = A|G' \oplus A|(F' \oplus E'')$. Or, suite au Lemme 3.1, $A|G' = A'|G'$ est à images des itérés fermées et, d’autre part, $F' \oplus E''$ étant de dimension finie, $A|(F' \oplus E'')$ est aussi à images des itérés fermées. Par conséquent, il résulte du Lemme 3.2 que $A$ est à images des itérés fermées.

**Remarque 4.2.** En général, une extension continue d’un opérateur nilpotent à images des itérés fermées n’est pas nécessairement à images des itérés fermées. En effet, soient $H'$ et $H''$ deux espaces de Hilbert isomorphes à un même espace de Hilbert séparable, $(e_n)_{n \in \mathbb{N}^*}$ (resp. $(f_n)_{n \in \mathbb{N}^*}$) une base hilbertienne de $H'$ (resp. $H''$) et $A \in \mathcal{B}(H' \oplus H'')$ défini pour tout $n \in \mathbb{N}^*$ par $A e_n = 0$ et $A f_n = \frac{1}{n} e_n$. Alors $A$ est nilpotent d’ordre 2 à image $\operatorname{Im} A = A(H'')$ non fermée ($A(H'')$ est dense dans $H'$ mais distinct de $H'$). Cependant, $A$ est une extension continue de l’endomorphisme nul (qui est à image fermée !) de $H'$ par l’endomorphisme nul de $H''$:

$$
A = \begin{bmatrix}
0 & B \\
0 & 0
\end{bmatrix}
$$

où $B$ est un opérateur continu de $H''$ dans $H'$ défini par $B f_n = \frac{1}{n} e_n$.

**Remarque 4.3.** Soient $E$ un espace de Banach complexe, $A \in \mathcal{B}(E)$ un opérateur nilpotent à images des itérés fermées et $F$ un sous-espace de dimension finie dans $E$. Si $F'$ est un sous-espace réduisant pour $A$ minimal
contenant $F$, alors $F'$ est de dimension finie et est unique à isomorphie près (voir K. Benabdallah et B. Charles [1]). Dans le cas où $F$ est cyclique ou lorsque $F$ est contenu dans le noyau de $A$, on démontre que le sous-module $(F, A|F')$ est unique à isomorphie près (voir respectivement [3] et [5]). Le cas où $F$ est quelconque reste cependant à étudier.

**Application.** Soient $A' \in \mathcal{B}(E')$ nilpotent à images des itérés fermées, $A'' \in \mathcal{B}(E'')$ un opérateur nilpotent “de dimension finie” et $A \in \mathcal{B}(E)$ une extension de $A'$ par $A''$. Supposons que $A$ vérifie (10). Soit $F'$ un sous-espace réductant pour $A'$ minimal contenant $A^{n_1}e_1, \ldots, A^{n_q}e_q$ et soit $G \in \text{Lat}A'$ tel que $E' = G' \oplus F'$, alors:

$$(E, A) = (G', A|G') \oplus (F' \oplus E'', T)$$

où $T = A|(F' \oplus E'')$ est une extension de $A|F'$ par $A''$.

Si $A''$ est cyclique (i.e., $q = 1$), alors $(F', A|F') = (F', A|F')$ étant unique à isomorphie près, il en est de même pour $(F' \oplus E'', T)$. Cela entraîne, d’après le Théorème 3.8, que $(G', A|G')$ est aussi unique à isomorphie près. On en déduit que l’étude des extensions d’opérateurs nilpotents à images des itérés fermées par un opérateur nilpotent cyclique se ramène à l’étude en dimension finie des extensions d’opérateurs nilpotents par un opérateur nilpotent cyclique.

En procédant de la même façon, on se ramène aussi au cas de la dimension finie lorsque $q$ est quelconque et que $A^{n_1}e_1, \ldots, A^{n_s}e_q$ sont dans Ker$A'$.

**Remarque 4.4.** Nous nous sommes basés pour la définition des extensions d’opérateurs linéaires sur B. Charles [2].

Je tiens à exprimer ma profonde gratitude au Professeur B. Charles pour ses conseils judicieux et son aide efficace qui m’ont permis l’achèvement de ce travail.

**References**


Received September 16, 1999.

DÉPARTEMENT DE MATHÉMATIQUES ET INFORMATIQUE

UNIVERSITÉ CHOUAÏB DOUKKALI

FACULTÉ DES SCIENCES, B.P.: 20

24000 EL JADIDA

MAROC

E-mail address: faouzi@ucd.ac.ma afaouzi@hotmail.com
CRITICAL-EXPONENT SOBOLEV NORMS AND THE SLICE THEOREM FOR THE QUOTIENT SPACE OF CONNECTIONS

PAUL M.N. FEEHAN

Following Taubes, we describe a collection of critical-exponent Sobolev norms, discuss their embedding and multiplication properties, and describe optimal Green’s operator estimates where the constants depend at most on the first positive eigenvalue of the covariant Laplacian of a $G$ connection and the $L^2$ norm of the connection’s curvature, for arbitrary compact Lie groups $G$. Using these critical-exponent norms, we prove a sharp, global analogue of Uhlenbeck’s Coulomb gauge-fixing theorem, where the usual product connection over a ball is replaced by an arbitrary reference connection over the entire manifold. We also prove a quantitative version of the conventional slice theorem for the quotient space of $G$ connections, with an invariant and sharp characterization of those points in the quotient space which are contained in the image of an $L^4$ ball in the Coulomb-gauge slice. Our gauge-fixing and slice theorems use $L^4_1$ distance functions on the quotient space and the estimate constants depend at most on the first positive eigenvalue of the covariant Laplacian of the reference connection and the $L^2$ norm of its curvature.

1. Introduction.

The use of certain “critical-exponent” Sobolev norms is an important feature of methods employed by Taubes to solve the anti-self-dual and related non-linear elliptic partial differential equations [23], [24], [25]. Indeed, the estimates one can obtain using these critical-exponent norms appear to be the best possible when one needs to bound the norm of a Green’s operator for a Laplacian, depending on a connection varying in a non-compact family, in terms of minimal data such as the first positive eigenvalue of the Laplacian or the $L^2$ norm of the curvature of the connection. Despite their utility, particularly in applications where an optimal analysis is required for gluing or degeneration problems (for example, when considering Uhlenbeck-bubbling families of anti-self-dual connections or PU(2) monopoles), these methods are not widely known. Following Taubes [21], [23], [24], [25] we describe a collection of critical-exponent Sobolev norms and general Green’s operator...
estimates depending only on first positive eigenvalues or the \( L^2 \) norm of the connection’s curvature. These estimates are especially useful both for the construction of gluing maps, in the case of either anti-self-dual connections \([24]\) or, more recently, in the case of PU(2) monopoles \([6], [7], [9]\) and for analyzing their asymptotic behavior with respect to Uhlenbeck limits of the underlying gluing data. We apply them here to prove an optimal slice theorem for the quotient space of connections. The result is “optimal” in the sense that if a point \([A]\) in the quotient space is known to be just \( L^2 \)-close enough to a reference point \([A_0]\) (see below for the precise statement), then \( A \) can be placed in Coulomb gauge relative to \( A_0 \), with all constants depending at most on the first positive eigenvalue of the covariant Laplacian defined by \( A_0 \) and the \( L^2 \) norm of the curvature of \( A_0 \). Such slice theorems are particularly advantageous when analyzing gluing maps and their differentials in situations (such as those of \([8], [9]\)) where the underlying gluing data is allowed to “bubble”. In this paper we shall for simplicity only consider connections over four-dimensional manifolds, but the methods and results can adapted to the case of manifolds of arbitrary dimension, as in \([26]\), to prove slice theorems applicable to cases where the reference connection is allowed to degenerate.

1.1. Critical-exponent Sobolev norms and the slice theorem. Suppose that \( X \) is a closed, Riemannian four-manifold, that \( G \) is a compact Lie group, and that \( \mathcal{B}_{k,p}^E = \mathcal{A}_{k,p}^E / \mathcal{G}_{k+1}^E \) is the quotient space of \( L_k^p \) connections on a \( G \) bundle \( E \) modulo the Banach Lie group of \( L_{k+1}^p \) gauge transformations. Here, the integer \( k \geq 1 \) and the Sobolev exponent \( 1 < p < \infty \) obey the constraint \((k + 1)p > 4\), so \( L_{k+1}^p(X) \subset C^0(X) \) and gauge transformations in \( \mathcal{G}_{k+1}^E \) are continuous. When \((k + 1)p = 4\) we have the “borderline”, “critical”, or “limiting case” of the Sobolev embedding theorem: \( L_{k+1}^p(X) \subset L^q(X) \) for all \( q < \infty \) but not \( q = \infty \).

A connection \( A \in \mathcal{A}_{k,p}^E \) is in Coulomb gauge relative to a reference connection \( A_0 \) if \( d^*_{A_0}(A - A_0) = 0 \) and it is a standard result that \( \mathbf{S}_{A_0} = A_0 + \text{Ker} \; d^*_{A_0} \subset \mathcal{A}_{k,p}^E \) provides a slice for the action of the gauge group \( \mathcal{G}_{k+1}^E \) \([2], [5], [10], [11], [13], [14], [16], [18]\). (See Proposition 3.4 for a detailed statement.) More exactly, if \( \mathbf{B}_{A_0}^{k,p}(\varepsilon) \) is the \( L_k^p \) ball in \( \mathbf{S}_{A_0} \) with center \( A_0 \) and \( L_{k,A_0}^p \)-radius \( \varepsilon \) and \( \text{Stab}_{A_0} \subset \mathcal{G}_{k+1}^E \) is the stabilizer of \( A_0 \), then the projection \( \pi : \mathbf{B}_{A_0}^{k,p}(\varepsilon)/\text{Stab}_{A_0} \to \mathbf{B}_{k,p}^E \) is a homeomorphism onto its image and thus contains a small enough \( L_k^p \) ball

\[
B_{[A_0]}^{k,p}(\eta) = \left\{ [A] \in \mathbf{B}_{k,p}^E : \text{dist}_{L_{k,A_0}^p}([A], [A_0]) < \eta \right\},
\]
where gauge-invariant distance functions on the \( G_E^{k+1,p} \)-quotient are defined by

\[
\text{dist}_{L^p_{k', A_0}} ([A], [A_0]) = \inf_{u \in G_E^{k+1,p}} \| u(A) - A_0 \|_{L^p_{k', A_0}},
\]

whenever \( L^p_k \subset L^p_{k'} \). One unsatisfactory aspect of the standard slice theorem concerns the dependence of the constants \( \epsilon([A_0], k, p) \) and \( \eta([A_0], k, p) \) above on the orbit \([A_0]\)—in particular on the curvature \( F_{A_0} \)—when \( k \) and \( p \) are large enough that gauge transformations in \( G_E^{k+1,p} \) are continuous. Even in the minimal cases, \( k = 1 \) and \( p > 2 \) or \( k = 2 \) and \( p = 2 \), the constants \( \epsilon, \eta \) depend unfavorably on \([A_0]\) when the curvature \( F_{A_0} \) bubbles. This makes it difficult to analyze the asymptotic behavior of Taubes’ gluing maps [20], [22], [23], [24] and their differentials on neighborhoods of points in the Uhlenbeck boundary of the moduli space of anti-self-dual connections, since the balls \( B_{A_0}^{k,p}(\epsilon) \) and \( B_{[A_0]}^{k,p}(\eta) \) tend to shrink as \([A_0]\) approaches an Uhlenbeck-boundary point. For example, if the connection \( A_0 \) is anti-self-dual, then its energy is bounded by a constant depending only on the topology of \( E \) via the Chern-Weil identity [5, §2.1.4], whereas \( \| F_{A_0} \|_{L^p} \) (with \( p > 2 \)) or \( \| F_{A_0} \|_{L^2_{1,A_0}} \) tends to infinity as the curvature of \( A_0 \) bubbles.

Our main purpose in this article is to prove a global analogue, Theorem 1.1, of Uhlenbeck’s local Coulomb gauge-fixing theorem [26, Theorems 1.3 & 2.1] and a corresponding slice theorem, Theorem 1.2, where the radii of the coordinate balls on the quotient \( B_{E}^{k,p} \) depend only on \( \| F_{A_0} \|_{L^2} \) and the least positive eigenvalue \( \nu_0[A_0] \) of the Laplacian \( d_{A_0}^* d_{A_0} \) on \( \Omega^0(\mathfrak{g}_E) \). The key difficulty in establishing Theorem 1.1 is to ensure that the constants depend at most on \( \| F_{A_0} \|_{L^2} \) and \( \nu_0[A_0] \); To guarantee this minimal dependence, we employ critical-exponent Sobolev norms (defined below) to circumvent the fact that when \( (k+1)p = 4 \) the standard Sobolev embedding and multiplication theorems fall just short of what one needs to give the quotient \( B_{E}^{k,p} = A_{E}^{k,p}/G_E^{k+1,p} \) a manifold structure (see Section 4). Such norms were introduced by Taubes for related purposes in [23].

1.2. Statement of results. For clarity, we now fix \( p = 2 \) and \( k \geq 2 \) and define the following distance functions on the quotient space \( B_{E}^{k} = A_{E}^{k}/G_E^{k+1} \) of \( L^2_k \) connections modulo \( L^2_{k+1} \) gauge transformations,

\[
\text{dist}_{L^2_{1,A_0}} ([A], [A_0]) = \inf_{u \in G_E^{k+1}} \left( \| u(A) - A_0 \|_{L^2_{1,A_0}} + \| d_{A_0}^* (u(A) - A_0) \|_{L^2} \right),
\]

\[
\text{dist}_{L^2_{1,A_0}} ([A], [A_0]) = \inf_{u \in G_E^{k+1}} \left( \| u(A) - A_0 \|_{L^2_{1,A_0}} + \| d_{A_0}^* (u(A) - A_0) \|_{L^2} \right),
\]

for \( k \geq 2 \).
where the norms of \(a \in \Omega^1(g_E)\) are defined by (see Equations (4.1), (4.2), and (4.3)),

\[
\|a\|_{L^2(X)} = \sup_{x \in X} \|\text{dist}^{-2}(x, \cdot) |a||_{L^1(X)},
\]

\[
\|a\|_{L^2(X)} = \sup_{x \in X} \|\text{dist}^{-1}(x, \cdot)|a|\|_{L^2(X)},
\]

\[
\|a\|_{L^4(X)} = \|a\|_{L^2(X)} + \|a\|_{L^2(X)},
\]

\[
\|a\|_{L^4,2(X)} = \|a\|_{L^2(X)} + \|a\|_{L^2(X)},
\]

\[
\|a\|_{L^2,2,0}^2(X) = \left(\|a\|_{L^2(X)}^2 + \|\nabla_{A_0} a\|_{L^2(X)}^2\right)^{1/2}.
\]

Here, \(\text{dist}(x, y)\) denotes the geodesic distance between points \(x, y \in X\). The distance function \(\text{dist}_{L^1,2,0}([A], [A_0])\) is bounded by scale invariant norms,

\[
\|a\|_{L^2(X)} + \|\nabla^g a\|_{L^2(X)} + \sup_{x \in X} \|\text{dist}^{-2}(x, \cdot)|d^g_{A_0} a||_{L^1(X)} < \varepsilon,
\]

since the \(L^{4/\ell}\) norm on \(\otimes^{\ell}(T^*X)\) is conformally invariant, while the third term is invariant under constant rescalings \(g \mapsto \tilde{g} = \lambda^{-2}g\) of the metric, as \(d^g_{A_0} a = \lambda^2 d^g_{A_0} a\), \(\text{dist}^{-2}g(x, y) = \lambda^2 \text{dist}^{-2}g(x, y)\) and \(dV_{\tilde{g}} = \lambda^{-4}dV_g\). Similarly for \(\text{dist}_{L^1,2,0}([A], [A_0])\). Like the \(L^4\) norm, the \(L^{2\ell}\) norm on one-forms is scale-invariant. Our first result is the following global analogue of Uhlenbeck’s theorem and complements results of Taubes in [23, §6]:

**Theorem 1.1.** Let \(X\) be a closed, smooth four-manifold with metric \(g\) and let \(G\) be a compact Lie group. Then there are positive constants \(c, \varepsilon\) with the following significance. Let \(E\) be a \(G\) bundle over \(X\) and suppose that \(k \geq 2\) is an integer. Given a point \([A_0]\) in \(\mathcal{B}_E^k\), let \(\nu_0[A_0]\) be the least positive eigenvalue of the Laplacian \(d^*_{A_0} d_{A_0}\) on \(\Omega^0(g_E)\) and set \(K_0 = (1 + \nu_0[A_0]^{-1}) (1 + \|F_{A_0}\|_{L^2}).\) Let \(\varepsilon_1\) be a constant satisfying \(0 < \varepsilon_1 \leq zK_0^{-2}(1 + \nu_0[A_0]^{-1/2})^{-1}\). Then:

1. For any \([A] \in \mathcal{B}_E^k\) with \(\text{dist}_{L^1,2,0}([A], [A_0]) < \varepsilon_1\), there is a gauge transformation \(u \in G_{E}^{k+1}\), unique up to an element of the stabilizer \(\text{Stab}_{A_0} \subset G_{E}^{k+1}\), such that:
   (a) \(d^*_{A_0} (u(A) - A_0) = 0\),
   (b) \(\|u(A) - A_0\|_{L^{2,4}} \leq cK_0 \text{dist}_{L^1,2,0}([A], [A_0])\).

2. For any \([A] \in \mathcal{B}_E^k\) with \(\text{dist}_{L^1,2,0}([A], [A_0]) < \varepsilon_1\), there is a gauge transformation \(u \in G_{E}^{k+1}\), unique up to an element of the stabilizer \(\text{Stab}_{A_0} \subset G_{E}^{k+1}\), such that:
   (a) \(d^*_{A_0} (u(A) - A_0) = 0\),
   (b) \(\|u(A) - A_0\|_{L^{2,4}} \leq cK_0 \text{dist}_{L^1,2,0}([A], [A_0])\).
(c) \( \|u(A) - A_0\|_{L^2_{1,A_0}} \leq cK_0 \text{dist}_{L^1_{1,A_0}}([A], [A_0]) \).

In Theorem 2.1 of [26] the \( L^2 \) norm of the curvature \( F_A \) of a local connection matrix \( A \) over the unit ball in \( \mathbb{R}^4 \) provides a natural (gauge-invariant) measure of the distance from \([A]\) to \([\Gamma]\), where \( \Gamma \) is the product connection. Uhlenbeck’s theorem guarantees the existence of an \( L^p_{k+1} \) gauge transformation \( u \) taking an \( L^p_{k} \) connection \( A \) on the product bundle over the unit four-ball, with product connection \( \Gamma \), to a connection \( u \) taking an \( L^p_{k} \) connection \( A \) on the product bundle over the unit four-ball, with product connection \( \Gamma \), to a connection \( u(A) - \Gamma = 0 \) and \( \|u(A) - \Gamma\|_{L^2_{4}} \leq c\|F_A\|_{L^2} \); one only requires that \( \|F_A\|_{L^2} \) be smaller than a universal constant.

We next have the following refinement of the standard slice theorem for the quotient space \( \mathcal{B}^k_E \). The observation that an \( L^4 \)-ball in \( \text{Ker} \, d_{A_0}^* \) provides a slice for \( G_{E}^{k+1} \) was pointed out to us Mrowka; that slightly smaller \( L^{2e+4} \) and \( L^2_{1,A_0} \) balls provide slices follows from the second of our two proofs of Theorem 1.1 in Section 8. For any \( \epsilon > 0 \), we define open balls

\[
B_{[A_0]}^{1,*,2}(\epsilon) = \left\{ [A] \in \mathcal{B}^k_E : \text{dist}_{L^2_{1,A_0}}([A], [A_0]) < \epsilon \right\} \subset \mathcal{B}^k_E,
\]

\[
B_{[A_0]}^{1,\pm,2}(\epsilon) = \left\{ [A] \in \mathcal{B}^k_E : \text{dist}_{L^2_{1,A_0}}([A], [A_0]) < \epsilon \right\} \subset \mathcal{B}^k_E,
\]

\[
B^4_{A_0}(\epsilon) = \left\{ A \in \mathcal{A}^k_E : d_{A_0}^*(A - A_0) = 0 \text{ and } \|A - A_0\|_{L^4(X)} < \epsilon \right\} \subset \mathcal{A}^k_{A_0},
\]

where \( S_{A_0} = \{A_0\} + \text{Ker}(d_{A_0}^*|_{L^2_{4}}) \subset \mathcal{A}^k_E \) is the slice through \( A_0 \). We let \( B_{[A_0]}^{1,\pm,2}(\epsilon) \) and \( B_{[A_0]}^{1,*,2}(\epsilon) \) denote the closed balls.

**Theorem 1.2.** Let \( X \) be a closed, smooth four-manifold with metric \( g \) and let \( G \) be a compact Lie group. Then there are positive constants \( c_1, c_2, z \) with the following significance. Let \( E \) be a \( G \) bundle over \( X \), let \( k \geq 2 \) be an integer, and suppose that \([A_0]\) \( \in \mathcal{B}^k_E \). Then:

1. For \( \varepsilon_0 \) such that \( 0 < \varepsilon_0 < z(1 + \nu_0[A_0]^{-1/2})^{-1} \), the projection \( \pi : \mathcal{B}^4_{A_0}(\varepsilon_0)/\text{Stab}_{A_0} \rightarrow \mathcal{B}^k_E \), \( A \mapsto [A] \), is a homeomorphism onto an open neighborhood of \([A_0]\) \( \subset \mathcal{B}^k_E \) and a diffeomorphism on the open subset where \( \text{Stab}_{A_0}/\text{Center}(G) \) acts freely;

2. For any constant \( \varepsilon_1 \) satisfying \( 0 < \varepsilon_1 \leq zK_0^{-2}(1 + \nu_0[A_0]^{-1/2})^{-1} \) we have the following inclusions of open neighborhoods in \( \mathcal{B}^k_E \):

\[
B_{[A_0]}^{1,\pm,2}(\varepsilon_1) \subset B_{[A_0]}^{1,*,2}(c_1\varepsilon_1) \subset \pi(B^4_{A_0}(c_2K_0\varepsilon_1)).
\]

That sharper versions of the standard slice theorem (as in [5], [10], [11], for example) would hold is suggested by related results of Taubes, namely [21, Lemma A.1] and [23, Lemma 6.5]: For example, they show that if \( u \) is an \( L^2_{2} \) gauge transformation intertwining \( L^2_{1} \) connections \( A_i, i = 1, 2, \)
obeying a slice condition \( d^*_A(A_i - A_0) = 0 \) defined by an \( L^2 \) connection \( A_0 \), then \( u \) is necessarily in \( C^0 \). Moreover, transition functions relating neighborhoods of the origin in \( \text{Ker}(d^*_A|_{L^2}) \) and \( \text{Ker}(d^*_A + a|_{L^2}) \), where \( a \) is \( L^2 \)-small, are constructed in [23, Lemma 6.5]; the constants depend only on \( \|F_{A_0}\|_{L^2} \) and \( \nu_0[A_0] \). (See [23, §6] for detailed statements and related results.) The proof of Theorem 1.1 makes use of methods developed in [21], [23], [24]. To illustrate applications of the methods of Sections 4 and 5 and to point to possible generalizations of the estimates in this article, we derive some elliptic estimates for the linearization of the anti-self-dual equation in Section 5.2.

1.3. Outline of the proofs. Assertion (1) of Theorem 1.2 is proved in Section 3. The proof that the projection map \( \pi : B^4_{A_0}(\varepsilon_0) \to B^k_E \) is a local diffeomorphism away from connections with non-minimal stabilizer essentially follows Uhlenbeck’s verification of “openness” in her proof of Theorem 2.1 in [26] via the method of continuity (see Lemma 3.6). The proof that the \( L^4 \) ball \( B^4_{A_0}(\varepsilon_0) \) injects into the quotient (see Lemma 3.7) was suggested to us by Mrowka. The remainder of our article is taken up with the proof of Theorem 1.1 and hence Assertion (2) of Theorem 1.2.

In Section 4 we introduce the family of critical-exponent Sobolev norms, \( L^k_{k,A_0} \), \( k = 0, 1, 2 \), used to complete the proof of Theorem 1.1 and in Section 5 we describe the crucial embedding theorems enjoyed by those Sobolev spaces, as well as estimates for the Green’s operator of the Laplacian \( d^*_A d_A \). In particular, \( L^p_{k,A_0} \subset L^k_{k,A_0} \), for every \( p > 2 \) while, in the other direction, \( L^k_{k,A_0} \subset C^0 \). The latter embedding is the key motivation for the definition of these norms and it greatly facilitates the derivation of Green’s operator estimates, in a wide number of applications in gauge theory [23], [24], with minimal dependence on the curvature of the connection \( A_0 \). The main ideas and embedding results in Sections 4 and 5 are due to Taubes [21], [23], [24], [25], so these sections are essentially expository. An earlier exposition from a somewhat different perspective, due to Donaldson, of Taubes’ methods and some applications appears in [4]. The estimates of Section 5 are stated only in the four-dimensional case. While we might expect all of them to hold, in some form, for higher dimensions we confine our attention to dimension four as our intended applications are primarily concerned with smooth four-manifold topology. In essence, the critical-exponent norms make a virtue out of necessity of the familiar fact that while the Green’s operator of the Laplacian \( d^* d \) on \( C^\infty(X) \) maps \( L^p(X) \) into \( L^{2p/(2-p)}(X) \) for \( 1 < p < 2 \), it does not map \( L^2(X) \) into \( L^\infty(X) \) [19, Chapter V]. We recall that an Orlicz space \( L_\varphi \) can be used to provide the “best target space” for an embedding of \( L^2_2(X) \) [1, Chapter 8]. Here, we may instead view \( L^2_{1,2}(X) \) as providing
the “best domain space” for an embedding into $L^\infty(X)$, since $L^p_0(X) \subset L^{\frac{p}{2}}_2(X) \subset L^\infty(X)$ for all $p > 2$.

We give two proofs of Theorem 1.1. For our first proof, in Section 6, we essentially follow the strategy of Uhlenbeck [26] and apply the method of continuity. The difficult step here (in establishing openness—see Section 6.3) is to prove that the intrinsic, gauge-invariant continuity. The difficult step here (in establishing openness—see Section 6.3) is to prove that the intrinsic, gauge-invariant $L^{\frac{p}{2}}_2$ and $L^{\frac{p}{2}}_2$ distances in the quotient $B_E^k$ bound the $L^{2,4}$ and $L^2_{1,A_0}$ norms in the slice $S_{A_0} \subset A^k_E$: This is the point in our first proof where we use the critical-exponent estimates derived in Section 5 to control gauge transformations. The proof of “closedness” uses a compactness argument and is given in Section 6.2. The proofs of Theorems 1.1 and 1.2 are completed in Section 6.3.

Our second proof of Theorem 1.1 occupies Sections 7 and 8. In Section 7 we show that the exponential map $\text{Exp} : \Omega^0(\frak{g}_E) \to G_E$ extends to a continuous map $\text{Exp} : L^{\frac{p}{2},2}_2(\frak{g}_E) \to L^{\frac{p}{2},2}_2(\frak{g}_E)$ and that the resulting space of $L^{\frac{p}{2},2}$-gauge transformations $G_E^{2,p}$ is a Banach Lie group. In particular, $L^{\frac{p}{2},2}$-gauge transformations are continuous and are contained in $G_E^{2,p}$ for every $p > 2$. The Sobolev multiplication and composition results for the critical-exponent norms then allow us to apply the inverse function theorem directly in Section 8, while still ensuring that all constants depend at most on $\nu_0[A_0]$ and $\|F_{A_0}\|_{L^2}$. We first use the compactness result of Section 6.1 to establish the existence of gauge transformations $w$ in $G_E^3$ which minimize the $L^{\frac{p}{2},2}_{1,A_0}$ and $L^{\frac{p}{2},2}_{1,A_0}$ distances in the quotient $B_E^k$. Then, assuming the norm $\|w(A) - A_0\|_{L^{2,4}}$ or $\|w(A) - A_0\|_{L^{2,4}_{1,A_0}}$ is sufficiently small, we use the Sobolev embedding and multiplication theorems of Sections 4, 5, and 7 and a quantitative version of the inverse function theorem to prove the existence of a gauge transformation $v \in G_E^3$ such that $d_{A_0}^1(u(A) - A_0) = 0$, $u = vw \in G_E^{k+1}$, and $\|u(A) - A_0\|_{L^{2,4}}$ and $\|u(A) - A_0\|_{L^{2,4}_{1,A_0}}$ are controlled by $\text{dist}_{L^{\frac{p}{2},2}_{1,A_0}}([A],[A_0])$ and $\text{dist}_{L^{\frac{p}{2},2}_{1,A_0}}([A],[A_0])$, respectively.

Acknowledgments. We warmly thank Tom Mrowka for helpful and stimulating discussions on gauge theory and his suggestion that $L^4$-balls inject into the quotient, and warmly thank Cliff Taubes for very helpful discussions and especially for explaining his ideas and methods in [21], [23]. We would also like to thank the Harvard Mathematics Department, where this article was completed in September 1996, and the National Science Foundation for their generous hospitality and support during the preparation of this article.

2. Preliminaries.

We assume throughout this article that $X$ is a closed, connected, smooth, four-manifold with Riemannian metric $g$. Let $G$ be a compact Lie group with matrix representation $\rho : G \subset \text{SO}(\mathbb{E}) = \text{SO}(r)$ where $\mathbb{E} \simeq \mathbb{R}^r$ as a real
inner product space, let $P$ be a principal $G$ bundle, and let $E = P \times_{\rho} \mathbb{E}$ be the corresponding Riemannian vector bundle associated to $P$ by the representation $\rho$. Let $\mathfrak{g}_E \subset \mathfrak{gl}(E)$ be the bundle of Lie algebras associated to $P$ via the adjoint representation $\text{Ad} : G \to \text{Aut}(\mathfrak{g})$ of $G$ on its Lie algebra $\mathfrak{g}$ and viewed as a subbundle of $\mathfrak{gl}(E)$ via the induced representation $\rho_* : \mathfrak{g} \subset \mathfrak{so}(\mathbb{E})$.

Given the covariant derivative $\nabla_A : C^\infty(E) \to C^\infty(T^*X \otimes E)$, we define the exterior covariant derivative $d_A : \Omega^i(E) \to \Omega^{i+1}(E)$ in the usual way by setting $d_A = \nabla_A$ on $\Omega^0(E) = C^\infty(E)$ and extending $d_A$ to $\Omega^i(E) = C^\infty(\Lambda^i \otimes E)$, where $\Lambda^i := \Lambda^i(T^*X)$, according to the rule $d_A(\omega \wedge v) = d\omega \wedge v + (-1)^i \omega \wedge d_A v$ for $\omega \in \Omega^i(X)$ and $v \in \Omega^j(E)$.

For any integer $k \geq 0$, exponent $1 \leq p \leq \infty$, and $L^p_k$ connection $A_0$ on $E$ we define the $L^p_k$ Sobolev completion, $L^p_k(\Lambda^i \otimes E)$, of $\Omega^i(E)$ with respect to the norm

$$\|s\|_{L^p_k,A_0}(X) := \left( \sum_{j=0}^{k} \|\nabla_{A_0}^j s\|_{L^p(X)}^p \right)^{1/p}.$$ 

We define the action of a $C^\infty$ gauge transformation $u \in \mathcal{G}_E$ on a $C^\infty$ connection $A$ on the bundle $E$ by pushforward, so $u(A) := A - (d_A u) u^{-1}$. Fix a connection $A_0 \in \mathcal{A}_E$, let $\mathcal{A}^k_E = A_0 + L^2_k(\Lambda^1 \otimes \mathfrak{g}_E)$, and define

$$\mathcal{G}^{k+1}_E := \{ u \in L^2_{k+1}(\mathfrak{gl}(E)) : u \in G \text{ a.e.} \} \subset L^2_{k+1}(\mathfrak{gl}(E)).$$

The space $\mathcal{G}^{k+1}_E$ is a Banach Lie group, with Lie algebra $T_1 \mathcal{G}^{k+1}_E \mathcal{G}^{k+1}_E = L^2_{k+1}(\mathfrak{g}_E)$, and acts smoothly on $\mathcal{A}^k_E$ with quotient $\mathcal{B}^k_E := \mathcal{A}^k_E / \mathcal{G}^{k+1}_E$ endowed with the quotient $L^2_k$ topology.

The stabilizer subgroup $\text{Stab}_A \subset \mathcal{G}^{k+1}_E$ for a connection $A$ on $E$ always contains the center $\text{Center}(G) \subset G$. We let $\mathcal{A}^{*,k}_E \subset \mathcal{A}^k_E$ denote the space of connections $A \in \mathcal{A}^k_E$ with minimal stabilizer $\text{Stab}_A = \text{Center}(G)$ and let $\mathcal{B}^{*,k}_E = \mathcal{A}^{*,k}_E / \mathcal{G}^{k+1}_E$. As usual, the stabilizer subgroup $\text{Stab}_A \subset \mathcal{G}_E$ can be identified with a closed subgroup of $G \subset \text{GL}(E|_x)$ for any point $x_0 \in X$ by parallel translation with respect to the connection $A$. Let $\text{stab}_A$ denote the Lie algebra of $\text{Stab}_A$, so $\text{stab}_A = \text{Ker}\{ d_A : L^2_{k+1}(\mathfrak{g}_E) \to L^2_k(\Lambda^1 \otimes \mathfrak{g}_E) \}$.

Throughout the article, we use $c$ or $z$ to denote positive constants which depend at most on the Riemannian manifold $(X, g)$ and the group $G$; constants may increase from one line to the next and are not renamed unless clarity demands otherwise.

3. The slice theorem.

In this section we prove the first assertion of Theorem 1.2—see Proposition 3.4 below—namely, that a small enough $L^4$-ball $\mathcal{B}^4_{A_0}(\varepsilon_0)/\text{Stab}_{A_0}$ provides a slice for the action of $\mathcal{G}^{k+1}_E$. The proof that the projection $\pi :$
Let \( \text{Stab}_\perp \rightarrow \mathcal{B}^k_E \) be injective (Lemma 3.7) was suggested to us by Mrowka.

Let \( k \geq 2 \) be an integer. The Banach Lie group \( \mathcal{G}^{k+1}_E \) has Lie algebra \( T_{i\mathbb{R}} \mathcal{G}^{k+1}_E = L^2_{k+1}(\mathfrak{g}_E) \) and exponential map \( \text{Exp} : L^2_{k+1}(\mathfrak{g}_E) \rightarrow \mathcal{G}^{k+1}_E \) given by \( \zeta \mapsto u = \text{Exp} \zeta \). Recall that \( \text{Stab}_A = \{ \gamma \in \mathcal{G}^{k+1}_E : \gamma(A) = A \} \) may be identified with a Lie subgroup of \( G \) and has Lie algebra \( \text{stab}_A = \ker(d_A|_{L^2_{k+1}}) \). The operator \( d_A^* : L^2_{k+2}(\Lambda^1 \otimes \mathfrak{g}_E) \rightarrow L^2_{k+1}(\mathfrak{g}_E) \) has closed range and we have an \( L^2 \)-orthogonal decomposition

\[
T_{i\mathbb{R}} \mathcal{G}^{k+1}_E = L^2_{k+1}(\mathfrak{g}_E)
\]

\[
= (\ker(d_A|_{L^2_{k+1}})) \perp \ker(d_A|_{L^2_{k+1}})
\]

\[
= \text{Im} \left( d_A^*|_{L^2_{k+2}} \right) \perp \ker \left( d_A|_{L^2_{k+1}} \right)
\]

\[
= \left( \ker \left( d_A|_{L^2_{k+1}} \right) \right) \perp \text{stab}_A.
\]

Let \( \text{Stab}_A^\perp = L^2_{k+1} \cap \text{Stab}_A = \text{Exp}((\ker d_A|_{L^2_{k+1}})^+) \), the second equality following from the Sobolev composition lemma. The subspace \( \text{Stab}_A^\perp \subset \mathcal{G}^{k+1}_E \) is closed and is a Banach submanifold of \( \mathcal{G}^{k+1}_E \) with codimension \( \dim \text{stab}_A \).

From Claim 3.5 below we see that \( \text{Stab}_A^\perp \) is a slice near \( i\mathbb{R} \in \mathcal{G}^{k+1}_E \) for the right action of \( \text{Stab}_A \) on \( \mathcal{G}^{k+1}_E \).

The map \( d_A : L^2_{k+1}(\mathfrak{g}_E) \rightarrow L^2_k(\Lambda^1 \otimes \mathfrak{g}_E) \) has closed range and so we have an \( L^2 \)-orthogonal decomposition

\[
T_A \mathcal{A}^k_E = L^2_k(\Lambda^1 \otimes \mathfrak{g}_E)
\]

\[
= \text{Im} \left( d_A|_{L^2_{k+1}} \right) \perp \ker \left( d_A^*|_{L^2_k} \right)
\]

\[
= \text{Im} \left( d_A|_{L^2_{k+1}} \right) \perp \text{K}_A,
\]

of the tangent space to the space of \( L^2_k \) connections at \( A \), where \( \text{K}_A = \ker(d_A^*|_{L^2_k}) \).

The slice \( S_A \subset \mathcal{A}^k_E \) through a connection \( A \) is given by \( S_A = A + \text{K}_A \). If \( \pi \) is the projection from \( \mathcal{A}^k_E \) onto \( \mathcal{B}^k_E = \mathcal{A}^k_E / \mathcal{G}^{k+1}_E \), denoted by \( A \mapsto [A] \), we let

\[
\mathcal{B}_A(\epsilon) = \{ A_1 \in S_A : \| A_1 - A \|_{L^2_{k,A}} < \epsilon \}
\]

\[
= A + \{ a \in \text{K}_A : \| a \|_{L^2_{k,A}} < \epsilon \}
\]

be the open \( L^2_k \)-ball in \( S_A \) with center \( A \) and \( L^2_{k,A} \)-radius \( \epsilon \). Similarly, we let

\[
\mathcal{B}^A_\epsilon(\epsilon) = \{ A_1 \in S_A : \| A_1 - A \|_{L^4} < \epsilon \}
\]

\[
= A + \{ a \in \text{K}_A : \| a \|_{L^4} < \epsilon \}
\]

be the open \( L^4 \)-ball in \( S_A \) with center \( A \) and \( L^4 \)-radius \( \epsilon \).
be the open ball in $S_A$ with center $A$ and $L^1$-radius $\varepsilon$.

The proof that the quotient space $A^k_E$ is Hausdorff makes use of the following well-known technical result [10, Proposition A.5]. Note that the space $G^2_E$ is neither a Banach Lie group nor does it act smoothly on $A^k_E$ for $k \geq 1$.

**Lemma 3.1.** Let $E$ be a Hermitian bundle over a Riemannian manifold $X$ and let $k \geq 2$ be an integer. Suppose $\{A_\alpha\}$ and $\{B_\alpha\}$ are sequences of $L^1$ unitary connections on $E$ and that $\{u_\alpha\}$ is a sequence in $G^2_E$ such that $u_\alpha(A_\alpha) = B_\alpha$. Then the following hold.

1. The sequence $\{u_\alpha\}$ is in $G^{k+1}_E$;
2. If $\{A_\alpha\}$ and $\{B_\alpha\}$ converge in $A^k_E$ to limits $A_\infty$, $B_\infty$, then there is a subsequence $\{\alpha^{'}\} \subset \{\alpha\}$ such that $\{u_{\alpha^{'}}\}$ converges in $G^{k+1}_E$ to $u_\infty$ and $B_\infty = u_\infty(A_\infty)$.

We shall need the following quantitative version of the inverse function theorem here and especially in Section 8:

**Theorem 3.2.** Let $\Phi : E \to F$ be a $C^\ell$ map of Banach spaces, for some $\ell \geq 1$, such that the differential $(D\Phi)_{x_0} : E \to F$ has a continuous inverse $(D\Phi)^{-1}_{x_0} : F \to E$ satisfying

$$
\| (D\Phi)^{-1}_{x_0} \| \leq K \quad \text{and} \quad \| (D\Phi)x - (D\Phi)_{x_0} \| \leq \frac{1}{2} K^{-1}, \quad \text{if} \quad \| x - x_0 \| \leq \delta,
$$

for some positive constants $K$ and $\delta$. Then the following hold:

1. The restriction of $\Phi$ to the ball $U = B^E(x_0, \delta)$ is injective and $\Phi(U) = V$ is an open set in $F$ containing the ball $B^F(\Phi(x_0), \delta/(2K));$

2. The inverse map $\Phi^{-1} : V \to U$ is $C^\ell$;

3. If $x_1, x_2 \in B^E(x_0, \delta)$, then $\| x_1 - x_2 \| \leq 2K\| \Phi(x_1) - \Phi(x_2) \|$.

For quantitative comparisons in this section, the following elementary fact will suffice:

**Lemma 3.3.** Let $E$, $F$ be Banach spaces and let $T \in \text{Hom}(E,F)$ have a right (left) inverse $S$. If $\tilde{T} \in \text{Hom}(E,F)$ satisfies $\| \tilde{T} - T \| < \| S \|^{-1}$, then $\tilde{T}$ also has a right (left) inverse.

**Proof.** If $S \in \text{Hom}(F,E)$ is a right inverse for $T$, so $TS = \text{id}_F$, then $\| (\tilde{T} - T)S \| \leq \| \tilde{T} - T \| \| S \| < 1$ and $\text{id}_E + (\tilde{T} - T)S$ is an invertible element of the Banach algebra $\text{End}(E)$. Define $\tilde{S} = S(1 + (\tilde{T} - T)S)^{-1}$, so $\tilde{S}T = \text{id}_E$ and $\tilde{S}$ is a right inverse for $\tilde{T}$. Similarly for left inverses. \hfill $\square$

This consequence of the usual characterization of invertible elements of a Banach algebra will be invoked in the proof of Lemma 3.6.

**Proposition 3.4.** Let $X$ be a closed, Riemannian four-manifold. Then there is a positive constant $z$ with the following significance. Let $E$ be a
G bundle over X. Suppose that \( k \geq 2 \) is an integer. Given \([A_0]\) in \( \mathcal{B}_E^k \), let \( \nu_0[A_0] \) be the least positive eigenvalue of the Laplacian \( \Delta^0_{A_0} \) and let \( \varepsilon_0 \) be a constant satisfying 0 < \( \varepsilon_0 < z(1 + \nu_0[A_0]^{-1/2})^{-1} \). Then:

1. The space \( \mathcal{B}_E^k \) is Hausdorff;
2. The subspace \( \mathcal{B}_E^{*,k} \subset \mathcal{B}_E^k \) is open and is a \( C^\infty \) Banach manifold with local parameterizations given by \( \pi : \mathcal{B}^1_{A_0}(\varepsilon_0) \to \mathcal{B}_E^{*,k} \);
3. The projection \( \pi : \mathcal{A}_E^{*,k} \to \mathcal{B}_E^{*,k} \) is a \( C^\infty \) principal \( \mathcal{G}_{E}^{k+1} / \text{Center}(G) \) bundle;
4. The projection \( \pi : \mathcal{B}^1_{A_0}(\varepsilon_0) / \text{Stab}_{A_0} \to \mathcal{B}_E^k \) is a homeomorphism onto an open neighborhood of \([A_0] \in \mathcal{B}_E^k \) and a diffeomorphism on the subset where \( \text{Stab}_{A_0} / \text{Center}(G) \) acts freely.

**Proof.** The stabilizer \( \text{Stab}_{A_0} \) acts freely on \( \mathcal{G}_{E}^{k+1} \) and thus on the Banach manifold \( \mathcal{G}_{E}^{k+1} \times \mathcal{S}^k_{A_0} \) by \( \gamma \cdot (u, A) = (u\gamma^{-1}, \gamma(A)) \) and so the quotient \( \mathcal{G}_{E}^{k+1} / \text{Stab}_{A_0} \mathcal{S}^k_{A_0} \) is again a Banach manifold. We define a smooth map

\[
\Psi : \mathcal{G}_{E}^{k+1} \times \text{Stab}_{A_0} \mathcal{S}^k_{A_0} \to \mathcal{A}^k_{E}, \quad [u, A] \mapsto u(A).
\]

Our main task is to show that the map \( \Psi \) is (i) a local diffeomorphism onto its image and (ii) injective upon restriction to a sufficiently small neighborhood \( \mathcal{G}_{E}^{k+1} / \text{Stab}_{A_0} \mathcal{B}^1_{A_0}(\varepsilon_0) \). Given \( \delta_0 > 0 \), let \( \mathcal{B}_{\text{id}_E}(\delta_0) \) be the ball \( \{u \in \mathcal{G}_{E}^{k+1} : \|u - \text{id}_E\|_{L^2_{k+1,A_0}} < \delta_0 \} \) and let \( \mathcal{B}^1_{\text{id}_E}(\delta_0) = \mathcal{B}_{\text{id}_E}(\delta_0) \cap \text{Stab}_{A_0}^1 \).

**Claim 3.5.** For small enough \( \delta = \delta(A_0, k) \), the ball \( \mathcal{B}_{\text{id}_E}(\delta) \) is diffeomorphic to an open neighborhood in \( \mathcal{B}^1_{\text{id}_E}(\delta) \times \text{Stab}_{A_0} \), with inverse map given by \((u_0, \gamma) \mapsto u = u_0\gamma\).

**Proof.** The differential of the multiplication map

\[
\text{Stab}_{A_0}^1 \times \text{Stab}_{A_0} \to \mathcal{G}_{E}^{k+1}, \quad (u_0, \gamma) \mapsto u_0\gamma,
\]

at \((\text{id}_E, \text{id}_E)\) is given by

\[
\text{Ker} \left( \frac{d_{A_0|L^2_{k+1}}} \right) \oplus \text{stab}_{A_0} \to L^2_{k+1}(\mathfrak{g}_E) \quad (\zeta, \chi) \mapsto u_0\zeta\gamma + u_0\gamma\chi,
\]

and so is just the identity map with respect to the \( L^2 \)-orthogonal decomposition (3.1) of the range. Hence, the Banach space implicit function theorem implies that there is a diffeomorphism from an open neighborhood of \((\text{id}_E, \text{id}_E)\) onto an open neighborhood of \( \text{Stab}_{A_0}^1 \). For small enough \( \delta \), we may suppose that if \( u \in \mathcal{B}_{\text{id}_E}(\delta) \), then \( u \) can be written uniquely as \( u = u_0\gamma \) with \( u_0 \in \mathcal{B}^1_{\text{id}_E}(\delta) \) and \( \gamma \in \text{Stab}_{A_0} \).

**Lemma 3.6.** For any \( 0 < \varepsilon_0 < \frac{1}{2}(1 + \nu_0[A_0]^{-1/2})^{-1} \), the map \( \Psi \) is a local diffeomorphism from \( \mathcal{G}_{E}^{k+1} \times \text{Stab}_{A_0} \mathcal{B}^1_{A_0}(\varepsilon_0) \) onto its image in \( \mathcal{A}^k_{E} \).
Proof. We first restrict the map \( \Psi \) to a neighborhood \( B_{id_E}(\delta_0) \times \text{Stab}_{A_0} S_{A_0} \), which is diffeomorphic to the neighborhood \( B^1_{id_E}(\delta) \times S_{A_0} \) in \( \text{Stab}_{A_0} \times S_{A_0} \) by Claim 3.5. The differential of the induced map
\[
(3.4) \quad \Psi : \text{Stab}_{A_0} \times S_{A_0} \to A^k_E, \quad (u, A) \mapsto u(A),
\]
at \((id_E, A) := (id_E, A_0 + a_0)\) is given by
\[
(D\Psi)_{(id_E,A)} : T_{id_E} \text{Stab}_{A_0} \oplus T_{A} S_{A_0} \to T_{A} A^k_E,
\]
\[
(\zeta, a) \mapsto -d_A \zeta + a = -d_A \zeta - [a_0, \zeta] + a,
\]
where we recall that \( T_{A} S_{A_0} = K_{A_0} = \text{Ker}(d^*_{A_0}|_{L^2_{A_0}}) \) and
\[
T_{id_E} \text{Stab}_{A_0} = \left( \text{Ker}\left(d_{A_0}|_{L^2_{k+1}}\right) \right)^\perp = \text{Im}\left( d^*_{A_0}|_{L^2_{k+2}}\right).
\]
Using the \( L^2 \)-orthogonal decomposition (3.2) of the range we see that the map
\[
-d_{A_0} \oplus id_E : \left( \text{Ker}\left(d_{A_0}|_{L^2_{k+1}}\right) \right)^\perp \oplus \text{Ker}\left(d^*_{A_0}|_{L^2_{k+1}}\right) \to \text{Im}\left( d_{A_0}|_{L^2_{k+1}}\right) \oplus \text{Ker}\left(d^*_{A_0}|_{L^2_{k+1}}\right)
\]
given by \((\zeta, b) \mapsto -d_{A_0} \zeta + b\) is a Hilbert space isomorphism. More explicitly, the operator
\[
d_{A_0} : \left( \text{Ker}\left(d_{A_0}|_{L^2_{k+1}}\right) \right)^\perp \to \text{Im}\left( d_{A_0}|_{L^2_{k+1}}\right) = \left( \text{Ker}\left(d^*_{A_0}|_{L^2_{k+1}}\right) \right)^\perp
\]
has a two-sided inverse
\[
G^0_{A_0} d^*_{A_0} : \text{Im}\left( d_{A_0}|_{L^2_{k+1}}\right) \to \left( \text{Ker}\left(d_{A_0}|_{L^2_{k+1}}\right) \right)^\perp,
\]
where \( G^0_{A_0} \) is the Green's operator for the Laplacian \( \Delta^0_{A_0} = d^*_{A_0} d_{A_0} \): Indeed, \( G^0_{A_0} d^*_{A_0} d_{A_0} = G^0_{A_0} \Delta^0_{A_0} \) is the \( L^2 \)-orthogonal projection \( \Pi^0_{A_0} \) from \( L^2_{k+1}(\Lambda^1 \otimes g_E) \) onto \( \left( \text{Ker}(d_{A_0}|_{L^2_{k+1}}) \right)^\perp \) and \( d_{A_0} G^0_{A_0} d^*_{A_0} \) is the \( L^2 \)-orthogonal projection \( \Pi^0_{A_0+1} = \text{id} - \Pi^0_{A_0} \) from \( L^2(\Lambda^1 \otimes g_E) \) onto \( \left( \text{Ker}(d_{A_0}|_{L^2_1}) \right)^\perp \), as
\[
d^*_{A_0} (\text{id} - d_{A_0} G^0_{A_0} d^*_{A_0}) = 0.
\]
For \( \zeta \in (\text{Ker}(d_{A_0}|_{L^2_{k+1}}))^\perp \) and \( b = d_{A_0} \zeta \in \text{Im}(d_{A_0}|_{L^2_{k+1}}) \), we have
\[
\|G^0_{A_0} d^*_{A_0} b\|_{L^2_{k+1,A_0}} = \|G^0_{A_0} \Delta^0_{A_0} \zeta\|_{L^2_{k+1,A_0}} = \|\Pi^0_{A_0} \zeta\|_{L^2_{k+1}} = \|\zeta\|_{L^2_{k+1,A_0}} \leq \|d_{A_0} \zeta\|_{L^2} + \|\zeta\|_{L^2} \leq \left(1 + \nu_0^{-1/2}\right) \|d_{A_0} \zeta\|_{L^2} = \left(1 + \nu_0^{-1/2}\right) \|b\|_{L^2}
\]
and so \( G^0_{A_0} d^*_{A_0} \) has \( \text{Hom}(L^2, L^2_{k+1,A_0}) \) operator norm bound
\[
\|G^0_{A_0} d^*_{A_0}\| \leq 1 + \nu_0^{-1/2}.
\]
The Sobolev embedding $L^2 \subset L^4$ and Kato’s inequality imply that
\[
\|d_A \zeta - d_{A_0} \zeta\|_{L^2} \leq \|a_0, \zeta\|_{L^2} \leq 2\|a_0\|_{L^4}\|\zeta\|_{L^4} \leq 2\|a_0\|_{L^4}\|\zeta\|_{L^2, A_0},
\]
and so $d_A - d_{A_0}$ has operator norm bound
\[
\|d_A - d_{A_0}\| \leq 2\|a_0\|_{L^4}.
\]
In particular, we see that $(D\Psi)^{-1}_{(id_E, A_0)} = G^0_{A_0} d^*_{A_0} \oplus \text{id} = G^0_{A_0} d^*_{A_0} \oplus G^0_{A_0} \Delta^0_{A_0}$ satisfies
\[
\|(D\Psi)^{-1}_{(id_E, A_0)}\| \leq 1 + \nu_0^{-1/2} \quad \text{and} \quad \|(D\Psi)_{(id_E, A)} - (D\Psi)_{(id_E, A_0)}\| \leq 2\|a_0\|_{L^4}.
\]
Hence, Lemma 3.3 implies that if $\|a_0\|_{L^4} < \frac{1}{2}(1 + \nu_0^{-1/2})^{-1}$, then the operator
\[
(D\Psi)_{(id_E, A)} : \left(\text{Ker} \left(d_{A_0}|_{L^2_{A_0}}\right)\right)^\perp \times \text{Ker} \left(d^*_{A_0}|_{L^2_{A_0}}\right) \rightarrow L^2(\Lambda^1 \otimes g_E)
\]
is an isomorphism from $L^2_k$ to $L^2$ and restricts to a bounded linear map from $L^2_k$ to $L^2_{k+1}$. Provided $(D\Psi)_{(id_E, A)} : L^2_{k+1} \rightarrow L^2_k$ is bijective, the open mapping theorem guarantees the existence of a bounded inverse $(D\Psi)^{-1}_{(id_E, A)} : L^2_{k+1} \rightarrow L^2_k$. If $(D\Psi)_{(id_E, A)}(\zeta, a) = 0$ for $(\zeta, a) \in L^2_{k+1}$, then $(\zeta, a)$ is zero in $L^2_k$ and thus zero in $L^2_{k+1}$, so $(D\Psi)_{(id_E, A)}$ is injective. If $b \in L^2_k(\Lambda^1 \otimes g_E)$, then $b = (D\Psi)_{(id_E, A)}(\zeta, a) = -d_A \zeta + a$ for some $(\zeta, a) \in (\text{Ker}(d_{A_0}|_{L^2_{A_0}}))^\perp \times \text{Ker}(d^*_{A_0}|_{L^2_{A_0}})$. As $d^*_{A_0} a = 0$, we have
\[
d^*_{A_0} d_A \zeta = -d^*_{A_0} b \in L^2_{k-1}
\]
and $d^*_{A_0} d_A : L^2_{k+1} \rightarrow L^2_{k-1}$ is an elliptic operator with $L^2_{k-1}$ coefficients. Thus, $\zeta \in L^2_{k+1}$, so $a = b + d_A \zeta \in L^2_k$, and $(D\Psi)_{(id_E, A)}$ is surjective.

Combining the above observations, we see that the operator
\[
(D\Psi)_{(id_E, A)} : \left(\text{Ker} \left(d_{A_0}|_{L^2_{k+1}}\right)\right)^\perp \oplus \text{Ker} \left(d^*_{A_0}|_{L^2_{k+1}}\right) \rightarrow L^2_k(\Lambda^1 \otimes g_E),
\]
is an isomorphism for all $A = A_0 + a_0$ with $\|a_0\|_{L^4} < \varepsilon_0 = \frac{1}{2}(1 + \nu_0^{-1/2})^{-1}$. So, by the Banach space implicit function theorem, there are positive constants $\varepsilon = \varepsilon(A, k)$ and $\delta = \delta(A, k)$ and an open neighborhood $U_A \subset A^4_E$ such that the map
\[
\Psi : B^+_{id_E}(\delta) \times B^4_A(\varepsilon) \rightarrow U_A, \quad (u, A_1) \mapsto u(A_1),
\]
with $B^4_A(\varepsilon) \subset B^4_{A_0}(\varepsilon_0)$, gives a diffeomorphism from an open neighborhood of $(id_E, A)$ onto an open neighborhood of $A$. In particular, we obtain a map $U_A \rightarrow \text{Stab}^+_{A_0}$, given by $A_1 \mapsto u = u_{A_1}$, such that
\[
\Psi^{-1}(A_1) = (u, u^{-1}(A_1)) \in B^+_{id_E}(\delta) \times B^4_A(\varepsilon) \subset \text{Stab}^+_{A_0} \times B^4_{A_0}(\varepsilon_0).
\]
Hence, for any \(A_1 \in U_A\) there is a unique \(u \in B_{\text{id}_E}^\infty(\delta)\) such that \(u^{-1}(A_1) - A_0 \in K_{A_0}\):

\[
d^*_A(u^{-1}(A_1) - A_0) = 0. \tag{3.5}
\]

The neighborhood \(B_{A_0}^4(\varepsilon)\) is \(\text{Stab}_{A_0}\)-invariant: If \(A \in B_{A_0}(\varepsilon)\) and \(\gamma \in \text{Stab}_{A_0}\), then

\[
\|\gamma(A) - A_0\|_{L^4} = \|A - \gamma^{-1}(A_0)\|_{L^4} = \|A - A_0\|_{L^4} < \varepsilon,
\]

and

\[
d^*_A(\gamma(A) - A_0) = \gamma\left(d^*_A(A - \gamma^{-1}(A_0))\right) = \gamma\left(d^*_A(A - A_0)\right) = 0,
\]

so \(\gamma(A) \in B_{A_0}(\varepsilon)\).

The group \(G_{E}^{k+1}\) acts on \(G_{E}^{k+1} \times S_{A_0}\) by \((u, A) \mapsto (vu, A)\), and so gives a diffeomorphism

\[
B_{\text{id}_E}(\delta) \times B_{A_0}^4(\varepsilon) \to B_{v}(\delta) \times B_{A_0}^4(\varepsilon), \quad (u, A) \mapsto (vu, A),
\]

and as this action commutes with the given action of \(\text{Stab}_{A_0}\), it descends to a diffeomorphism

\[
B_{\text{id}_E}(\delta) \times_{\text{Stab}_{A_0}} B_{A_0}^4(\varepsilon) \to B_{v}(\delta) \times_{\text{Stab}_{A_0}} B_{A_0}^4(\varepsilon), \quad [u, A] \mapsto [vu, A],
\]

for each \(v \in G_{E}^{k+1}\). Consequently, the \(G_{E}^{k+1}\)-equivariant map

\[
G_{E}^{k+1} \times_{\text{Stab}_{A_0}} B_{A_0}^4(\varepsilon) \to A_{E}^k
\]

is a local diffeomorphism onto its image, as desired. \(\square\)

Plainly, \([\gamma(A)] = [A]\) for each \(\gamma \in \text{Stab}_{A_0}\) and \(A \in B_{A_0}^4(\varepsilon)\), and hence, the projection \(\pi : B_{A_0}^4(\varepsilon) \to A_{E}^k\) factors through \(B_{A_0}^4(\varepsilon)/\text{Stab}_{A_0}\).

**Lemma 3.7.** There is a positive constant \(z\) with the following significance. Let \(\nu_0[A_0]\) be the least positive eigenvalue of the Laplacian \(\Delta^0_{A_0}\). Then for any constant \(\varepsilon_0\) satisfying \(0 < \varepsilon_0 < z(1 + \nu_0[A_0]^{-1/2})^{-1}\), the projection map \(\pi : B_{A_0}^4(\varepsilon_0)/\text{Stab}_{A_0} \to B_{E}^k\) is injective.

**Proof.** Suppose \(A_i \in B_{A_0}^4(\varepsilon)\) for \(i = 1, 2\) and that \([A_1] = [A_2] \in B_{E}^k\), so \(u(A_1) = A_2\) for some \(u \in G_{E}^{k+1}\). Since \(u(A_0) = A_0 - (d_{A_0}u)u^{-1}\), we see that \(u \in \text{Stab}_{A_0}\) if and only \(d_{A_0}u = 0\). Here, we view \(u \in L_{k+1}^2(\text{gl}(E))\) via the isometric embedding \(G_{E}^{k+1} \subset L_{k+1}^2(\text{gl}(E))\) and write

\[
u_0 \in (\text{Ker } d_{A_0})^\bot\text{ and } \gamma \in \text{Ker } d_{A_0}.\]

We claim that \(u_0 = 0\), so \(u = \gamma \in \text{Stab}_{A_0}\).
Since \( u(A_1) := A_1 - (d_{A_1} u) u^{-1} = A_2 \), we have \( A_2 u = A_1 u - d_{A_1} u = A_1 u - d_{A_0} u - [A_1 - A_0, u] \), and therefore
\[
d_{A_0} u_0 = d_{A_0} u = u(A_1 - A_0) - (A_2 - A_0) u.
\]
Since \( d_{A_0}^* (A_1 - A_0) = 0 \) for \( i = 1, 2 \), we obtain
\[
d_{A_0}^* d_{A_0} u_0 = - (* (d_{A_0} u \wedge * (A_1 - A_0)) + u d_{A_0}^* (A_1 - A_0)) - (d_{A_0}^* (A_2 - A_0)) u + *(*(A_2 - A_0) \wedge d_{A_0} u)
\]
\[
= - (* (d_{A_0} u_0 \wedge * (A_1 - A_0)) + *(*(A_2 - A_0) \wedge d_{A_0} u_0)).
\]
Integrating by parts gives
\[
\|d_{A_0} u_0\|_{L^2}^2 = (d_{A_0}^* d_{A_0} u_0, u_0)_2 \leq \|d_{A_0}^* d_{A_0} u_0\|_{L^{4/3}} \|u_0\|_{L^4}.
\]
Kato’s inequality and the embedding \( L^2 \subset L^4 \) gives \( \|u_0\|_{L^4} \leq c(\|d_{A_0} u_0\|_{L^2} + \|u_0\|_{L^2}) \), so the eigenvalue estimate \( \|u_0\|_{L^2} \leq \nu_0^{-1/2} \|d_{A_0} u_0\|_{L^2} \) gives \( \|u_0\|_{L^4} \leq c(1 + \nu_0^{-1/2}) \|d_{A_0} u_0\|_{L^2} \) and thus
\[
\|d_{A_0} u_0\|_{L^2}^2 \leq \left( 1 + \nu_0^{-1/2} \right) \|d_{A_0}^* d_{A_0} u_0\|_{L^{4/3}} \|d_{A_0} u_0\|_{L^2}.
\]
Therefore, if \( d_{A_0} u_0 \neq 0 \), the preceding expression for \( d_{A_0}^* d_{A_0} u_0 \) yields
\[
\|d_{A_0} u_0\|_{L^2} \leq c \left( 1 + \nu_0^{-1/2} \right) \|d_{A_0}^* d_{A_0} u_0\|_{L^{4/3}} \|d_{A_0} u_0\|_{L^2} \leq c \left( 1 + \nu_0^{-1/2} \right) \|A_1 - A_0\|_{L^4} + \|A_2 - A_0\|_{L^4},
\]
and so we have
\[
1 \leq c \left( 1 + \nu_0^{-1/2} \right) (\|A_1 - A_0\|_{L^4} + \|A_2 - A_0\|_{L^4}) \leq c \left( 1 + \nu_0^{-1/2} \right) \varepsilon_0
\]
which gives a contradiction for \( \varepsilon_0 < c^{-1} (1 + \nu_0^{-1/2})^{-1} \). \qed

We now return to consider the local diffeomorphism \( \Psi \) of Lemma 3.6. Suppose \( \Psi[u_1, A_1] = \Psi[u_2, A_2] \in \mathcal{A}_E^k \), where \( [u_1, A_1], [u_2, A_2] \in G_E^{k+1} \times \text{Stab}_{A_0} \mathcal{B}_E^4(\varepsilon_0) \), and so \( u_1(A_1) = u_2(A_2) \in \mathcal{A}_E^k \) and hence \( [A_1] = [A_2] \in \mathcal{B}_E^k \). Provided \( \varepsilon_0 \) also satisfies the constraints of Lemma 3.7, we have \( u_2^{-1} u_1 = \gamma \in \text{Stab}_{A_0} \) and \( \gamma(A_1) = A_2 \). Hence \( [u_2, A_2] = [u_1 \gamma^{-1}, \gamma(A_1)] = [u_1, A_1] \), so \( \Psi \) is injective and therefore a diffeomorphism onto \( \mathcal{A}_E^k \).

The map \( \pi : \mathcal{B}_E^4(\varepsilon_0) / \text{Stab}_{A_0} \to \mathcal{B}_E^k \) can be factored as the composition of the inclusion \( A \to (\text{id}_E, A) \) of \( \mathcal{B}_E^4(\varepsilon_0) \) into \( G_E^{k+1} \times \mathcal{B}_E^4(\varepsilon_0) \), the projection onto the Stab_{A_0}-quotient \( G_E^{k+1} \times \text{Stab}_{A_0} \mathcal{B}_E^4(\varepsilon_0) \), the diffeomorphism \( \Psi \) of \( G_E^{k+1} \times \text{Stab}_{A_0} \mathcal{B}_E^4(\varepsilon_0) \) with \( \mathcal{A}_E^k \) and the projection from \( \mathcal{A}_E^k \) onto the \( G_E^{k+1} \)-quotient \( \mathcal{B}_E^k = \mathcal{A}_E^k / G_E^{k+1} \). Hence, \( \pi \) is a homeomorphism onto an open neighborhood of \( [A_0] \) in \( \mathcal{B}_E^k \) and a diffeomorphism on the open subset where \( \text{Stab}_{A_0} / \text{Center}(G) \) acts freely.
Claim 3.8. The quotient space $B_k^E$ is Hausdorff.

Proof. Let $\Gamma$ be the subspace $\{\{A, u(A)\} : A \in A_k^E \text{ and } u \in G_{E^+}^{k+1}\}$ of $A_k^E \times A_k^E$. If $\{(A_\alpha), u_\alpha(A_\alpha)\}$ is a sequence in $\Gamma$ which converges in $L^2_k$ to a point $(A_\infty, B_\infty)$, then Lemma 3.1 implies that there is a subsequence $\{\alpha'\} \subset \{\alpha\}$ such that $\{u_\alpha\}$ converges in $L^2_{k+1}$ to $u_\infty \in G_{E^+}^{k+1}$ and $u_\infty(A_\infty)=B_\infty$. Thus, $\Gamma$ is closed and $A_k^E/G_{E^+}^{k+1}$ is Hausdorff. □

Claim 3.8 gives Assertion (1) of the proposition and Assertions (2), (3), and (4) now follow from the preceding arguments and Lemma 3.7. This completes the proof of the proposition. □


We now describe the basic properties of the critical-exponent norms and corresponding Banach spaces introduced by Taubes in [21], [23], [24], [25]. In particular, we give the basic embedding, multiplication, and composition lemmas we need to complete the proof of our slice theorem. We shall make frequent use of the pointwise Kato inequality, $|d|v| |\leq |\nabla_A v|$ for $v \in \Omega^1(E)$, so that the norms of the embedding and multiplication maps depend at most on the Riemannian manifold $(X, g)$. Moreover, for simplicity, we confine our attention to the case of closed four-manifolds: There are obvious analogues of the Sobolev lemmas described here for any $n$-manifold, with $n>2$. Similarly, extensions are possible to the case of complete manifolds bounded geometry (bounded curvature and injectivity radius uniformly bounded from below)—see [1], [3] for further details for Sobolev embedding results in those situations and for the construction of Green kernels. We refer the reader to the monograph of R. Adams [1] for a comprehensive treatment of Sobolev spaces and to that of E. Stein [19] for a treatment based on potential functions.

Throughout this section, $A, B$ denote $C^\infty$ orthogonal connections on Riemannian vector bundles $E, F$ over $X$ with $C^\infty$ sections $u, v$, respectively. We first have the following analogues of the $L^2$ and $L^4$ norms,

$$
\|u\|_{L^2(X)} = \sup_{x \in X} |\text{dist}^{-2}(x, \cdot)u||_{L^1(X)},
$$

$$
\|u\|_{L^4(X)} = \sup_{x \in X} |\text{dist}^{-1}(x, \cdot)u||_{L^2(X)},
$$

where $\text{dist}(x, y)$ denotes the geodesic distance between points $x$ and $y$ in $X$ defined by the metric $g$; these norms have the same behavior as the $L^2$ and $L^4$ norms with respect to constant rescalings of the metric $g$—the $L^4$ norm on two-forms and the $L^{2d}$ norm on one-forms are scale invariant. Indeed, one sees this by noting that if $g \mapsto \tilde{g} = \lambda^{-2}g$, then $\text{dist}_{\tilde{g}}(x, y) = \lambda^{-1}\text{dist}_g(x, y)$ and $dV_{\tilde{g}} = \lambda^{-4}dV_g$, while for any $a \in \Omega^1(E)$ and $v \in \Omega^2(E)$, we have $|a|_{\tilde{g}} = \lambda|a|_g$, and $|v|_{\tilde{g}} = \lambda^2|v|_g$. 


Next, we define analogues of the $L^2_k$ and $L^2_2$ norms
\[
\|u\|_{L^2_k(A)(X)} = \|\nabla_A u\|_{L^2(X)} + \|u\|_{L^2(X)},
\]
\[
\|u\|_{L^2_2(A)(X)} = \|\nabla^2_A u\|_{L^2(X)} + \|\nabla_A u\|_{L^2(X)} + \|u\|_{L^2(X)},
\]
and set
\[
\|u\|_{L^2_{1,A}(X)} = \|\nabla_A u\|_{L^2(X)} + \|u\|_{L^2(X)} + \|u\|_{L^1(X)},
\]
\[
\|u\|_{L^2_{2,A}(X)} = \|\nabla^2_A u\|_{L^2(X)} + \|u\|_{L^2(X)} + \|u\|_{L^2(X)},
\]
where $\nabla^*_A = -\ast \nabla_A^* : \Omega^1(E) \rightarrow \Omega^0(E)$ is the $L^2$-adjoint of the map $\nabla_A : \Omega^0(E) \rightarrow \Omega^1(E)$.

Finally, we define analogues of the $C^0 \cap L^2_2$ norm
\[
\|u\|_{C^0 \cap L^2_2(A)(X)} = \|u\|_{C^0(X)} + \|u\|_{L^2_2(A)(X)},
\]
and set
\[
\|u\|_{L^1_{1,A}(X)} = \|u\|_{L^1 \cap L^2(A)(X)} = \|u\|_{L^1(X)} + \|u\|_{L^2(X)},
\]
\[
\|u\|_{L^2_{3,A}(X)} = \|u\|_{L^2 \cap L^2(A)(X)} = \|u\|_{L^2(X)} + \|u\|_{L^2(X)},
\]
\[
\|u\|_{L^2_{1,A}(X)} = \|u\|_{L^1 \cap L^2(A)(X)} = \|u\|_{L^1_1(A)(X)} + \|u\|_{L^2_1(A)(X)} + \|u\|_{L^2_2(A)(X)}.
\]

It might have appeared, at first glance, a little more natural to continue the obvious pattern and instead define $\|u\|_{L^2_{1,A}(X)}$ using $\|\nabla^2_A u\|_{L^2(X)}$: As we shall see below, though, the given definition is most useful in practice. For related reasons, if $u \in \Omega^1(E) = \Omega^0(\Lambda^1 \otimes E)$, it is convenient to define the norm $\|u\|_{L^2_{1,A}(X)}$ by
\[
\|u\|_{L^2_{1,A}(X)} = \|\nabla^*_A u\|_{L^2(X)} + \|u\|_{L^2(X)} + \|u\|_{L^2(X)}.
\]

Let $L^2_k(X)$ be the Banach space completion of $C^\infty(X)$ with respect to the norm $\|\cdot\|_{L^2_k}$ and similarly define the remaining Banach spaces above.

We have the following extensions of the standard Sobolev embedding theorem \[10, 15\]: Their proofs are given in the next section. See also \[4, 17, 21, 23, \S 6, 24, \S 4, \& \S 5\], and \[25, \text{Lemma 4.7}\].

**Lemma 4.1.** The following are continuous embeddings:

1. $L^2_k(E) \subset L^2_k(E)$, for $k = 0, 1, 2$ and all $p > 2$;
2. $L^q(E) \subset L^2(E)$, for all $q > 4$;
3. $L^2_1(E) \subset L^2_2(E)$.

In the reverse direction we have:

**Lemma 4.2.** The following are continuous embeddings:

1. $L^2(E) \subset C^0 \cap L^2_1(E)$ and $L^2(E) \subset L^2(E)$;
2. $L^2_2(E) \subset C^0 \cap L^2_1(E)$. 
We next consider the extension of the standard Sobolev multiplication lemma \cite{10}, \cite{15}. While there is no continuous multiplication map $L^2_2 \times L^2_2 \to L^2_2$, it is worth observing that there is a continuous bilinear map $C^0 \cap L^2_2(E) \times C^0 \cap L^2_2(F) \to C^0 \cap L^2_2(E \otimes F)$ given by $(u, v) \mapsto u \otimes v$. Note that for $u \in \Omega^0(E)$ and $v \in \Omega^0(F)$ we have

\begin{equation}
\nabla^2_{A \otimes B}(u \otimes v) = (\nabla^A u) \otimes v + 2\nabla_A u \otimes \nabla_B v + u \otimes \nabla^2_B v,
\end{equation}

and hence the multiplication maps

\begin{equation}
\nabla^*_{A \otimes B} \nabla_{A \otimes B}(u \otimes v) = (\nabla^* A u) \otimes v + ((\ast \nabla A u) \wedge \nabla B v)
- ((\ast \nabla A u) \wedge \nabla B v) + u \otimes \nabla^*_B \nabla B v.
\end{equation}

Similarly, for $u \in \Omega^0(\Lambda^1 \otimes E)$ and $v \in \Omega^0(F)$, we have

\begin{equation}
\nabla^*_{A \otimes B}(u \otimes v) = (\nabla^* A u) \otimes v + (*u \wedge \nabla B v).
\end{equation}

In particular, we see that if $u, v \in \Omega^0(\mathfrak{gl}(E))$, then

\begin{equation}
\nabla^*_A \nabla_A (uv) = (\nabla^* A \nabla A u)v + (*u \wedge \nabla A v)
- (*u \wedge (*A v)) + u(\nabla^*_A \nabla A v),
\end{equation}

an identity we will need in the next section.

**Lemma 4.3.** Let $\Omega^0(E) \times \Omega^0(F) \to \Omega^0(E \otimes F)$ be given by $(u, v) \mapsto u \otimes v$.

Then the following hold.

1. The map $C^0(E) \otimes L^2(E \otimes F) \to L^2(E \otimes F)$ is continuous;
2. The map $L^{22}(E) \otimes L^{22}(E \otimes F) \to L^2(E \otimes F)$ is continuous;
3. The spaces $L^2_1(E)$, $L^2_2(F)$, and $L^2_2(E)$ are $L^2_2(E)$-modules;
4. The spaces $L^2_1(F)$, $L^2_2(F)$, and $L^2_2(E)$ are $L^2_2(E)$-modules.

The conclusions continue to hold for $\Omega^1(E)$ in place of $\Omega^0(E)$ and the norms on $L^2_1(\Lambda^1 \otimes E)$ and $L^2_2(\Lambda^1 \otimes E)$ defined via (4.4).

**Proof.** Let $u \in C^\infty(E)$ and $v \in C^\infty(F)$ and denote the covariant derivatives on $E$, $F$, and $E \otimes F$ by $\nabla$. Using $\nabla(u \otimes v) = (\nabla u)v + u \otimes \nabla v$ and the embedding $L^2_2(E) \subset C^0(E)$, we see that

\begin{align*}
\|u \otimes v\|_{L^2} &\leq \|u\|_{C^0} \|v\|_{L^2} \quad \text{and} \quad \|u \otimes v\|_{L^2} \leq \|u\|_{L^{22}} \|v\|_{L^{22}}, \\
\|\nabla(u \otimes v)\|_{L^2} &\leq \|\nabla u\|_{L^{22}} \|v\|_{L^{22}} + \|u\|_{C^0} \|\nabla v\|_{L^2} \leq c \|u\|_{L^2_2} \|v\|_{L^2_2} \\
\|\nabla(u \otimes v)\|_{L^2} &\leq \|\nabla u\|_{L^4} \|v\|_{L^4} + \|u\|_{C^0} \|\nabla v\|_{L^2} \leq c \|u\|_{C^0 \cap L^2_2} \|v\|_{L^2_2},
\end{align*}

and hence the multiplication maps $C^0 \times L^2 \to L^2$, $L^{22} \times L^{22} \to L^2$, and $L^2_2 \times L^2_1 \to L^2_2$ are continuous. Moreover,

\begin{align*}
\|\nabla(u \otimes v)\|_{L^2} &\leq \|\nabla u\|_{L^4} \|v\|_{L^4} + \|u\|_{C^0} \|\nabla v\|_{L^2} \leq c \|u\|_{C^0 \cap L^2_2} \|v\|_{L^2_2},
\end{align*}

and so, using the embedding $L^2_2 \subset C^0$, the multiplication $L^2_2 \times L^2_1 \to L^2_1$ is continuous. Thus, $L^2_2$ is an $L^2_2$-module and $L^2_2$ is an $L^2_2$-module.
Finally, to see that \( L^1_2 \) and \( L^{5,2}_2 \) are algebras, we use the identities (4.5), noting that
\[
\|\nabla^* \nabla (u \otimes v)\|_{L^2} \leq \|\nabla^* \nabla u\|_{L^4} \|v\|_{C^0} + 2\|\nabla u\|_{L^4} \|\nabla v\|_{L^2} + \|u\|_{C^0} \|\nabla^* \nabla v\|_{L^2} \\
\leq c\|u\|_{L^2_2} \|v\|_{L^2_2},
\]
so the multiplication \( L^1_2 \times L^1_2 \to L^1_2 \) is continuous, while
\[
\|\nabla^2 (u \otimes v)\|_{L^2} \leq \|\nabla^2 u\|_{L^2} \|v\|_{C^0} + 2\|\nabla u\|_{L^4} \|\nabla v\|_{L^4} + \|u\|_{C^0} \|\nabla^2 v\|_{L^2} \\
\leq c\|u\|_{C^0 \cap L^2_2} \|v\|_{C^0 \cap L^2_2}.
\]
The embedding \( L^{5,2}_2 \subset C^0 \) now implies that the multiplication \( L^{5,2}_2 \times L^{5,2}_2 \to L^{5,2}_2 \) is continuous.

5. Critical-exponent Sobolev embeddings and estimates for Green’s operators.

We continue the notation and assumptions of Section 4. Our goal in this section is to prove the Sobolev embedding Lemmas 4.1 and 4.2, and to derive estimates for the Green’s operator \( G_A \) of the Laplacian \( \nabla^*_A \nabla_A \) on \( \Omega^0(E) \). The key estimates described in this section are due to Taubes and they arise, in a variety of contexts, in the proofs of [17, Lemma 5.4], [21, Equation (2.14) & Lemmas 3.5, 3.6, & A.3], [23, Equation (3.4b) & Lemma 6.2], [24, Lemma 5.6], and [25, §4(c), (d), (e)]. However, we find it convenient to collect them here—together with some useful extensions and generalizations—both for the purposes of the present article and applications in [8], [9].

5.1. Estimates for the covariant Laplacian \( \nabla^*_A \nabla_A \). Let \( G(x, y) \) be the kernel function for the Green’s operator \((d^s d + 1)^{-1}\) of the Laplacian \(d^s d + 1\) on \( C^\infty(X) \). The kernel \( G(x, y) \) of \((d^s d + 1)^{-1}\) behaves like \( \text{dist}^{-2}(x, y) \) as \( \text{dist}(x, y) \to 0 \) (see [25, Lemma 4.7] and [24, §5]):

**Lemma 5.1.** The kernel \( G(x, y) \) is a positive \( C^\infty \) function away from the diagonal in \( X \times X \) and as \( \text{dist}(x, y) \to 0 \),
\[
G(x, y) = \frac{1}{4\pi^2 \text{dist}^2(x, y)} + o(\text{dist}^{-2}(x, y)).
\]

**Proof.** These and other properties of \( G \) are obtained by explicitly constructing \( G \) from an initial choice of parametrix \( H \) for \( d^s d + 1 \) using the method of [3, §4.2.2–3], where the kernel for the Green’s operator for \( d^s d \) is constructed. Recall from [19, p. 132] that the kernel \( G_0(x, y) \) for \((d^s d + 1)^{-1}\) on \( \mathbb{R}^4 \) with its standard metric satisfies
\[
G_0(x, y) = \frac{1}{4\pi^2 |x - y|^2} + o(|x - y|^{-2}), \quad |x - y| \to 0.
\]
The kernel \( G \) is now constructed using \( G_0 \) by following the method of [3, §4.2.2–3].
Hence, for all \( f \) at a point \( x \) via integration by parts. Therefore, \( u \) 

\[
-\Delta u \leq c \left( \operatorname{dist}^{-2}(x, y) \right),
\]

Consequently, for all \( u \in \Omega^0(E) \), we have 

\[
c^{-1} \| u \|_{L^2(X)} \leq \| G \|_{C^0(X)} \leq c \| u \|_{L^4(X)}.
\]

Lemma 4.2 will follow from the next estimate; a similar inequality is stated as Equation (3.4) in [24]; see [17, Lemma 5.4(a)] for a related result on \( \mathbb{R}^3 \).

**Lemma 5.2.** For all \( f \in L^2_4(\mathbb{R}^4) \), where \( \mathbb{R}^4 \) has its standard metric, 

\[
\sup_{x \in \mathbb{R}^4} \| \text{dist}^{-1}(x, \cdot) f \|_{L^2(\mathbb{R}^4)} \leq \frac{1}{2} \| \nabla f \|_{L^2(\mathbb{R}^4)}.
\]

Suppose \( X \) be a closed, oriented, Riemannian four-manifold. Then there is a positive constant \( c \) such that for all \( f \in L^2_4(X) \), 

\[
\sup_{x \in X} \| \text{dist}^{-1}(x, \cdot) f \|_{L^2(X)} \leq c \| f \|_{L^2_4(X)}.
\]

**Proof.** Let \( f \in C^\infty_0(\mathbb{R}^4) \) and let \( x = (r, \theta) \) denote polar coordinates centered at a point \( x_0 \in \mathbb{R}^4 \), so \( r = |x - x_0| \). Then 

\[
\int_{\mathbb{R}^4} r^{-2} |f|^2 \, dx = \int_{S^3} \int_{\mathbb{R}} r f^2 \, dr \, d\theta
\]

\[
= \frac{1}{2} \int_{S^3} \int_{\mathbb{R}} \frac{d^2}{dr^2} f^2 \, dr \, d\theta = \frac{1}{2} \int_{S^3} \int_{\mathbb{R}} r^2 \frac{\partial f}{\partial r} \, dr \, d\theta,
\]

via integration by parts. Therefore, 

\[
\int_{\mathbb{R}^4} r^{-2} |f|^2 \, dx = -\frac{1}{2} \int_{\mathbb{R}^4} r^{-1} f \frac{\partial f}{\partial r} \, dx
\]

\[
\leq \frac{1}{2} \left( \int_{\mathbb{R}^4} r^{-2} |f|^2 \, dx \right)^{1/2} \left( \int_{\mathbb{R}^4} |\nabla f|^2 \, dx \right)^{1/2}.
\]

Hence, for all \( f \in C^\infty_0(\mathbb{R}^4) \) we have 

\[
\| \text{dist}^{-1}(x_0, \cdot) f \|_{L^2(\mathbb{R}^4)} \leq \frac{1}{2} \| \nabla f \|_{L^2(\mathbb{R}^4)},
\]

and taking the supremum over \( x_0 \in \mathbb{R}^4 \) yields the first assertion.

For a closed Riemannian manifold \( X \), choosing a smooth partition of unity for \( X \) and applying first assertion (when \( X = \mathbb{R}^4 \)) to each patch then yields the second assertion. \( \square \)

**Proof of Lemma 4.1.** Define \( 1 \leq p' < 2 \) by setting \( 1 = 1/p + 1/p' \). Then Hölder’s inequality implies that 

\[
\| \text{dist}^{-2}(x, \cdot) |u| \|_{L^1} \leq \| \text{dist}^{-2}(x, \cdot) \|_{L^{p'}} \| u \|_{L^p} \leq C \| u \|_{L^p},
\]

\[
\| \text{dist}^{-1}(x, \cdot) |u| \|_{L^2} \leq \| \text{dist}^{-1}(x, \cdot) \|_{L^{2p'}} \| u \|_{L^{2p}} \leq C \| u \|_{L^{2p}},
\]

and taking the supremum over \( x_0 \in \mathbb{R}^4 \) yields the first assertion.
which gives Assertions (1) and (2). By Lemma 5.2 and Kato's inequality, 
\[ |d|u| \leq |\nabla_A u|, \]
we see that
\[
\sup_{x \in X} \| \text{dist}^{-1}(x, \cdot)u \|_{L^p} = \sup_{x \in X} \| \text{dist}^{-1}(x, \cdot)|u| \|_{L^p} \\
\leq C(\|d|u|\|_{L^p} + \|u\|_{L^p}) \\
\leq C(\|\nabla_A u\|_{L^p} + \|u\|_{L^p}).
\]

Taking \( p = 2 \) gives Assertion (3). \( \square \)

Lemma 4.2 will follow from the estimates below; the key estimates (1) and (2) in Lemma 5.3 below and the estimates (1), (2), and (3) in Lemma 5.4 are essentially those of Lemma 6.2 in [23], except that the dependence of the constant on \( \|F_A\|_{L^2} \) is made explicit, but the argument is the same as that of [23].

**Lemma 5.3.** Let \( X \) be a closed, oriented four-manifold with metric \( g \). Then there is a constant \( c \) with the following significance. Let \( E \) be a Riemannian vector bundle over \( X \) and let \( A \) be an orthogonal \( L^2_2 \) connection on \( E \) with curvature \( F_A \). Then \( L^2_1(E) \subset C^0 \cap L^2_1(E) \) and the following estimates hold:

1. \( \|\nabla_A u\|_{L^2_2(X)} + \|u\|_{C^0(X)} \leq c\|\nabla_A^* \nabla_A u\|_{L^2_1(X)} + \|u\|_{L^2_1(X)}, \)
2. \( \|\nabla_A u\|_{L^2_2(X)} + \|u\|_{C^0(X)} \leq c\|\nabla_A^* \nabla_A u\|_{L^2_1(X)} + \|u\|_{L^2_1(X)}, \)
3. \( \|u\|_{L^1_1(X)} \leq c\|u\|_{L^2_1(X)}, \)
4. \( \|u\|_{L^2_1(X)} \leq c\|u\|_{L^2_2(X)}, \)
5. \( \|\nabla_A u\|_{L^2_1(X)} \leq c\|\nabla_A u\|_{L^2_2(X)}. \)

**Proof.** For any \( u \in C^\infty(E) \) there is the following pointwise identity [10, p. 93],
\[
|\nabla_A u|^2 + \frac{1}{2} d^* d|u|^2 = \langle \nabla_A^* \nabla_A u, u \rangle,
\]
and thus:
\[
|\nabla_A u|^2 + \frac{1}{2} (1 + d^* d)|u|^2 = \langle \nabla_A^* \nabla_A u, u \rangle + \frac{1}{2}|u|^2.
\]

Using this identity and the fact that \( \int_X G(x, \cdot) (d^* d + 1)|u|^2 dV = |u|^2(x) \), we obtain
\[
\int_X G(x, \cdot) |\nabla_A u|^2 dV + \frac{1}{2}|u|^2(x) \\
\leq \int_X G(x, \cdot) \langle \nabla_A^* \nabla_A u, u \rangle dV + \frac{1}{2} \int_X G(x, \cdot) |u|^2 dV.
\]

Therefore, from (5.1), we have
\[
\|\nabla_A u\|_{L^2}^2 + \|u\|_{C^0}^2 \leq c\|\nabla_A^* \nabla_A u\|_{L^1} + c\|u\|_{L^2}^2 + c\|u\|_{C^0}^2.
\]
Consequently, using rearrangement with the last term, we see that

$$\|\nabla u\|_{L^2} + \|u\|_{C^0} \leq c\|\nabla A \nabla u\|_{L^2} + c\|u\|_{L^2},$$

giving (1). Combining this estimate with the embedding and interpolation inequalities, \(\|u\|_{L^2} \leq c\|u\|_{L^4} \leq c\|u\|_{L^2}^{1/2}\|u\|_{C^0}^{1/2}\), and again using rearrangement with the last term yields the bound in (2). Since \(X\) is closed, for all \(x \neq y\) we have \(\text{dist}(x,y) \leq M < \infty\), so

$$\int_X \text{dist}^{-2}(x,\cdot)|u| \, dV \geq M^{-2} \int_X |u| \, dV,$$

and this gives the estimates in (3), (4), and (5). \(\blacksquare\)

**Proof of Lemma 4.2.** From Lemma 5.3 we have the estimate

$$\|u\|_{C^0} \leq c\|u\|_{L^2(A_0^*)},$$

for any \(u \in C^\infty(E)\). Let \(\{u_m\}\) be a sequence in \(C^\infty(E)\) converging to \(u \in L^2_2(E)\). The sequence \(\{u_m\}\) is Cauchy in \(L^2_2(E)\) and applying the preceding estimate to the differences \(u_{m_2} - u_{m_1}\), we see that it is Cauchy in the Banach space \(C^0(E)\) and so the limit \(u\) lies in \(C^0(E)\). The same argument, with estimates (1) and (5) of Lemma 5.3, shows that \(u \in L^2_2(E)\) and this yields Assertion (2) of the lemma. Assertion (1) follows in the same manner. \(\blacksquare\)

**Lemma 5.4.** Continue the hypotheses of Lemma 5.3. Then for any \(u \in (C^0 \cap L^2_2)(E)\), we have

\[
\begin{align*}
(1) & \quad \|\nabla_A^2 u\|_{L^2(X)} \leq \|\nabla_A \nabla_A u\|_{L^2(X)} + c\|F_A\|_{L^2(X)} \|\nabla u\|_{L^4(X)} \\
& \quad \quad \quad + \|F_A\|_{L^2(X)} \|u\|_{C^0(X)}, \\
(2) & \quad \|\nabla_A u\|_{L^4(X)} \leq \|u\|_{C^0(X)}^{1/2} \left(\|\nabla_A \nabla_A u\|_{L^2(X)} + 2\|\nabla_A u\|_{L^2(X)}\right)^{1/2}, \\
(3) & \quad \|\nabla_A^2 u\|_{L^2(X)} \leq 2\|\nabla_A \nabla_A u\|_{L^2(X)} + c\|F_A\|_{L^2(X)} \|u\|_{C^0(X)}.
\end{align*}
\]

**Proof.** The Bochner-Weitzenböck formula for the covariant Laplacian [11, Appendix, Theorem II.1] asserts that

\[d_A^* d_A - d_A d_A^* = \nabla_A^* \nabla_A + \{F_A, \cdot\},\]

where we use \(\{\cdot, \cdot\}\) to denote a certain bilinear map whose precise form is unimportant here. Integrating by parts and noting that \(d_A = \nabla_A\) and \(d_A^* d_A = \nabla_A^* \nabla_A\) on \(\Omega^0(X, V)\) and \(F_A = d_A \circ d_A^*\) gives

\[
\begin{align*}
\|\nabla_A^2 u\|_{L^2}^2 & = \langle \nabla_A^* \nabla_A \nabla_A u, \nabla_A u \rangle_{L^2} \\
& = \left\langle (d_A^* d_A + d_A d_A^*) d_A u, d_A u \right\rangle_{L^2} - \left\langle \{F_A, d_A u\}, \nabla_A u \right\rangle_{L^2} \\
& = \left\langle d_A^* F_A u, d_A u \right\rangle_{L^2} + \left\langle d_A (d_A^* d_A) u, d_A u \right\rangle_{L^2} - \left\langle \{F_A, \nabla_A u\}, \nabla_A u \right\rangle_{L^2} \\
& = \left\langle F_A u, F_A u \right\rangle_{L^2} + \left\langle \nabla_A \nabla_A u, \nabla_A^* \nabla_A u \right\rangle_{L^2} - \left\langle \{F_A, \nabla_A u\}, \nabla_A u \right\rangle_{L^2},
\end{align*}
\]
Therefore, applying Hölder’s inequality, we find that
\[ \|\nabla^2_A u\|_{L^2}^2 \leq \|\nabla^*_A \nabla A u\|_{L^2}^2 + c\|F_A\|_{L^2}\|\nabla A u\|_{L^2}^2 + \|F_A\|_{L^2}^2\|u\|_{C^0}^2, \]
and taking square roots gives the desired bound in (1).

We now use integration by parts and Kato’s inequality \(|d|u| \leq |\nabla A u|\) to obtain an \(L^4\) bound on \(d_A u\):
\[
\|d_A u\|_{L^4}^4 = (d_A u, |d_A u|^2 d_A u)_{L^2}
= (u, |d_A u|^2 d_A^* d_A u)_{L^2} + 2 (u, |d_A u| d_A u \wedge d_A u)_{L^2},
= \|u\|_{C^0} \|d_A u\|_{L^4}^2 \|d_A^* d_A u\|_{L^2} + 2 \|u\|_{C^0} \|d_A u\|_{L^2} \|\nabla A d_A u\|_{L^2},
\]
and so, if \(d_A u \neq 0\),
\[
\|d_A u\|_{L^4} \leq \|u\|_{C^0}^{1/2} (\|d_A^* d_A u\|_{L^2} + 2 \|\nabla^2_A u\|_{L^2})^{1/2},
\]
which gives the desired estimate in (2).

By combining the \(L^4\) estimate for \(\nabla A u\) with the \(L^2\) estimate for \(\nabla^2_A u\), we obtain
\[
\|\nabla^2_A u\|_{L^2} \leq \|\nabla^*_A \nabla A u\|_{L^2} + \|F_A\|_{L^2}\|u\|_{C^0} + c\|F_A\|_{L^2}^{1/2}\|u\|_{C^0}^{1/2} (\|\nabla^*_A \nabla A u\|_{L^2} + \|\nabla^2_A u\|_{L^2})^{1/2}.
\]

We now use rearrangement with the last term above to give
\[
\|\nabla^2_A u\|_{L^2} \leq 2 \|\nabla^*_A \nabla A u\|_{L^2} + c\|F_A\|_{L^2}\|u\|_{C^0},
\]
and this establishes the desired bound in (3).

**Lemma 5.5.** Continue the hypotheses of Lemma 5.3. Then for any \(u \in L^2\)\(_{\text{rad}}\)(\(E\)), we have:
\[
\|u\|_{L^2_{\text{rad}}(X)} + \|u\|_{C^0(X)} \leq c(1 + \|F_A\|_{L^2(X)}) \left(\|\nabla^*_A \nabla A u\|_{L^2(X)} + \|u\|_{L^2(X)}\right).
\]

**Proof.** From Assertion (3) of Lemma 5.4 we have the estimate
\[
\|\nabla^2_A u\|_{L^2} \leq 2 \|\nabla^*_A \nabla A u\|_{L^2} + c\|F_A\|_{L^2}\|u\|_{C^0},
\]
while integration by parts gives
\[
\|\nabla A u\|_{L^2} = (\nabla^*_A \nabla A u, u)_{L^2}^{1/2} \leq \frac{1}{\sqrt{2}} (\|\nabla^*_A \nabla A u\|_{L^2} + \|u\|_{L^2}).
\]
According to Lemma 5.3 we have
\[
\|u\|_{C^0} \leq c\|\nabla^*_A \nabla A u\|_{L^2} + c\|u\|_{L^2},
\]
and therefore the desired bound follows by combining these estimates. \(\square\)
The above lemmas lead to the following estimates for the Green’s operator
\( G_A : L^{2,2}(E) \to L^{2,2}(E) \) of the Laplacian \( \nabla^*_A \nabla_A : L^{2,2}(E) \to L^{2,2}(E) \).
For \( u \in \Omega^0(E) \) define
\[
\| u \|_{L^{2,2}_A(X)} = \| \nabla_A^* \nabla_A u \|_{L^{2,2}(X)} + \| u \|_{L^2(X)},
\]
and observe that this is equivalent to the \( L^{2,2}_A \) norm defined in Section 4, although the comparison depends on the \( L^2 \) norm of the curvature \( F_A \).

**Lemma 5.6.** Continue the hypotheses of Lemma 5.3. Let \( \nu_0[A] \) be the least positive eigenvalue of the Laplacian \( \nabla^*_A \nabla_A \). Then for any \( u \in L^{2,2} \cap (\text{Ker} \nabla^*_A \nabla_A)^\perp \), we have:

1. \( \| G_A u \|_{L^{2,2}_A(X)} \leq c(1 + \nu_0[A]^{-1})\| u \|_{L^2(X)} \),
2. \( \| F_A u \|_{L^{2,2}_A(X)} \leq c(1 + \nu_0[A]^{-1})\| u \|_{L^2(X)} \),
3. \( \| G_A u \|_{L^{2,2}_A(X)} \leq (1 + \nu_0[A]^{-1})(1 + \| F_A \|_{L^2(X)})\| u \|_{L^2(X)} \).

**Proof.** The first and second assertions follow from Lemma 5.3, the fact that \( \nabla^*_A \nabla_A G_A u = u \) for \( u \in (\text{Ker} \nabla^*_A \nabla_A)^\perp \), and the eigenvalue estimate \( \| u \|_{L^2} \leq \nu_0[A]^{-1}\| \nabla_A^* \nabla_A u \|_{L^2} \), while the third assertion follows from the first and Lemma 5.5. \( \Box \)

### 5.2. Elliptic estimates for \( d_A^+ + d_A^- \)

To illustrate their application and to point to possible extensions, we note that the estimates of Section 5.1 for the covariant Laplacian \( \nabla^*_A \nabla_A = d_A^* d_A \) on \( \Omega^0(E) \) naturally extend to give estimates for the covariant Laplacians \( \nabla^*_A \nabla_A \) on \( \Omega^0(\Lambda^\ell(E) = \Omega^0(\Lambda^\ell \otimes E) \). Estimates for \( \nabla_A \nabla_A \) on \( \Omega^0(\Lambda^1 \otimes E) \) and \( \Omega^0(\Lambda^+ \otimes E) \) are of particular interest since these can in turn be profitably compared (via the Bochner-Weitzenböck formulas [10, Equations (6.25) & (6.26)], as in [24]) with the remaining Laplacians defined by the elliptic deformation complex for the anti-self-dual equation [5], [10], namely \( d_A d_A^* + d_A^* d_A^+ \) on \( \Omega^0(\Lambda^1 \otimes g_E) \) and \( d_A^+ d_A^* \) on \( \Omega^0(\Lambda^+ \otimes g_E) \). Indeed, if \( B_1 \) and \( B_+ \) are the Levi-Civita connections on \( \Lambda^1 \) and \( \Lambda^+ \) induced by the Levi-Civita connection on \( TX \) for the metric \( g \), then the curvature “\( F_A \)” in the estimates of the preceding subsection is simply replaced by [12, p. 165]

\[
F_{B_1} = F_{B_1} \otimes \text{id}_{g_E} + \text{id}_{A_1} \otimes F_A,
\]

\[
F_{B_+} = F_{B_+} \otimes \text{id}_{g_E} + \text{id}_{A_+} \otimes F_A,
\]

where \( F_{B_1} \) and \( F_{B_+} \) are expressed in terms of the Riemann curvature tensor Rm and where we abuse notation slightly and denote the connections on \( E \) and \( g_E \) both by \( A \). (See [10, Appendix C] and [11, Appendix II].) In the interests of brevity we shall confine our attention to the case of \( L^p \) estimates with \( p = 2 \), though the methods can be modified to obtain estimates for \( p \neq 2 \) (some work is required—see [5, p. 426] for hints).
In order to compute the required elliptic estimates for \(d^+_A\) we will need the Bochner-Weitzenböck formulas,
\[
\begin{align}
(5.6) & \quad d_Ad^+_A + 2d^*_Ad^+_A = \nabla^*_A \nabla_A + \{ \text{Ric}, \cdot \} - 2\{ F^-_A, \cdot \}, \\
(5.7) & \quad 2d^+_Ad^*_A = \nabla^*_A \nabla_A - 2\{ W^+, \cdot \} + \frac{R}{3} + \{ F^+_A, \cdot \},
\end{align}
\]
for the Laplacians on \(\Omega^1(g_E)\) and \(\Omega^+(g_E)\) \cite[Equations (6.25) & (6.26)]{10}; here, \(\text{Ric}\), \(W^+\), and \(R\) are the Ricci, self-dual Weyl, and scalar curvatures of the Riemannian metric \(g\) on \(X\). In applications to the degeneration of anti-self-dual or “almost anti-self-dual” connections \(A\) as in \cite{20, 22, 23, 24}, we can usually arrange to have a uniform \(L^\infty\) bound on \(F^+_A\), but not a uniform \(L^p\) bound on \(F^-_A\) when \(p > 2\). We derive estimates in the remainder of this subsection with such applications and assumptions in view. To illustrate the nature of the difficulty we first derive a naive estimate for \(a \in L^2_{1,A}(g_E)\) in terms of the \(L^2\) norm of \((d^*_A + d^+_A)a\):

**Lemma 5.7.** Let \(X\) be a closed, oriented four-manifold with metric \(g\). Then there is a constant \(c\) with the following significance. Let \(E\) be a Riemannian vector bundle over \(X\) and let \(A\) be an orthogonal \(L^2\) connection on \(E\) with curvature \(F_A\). Then for any \(a \in L^2_{1}((\Lambda^1 \otimes g_E))\),
\[
\|a\|_{L^2_{1,A}(X)} \leq \sqrt{2}\|d^*_A + d^+_A\|a\|_{L^2(X)} \\
+ c \left(1 + \|F^-_A\|_{C^0(X)}\right)^{1/2} \|a\|_{L^2(X)},
\]
(5.8)

If \(a\) is \(L^2\)-orthogonal to \(\text{Ker} d^+_A\), so that \(a = d^*_A v\) for some \(v \in L^2_{1}(\Lambda^+ \otimes g_E)\), then
\[
\|d^*_Av\|_{L^2_{1,A}(X)} \leq \sqrt{2}\|d^+_A d^*_A v\|_{L^2(X)} + c \left(1 + \|F^-_A\|_{C^0(X)}\right)^{1/2} \|v\|_{L^2(X)} \\
+ \|F^+_A\|_{C^0(X)} \|v\|_{L^2(X)}.
\]
(5.9)

**Proof.** From the Bochner-Weitzenböck formula for \(d_Ad^*_A + 2d^*_Ad^+_A\) in (5.6) and integration by parts, we have:
\[
\|\nabla_A a\|_{L^2}^2 = (\nabla^*_A \nabla_A a, a) \\
= (d_Ad^*_A a, a) + 2(d^*_A d^+_A a, a) - \{ \text{Ric}, a \}, a\) + 2\{F^-_A, a\}, a\)
\leq \|d^*_A a\|_{L^2}^2 + 2\|d^+_A a\|_{L^2}^2 + c \left(1 + \|F^-_A\|_{C^0}\right) \|a\|_{L^2}^2
\]
which gives (5.8). If \(a = d^*_A v\), then \(d^*_A d^*_A v = (d^+_A d_A)^* v = (F^+_A)^* v\), so that
\[
\|d^*_A d^*_A v\|_{L^2} \leq \|F^+_A\|_{C^0} \|v\|_{L^2}.
\]
Thus, (5.9) follows from (5.8) and the above inequality. \hfill \Box

Since \(d^*_A + d^+_A\) is an elliptic operator, estimates of the above form follow from the general theory of linear elliptic operators. However, the preceding elementary derivation using the Bochner-Weitzenböck formula gives us a
constant whose dependence on the curvature terms $F_A^-$ and $F_A^+$ is made explicit. In particular, we see that the estimate is only useful when we have a uniform $C^0$ bound on $F_A^-$ independent of $A$, which is not possible when $A$ bubbles. At the cost of introducing a slightly stronger norm than the $L^2$ norm on the right hand side of the estimate above, we can derive an $L^2_{1,A}$ bound for $a = d_A^* v$ with an estimate constant depending on $\|F_A^-\|_{L^2(X)}$ rather than $\|F_A^-\|_{C^0(X)}$. Specifically, Equation (5.5) and Lemma 5.5 give the following $L^2_{2,A}$ estimates for sections of $\Lambda^+ \otimes g_E$:

**Lemma 5.8.** Continue the hypotheses of Lemma 5.7. Then the following estimate holds for any $v \in L^2_{2,A}(\Lambda^+ \otimes g_E)$:

$$\|v\|_{L^2_{2,A}(X)} + \|v\|_{C^0(X)} \leq c(1 + \|F_A\|_{L^2(X)})(\|\nabla_A^* \nabla_A v\|_{L^2(X)} + \|v\|_{L^2(X)}).$$

We now replace the covariant Laplacian $\nabla_A^* \nabla_A$ in the estimates of Lemma 5.8 by the Laplacian $d_A^+ d_A^*$ via the Bochner formula (5.7) to give:

**Lemma 5.9.** Continue the hypotheses of Lemma 5.8. Then there is a positive constant $\varepsilon = \varepsilon(c)$ such that the following holds. If $\|F_A^+\|_{L^2(X)} < \varepsilon$, then

$$\|v\|_{L^2_{2,A}(X)} + \|v\|_{C^0(X)} \leq c(1 + \|F_A\|_{L^2(X)})(\|d_A^+ d_A^* v\|_{L^2(X)} + \|v\|_{L^2(X)}).$$

**Proof.** From (5.5) and Lemma 4.3 we have

$$\|\nabla_A^* \nabla_A v\|_{L^2} \leq 2\|d_A^+ d_A^* v\|_{L^2} + c\|v\|_{L^2} + c\|F_A^+\|_{L^2} \|v\|_{C^0}.$$

Combining the preceding estimate with that of Lemma 5.8, together with the embedding and interpolation inequalities $\|v\|_{L^2} \leq c\|v\|_{L^1} \leq c\|v\|_{L^2}^{1/2} \|v\|_{C^0}^{1/2}$, and using rearrangement with the last term yields the desired bound. In particular, by choosing $\varepsilon(c)$ small enough that $c\|F_A^+\|_{L^2} \|v\|_{C^0} \leq 1/2$, we may use rearrangement to bring the right-hand term $\|v\|_{C^0}$ to the left-hand side. \qed

Since $\|d_A^* v\|_{L^2_{1,A}} \leq \|v\|_{L^2_{2,A}}$, Lemma 5.9 yields an $L^2_{1,A}$ estimate for $d_A^* v$:

**Corollary 5.10.** Continue the hypotheses of Lemma 5.9. Then:

$$\|d_A^* v\|_{L^2_{1,A}(X)} \leq c(1 + \|F_A\|_{L^2(X)})(\|d_A^+ d_A^* v\|_{L^2(X)} + \|v\|_{L^2(X)}).$$

Note that if $a \in \Omega^1(g_E)$ is $L^2$-orthogonal to Ker $d_A^+$, so that $a = d_A^* v$ for some $v \in \Omega^+(g_E)$, and Ker $d_A^+ d_A^{*+} = 0$, then the estimate of Corollary 5.10 can be written in the more familiar form

$$(5.10) \quad \|a\|_{L^2_{2,A}(X)} \leq c(1 + \|F_A\|_{L^2(X)})(\|d_A^* a\|_{L^1(X)} + \nu_2[A]^{-1/2} \|a\|_{L^2(X)}),$$

where we make use of the eigenvalue estimate $\|v\|_{L^2} \leq \nu_2[A]^{-1/2} \|d_A^* v\|_{L^2}$; the term $d_A^* a$ above can be replaced by $(d_A^+ + d_A^*) a$ without changing the estimate constants. Here, $\nu_2[A]$ is the least positive eigenvalue of the Laplacian $d_A^+ d_A^{*+}$. 

In this section we complete the proof of Theorem 1.1, and hence the proof of Theorem 1.2, using the method of continuity. The strategy broadly follows that of Uhlenbeck’s proof of Theorem 2.1 in [26]. The main new technical difficulty, not present in [26], is the need to compare distances in the Coulomb-gauge slice $S_{\mathcal{A}_0} \subset A^k_E$ through the connection $A_0$ and gauge-invariant distances in $\mathcal{B}_E^k$ from the point $[A_0]$. It is at this stage of the method of continuity (in proving openness—see Lemma 6.6)—that we need to employ the special norms and Green’s operator estimates described in Sections 4 and 5 in order to achieve the requisite $C^0$ control of gauge transformations; the proof of closedness works, as one would expect, with standard Sobolev $L^4$ and $L^2$ norms. In [26], the $L^2$ norm of the curvature $F_A$ essentially serves as a gauge-invariant $L^2$ measure of distance from $[A]$ to $[\Gamma]$, where $\Gamma$ is the product connection on the product $G$ bundle over the unit ball. Our goal in this section is to prove:

**Theorem 6.1.** Let $X$ be a closed, smooth four-manifold with metric $g$ and let $G$ be a compact Lie group. Then there are positive constants $c, z$ with the following significance. Let $E$ be a $G$ bundle over $X$ and suppose that $k \geq 2$ is an integer. Given a point $[A_0]$ in $\mathcal{B}_E^k$, let $\nu_0[A_0]$ be the least positive eigenvalue of the Laplacian $\nabla_{A_0}^* \nabla_{A_0}$ on $\Omega^0(g_E)$ and set $K_0 = (1 + \nu_0[A_0]^{-1})(1 + \|F_{A_0}\|_{L^2}).$ Let $\varepsilon_1$ be a constant satisfying $0 < \varepsilon_1 \leq zK_0^{-2}(1 + \nu_0[A_0]^{-1})^{-1}$. Then the following hold:

1. For any $[A] \in \mathcal{B}_E^k$ with $\text{dist}_{L^2_{1,A_0}}([A], [A_0]) < \varepsilon_1$, then $u \in G_{E}^{k+1}$ exists such that
   a. $d_{A_0}^*(u(A) - A_0) = 0$,
   b. $\|u(A) - A_0\|_{L^{2^*}} \leq cK_0 \text{dist}_{L^2_{1,A_0}}([A], [A_0]).$

2. For any $[A] \in \mathcal{B}_E^k$ with $\text{dist}_{L^2_{1,A_0}}([A], [A_0]) < \varepsilon_1$, then $u \in G_{E}^{k+1}$ exists such that
   a. $d_{A_0}^*(u(A) - A_0) = 0$,
   b. $\|u(A) - A_0\|_{L^{2^*}} \leq cK_0 \text{dist}_{L^2_{1,A_0}}([A], [A_0]),$
   c. $\|u(A) - A_0\|_{L^2_{1,A_0}} \leq cK_0 \text{dist}_{L^2_{1,A_0}}([A], [A_0]).$

Our first proof of Theorem 6.1, via the method of continuity, occupies Sections 6.1, 6.2 and 6.3. A rather different proof, via a direct application of the inverse function theorem using $L^2$ gauge transformations, is given in Section 8.

6.1. Distance functions on the quotient space. Our first task is to verify the existence of minimizing gauge transformations $u \in G_{E}^{k+1}$ for the
family of distance functions on $E^{2}_{2}$ defined above: This is established in Lemma 6.3 and the proof uses the following version of Uhlenbeck’s weak compactness theorem.

**Proposition 6.2.** Let $X$ be a closed, smooth, Riemannian four-manifold, let $G$ be a compact Lie group, let $M$ be a positive constant, let $A_{0}$ be an $L^{2}_{2}$ connection on a $G$ bundle $E$ over $X$. If $\{A_{\alpha}\}$ is a sequence of $L^{2}_{2}$ connections on $E$ such that $\|F_{A_{\alpha}}\|_{L^{2}_{1,A_{0}}} \leq M$, then there is a subsequence $\{\alpha'\} \subset \{\alpha\}$ and a sequence of $L^{2}_{3}$ gauge transformations $\{u_{\alpha'}\}$ such that $u_{\alpha'}(A_{\alpha'})$ converges weakly in $L^{2}_{3,A_{0}}$ and strongly in $L^{p}_{1,A_{0}}$, for $1 \leq p < 4$, to an $L^{2}_{2}$ connection $A_{\infty}$ on $E$.

**Proof.** From the Sobolev embedding $L^{2}_{1} \subset L^{p}$, $2 < p < 4$, we obtain a uniform $L^{p}$ bound $\|F_{A_{\alpha}}\|_{L^{p}} \leq cM$ and so, according to [26, Theorem 3.6], there is a subsequence $\{\alpha'\} \subset \{\alpha\}$ and a sequence of $L^{2}_{2}$ gauge transformations $\{u_{\alpha'}\}$ such that $u_{\alpha'}(A_{\alpha'})$ converges weakly in $L^{p}_{1,A_{0}}$ to an $L^{2}_{1}$ connection $A_{\infty}$ on $E$. The stronger conclusion above is obtained simply by reworking the proof of Theorem 3.6 in [26], using the following local estimate for the connections $A_{\alpha}$ over small balls $B \subset X$. Theorem 2.1 of [26] provides a sequence of local trivializations $v_{\alpha}: P|_{B} \to B \times G$ such that $a_{\alpha} = v_{\alpha}(A_{\alpha}) - \Gamma$ satisfies $d^{\ast}a_{\alpha} = 0$ and

$$\|a_{\alpha}\|_{L^{2}_{2}(B)} \leq c\|F_{A_{\alpha}}\|_{L^{p}(B)}, \quad 2 \leq p < 4,$$

where $\Gamma$ is the product connection. Now $F_{A_{\alpha}} = da_{\alpha} + a_{\alpha} \wedge a_{\alpha}$, so

$$\|a_{\alpha}\|_{L^{2}_{1}(B)} \leq \|da_{\alpha}\|_{L^{2}_{1}(B)} + \|a_{\alpha}\|_{L^{2}_{1}(B)} \leq \|a_{\alpha} \wedge a_{\alpha}\|_{L^{2}_{1}(B)} + \|F_{A_{\alpha}}\|_{L^{2}_{1}(B)} + \|a_{\alpha}\|_{L^{2}_{1}(B)}.$$

Now, using the multiplication $L^{6} \times L^{3} \to L^{2}$, the embeddings $L^{1}_{1} \subset L^{12/5} \subset L^{6}$ and $d(a_{\alpha} \wedge a_{\alpha}) = da_{\alpha} + a_{\alpha} - a_{\alpha} \wedge da_{\alpha}$, we have

$$\|d(a_{\alpha} \wedge a_{\alpha})\|_{L^{2}} \leq c\|da_{\alpha}\|_{L^{3}}\|a_{\alpha}\|_{L^{6}} \leq c\|a_{\alpha}\|_{L^{3}}^{2},$$

while $\|a_{\alpha} \wedge a_{\alpha}\|_{L^{2}} \leq \|a_{\alpha}\|_{L^{4}(B)}^{2} \leq c\|a_{\alpha}\|_{L^{2}_{1}(B)}^{2}$. Hence, we obtain

$$\|a_{\alpha}\|_{L^{2}_{1}(B)} \leq c\|F_{A_{\alpha}}\|_{L^{2}_{1}(B)} + c\|a_{\alpha}\|_{L^{2}_{1}(B)}^{2} + \|a_{\alpha}\|_{L^{2}_{1}(B)} \leq c\|F_{A_{\alpha}}\|_{L^{2}_{1}(B)}^{2}(1 + \|F_{A_{\alpha}}\|_{L^{2}_{1}(B)}) \leq c\|F_{A_{\alpha}}\|_{L^{2}_{1,A_{0}}(B)(1 + \|F_{A_{\alpha}}\|_{L^{2}_{1,A_{0}}(B)})(1 + \|A_{0} - \Gamma\|_{L^{2}}).$$

In particular, the sequence of Coulomb-gauge, local connection matrices $\{a_{\alpha}\}$ is bounded in $L^{2}_{2}(B)$, so we can extract a weakly $L^{2}_{2}(B)$-convergent and strongly $L^{p}_{1}(B)$-convergent subsequence, via the compactness of embedding $L^{2}_{2}(B) \subset L^{p}_{1}(B)$ when $1 \leq p < 4$. The patching argument used to complete the proof of Uhlenbeck’s theorem now proceeds exactly as in [26] to give the desired conclusion. □
The proposition is used to extract the desired convergence in the next lemma.

**Lemma 6.3.** For any points \([A_0], [A] \) in \(\mathcal{B}_E^k\) there are gauge transformations such that the following equalities hold:

1. \(\text{dist}_{L^4}([A], [A_0]) = \|u(A) - A_0\|_{L^4(X)}, \ u \in \mathcal{G}_E^3\)
2. \(\text{dist}_{L^2_1(A_0^q)}([A], [A_0]) = \|v(A) - A_0\|_{L^2_1(A_0)}, \ v \in \mathcal{G}_E^3\)
3. \(\text{dist}_{L^2_1(A_0)}([A], [A_0]) = \|w(A) - A_0\|_{L^2_1(A_0)}, \ w \in \mathcal{G}_E^3\)
4. \(\text{dist}_{L^2_1(A_0)}([A], [A_0]) = \|w_\ell(A) - A_0\|_{L^2_1(A_0)}, \ \ell = 1, 3 \leq \ell \leq k\)

where \(w_1 \in \mathcal{G}_E^3\) and \(w_3 \in \mathcal{G}_E^{k+1}\) in (4). The above distance functions (including the \(\ell = 2\) distance function in (4)) are continuous with respect to the quotient \(L^2_k\) topology on \(\mathcal{B}_E^k\).

**Proof.** Consider (1). Let \(\{u_\alpha\}\) be a minimizing sequence in \(\mathcal{G}_E^{k+1}\), so \(\|u_\alpha(A) - A_0\|_{L^4} \) converges to \(\text{dist}_{L^4}([A], [A_0])\) as \(\alpha \to \infty\). Setting \(B_\alpha = u_\alpha(A) = A - (d_Au_\alpha)u_\alpha^{-1} \in \mathbb{A}^E_1\), we see that \(B_\alpha u_\alpha = Au_\alpha - d_Au_\alpha = Au_\alpha - A_0u_\alpha - [A - A_0, u_\alpha]\), and thus

\[
(6.1) \quad d_{B_\alpha} u_\alpha = u_\alpha(A - A_0) - (B_\alpha - A_0) u_\alpha.
\]

Therefore, as \(\|u_\alpha\|_{C^0} \leq c(G)\), we have

\[
\|\nabla_A_\alpha u_\alpha\|_{L^4} \leq c(\|A - A_0\|_{L^4} + \|B_\alpha - A_0\|_{L^4}),
\]

so the sequence \(\{u_\alpha\} \subset L^4_1(A_0)\) is bounded in \(L^4_1(A_0)\). So, passing to a subsequence, we may suppose that \(\{u_\alpha\}\) converges weakly in \(L^q_1(A_0)\) and strongly in \(L^p_1(A_0)\), via the compact embedding \(L^4_1 \subset L^q_1\), for any \(1 \leq p < \infty\), to a limit \(u \in L^4_1\).

We also have \(F_{A_\alpha} = F_{u_\alpha(A)} = u_\alpha F_A u_\alpha^{-1}\), so \(\|F_{A_\alpha}\|_{L^2} \leq \|F_A\|_{L^2}\) and as \(\nabla_A F_{A_\alpha} = (\nabla_A u_\alpha) \otimes F_A u_\alpha^{-1} + u_A (\nabla_A F_A) u_\alpha^{-1} - u_A F_A \otimes u_\alpha^{-1} (\nabla_A u_\alpha) u_\alpha^{-1}\), we see that

\[
(6.2) \quad \|\nabla_A F_{A_\alpha}\|_{L^2} \leq c(\|\nabla_A u_\alpha\|_{L^4} \|F_A\|_{L^4} + \|\nabla_A F_A\|_{L^2})
\]

\[
\leq c(1 + \|u_\alpha\|_{L^4_1(A_0)}) \|F_A\|_{L^2_1(A_0)}.
\]

Hence, the sequence of \(L^2_1\) connections \(\{A_\alpha\}\) has curvature uniformly bounded in \(L^2_1(A_0)\): Proposition 6.2 implies, after passing to a subsequence, that the sequence \(\{A_\alpha\}\) converges weakly in \(L^2_1(A_0)\) and strongly in \(L^p_1(A_0)\), for \(1 \leq p \leq \frac{4}{3}\), to an \(L^2_1\) connection \(B\) on \(E\). From (6.1) we obtain

\[
(6.3) \quad d_{A_\alpha} u = u(A - A_0) - (B - A_0) u,
\]

a first-order linear elliptic equation in \(u\) with \(L^2_1\) coefficients. Therefore, \(u \in L^2_3(\mathfrak{gl}(E))\) and \(B = u(A) = A - (d_Au)u^{-1}\) lies in \(\mathbb{A}^E_1\). It is not a priori
clear that the limit $u$ actually lies in $G^3_E$ (since the convergence was only weakly $L^4_{1,A_0}(\mathfrak{gl}(E))$ and strongly $L^3(\mathfrak{gl}(E))$). However, the argument of the last paragraph in the proof of Lemma 4.2.4 in [5, p. 130] applies (using the compactness of the structure group $G$) and shows that the limit gauge transformation $u$ lies in $G^3_E$. Since $B_\alpha = u_\alpha(A)$ converges strongly in $L^p_{1,A_0}$ to $u(A)$ we now have
\[
\text{dist}_{L^4}([A], [A_0]) = \lim_{\alpha \to \infty} \|u_\alpha(A) - A_0\|_{L^4} = \|u(A) - A_0\|_{L^4},
\]
as required in (1). The same argument proves Assertions (2) and (3) and Assertion (4) when $\ell = 1$. The case $\ell \geq 3$ in (4) is straightforward as we can now apply Lemma 3.1 to obtain the desired convergence.

It remains to check $L^k_\mu$-continuity. We just consider (1), as the remaining cases are identical. If $[A_\alpha] \in B^4_E$ is a sequence converging to $[A_\infty] \in B^4_E$, then there is a sequence of gauge transformations $s_\alpha \in G^{k+1}_E$ such that $s_\alpha(A_\alpha)$ converges in $L^2_{k,A_0}$ to $A_\infty \in A^4_E$ and, in particular, in $W^{1,4}$. But then
\[
\text{dist}_{L^4}([A_\alpha], [A_0]) - \text{dist}_{L^4}([A_\infty], [A_0]) = \text{dist}_{L^4}([s_\alpha(A_\alpha)], [A_0]),
\]
and so
\[
\lim_{\alpha \to \infty} \text{dist}_{L^4}([A_\alpha], [A_0]) = \text{dist}_{L^4}([A_\infty], [A_0]),
\]
as desired. \hfill \Box

6.2. Closedness. Let $\mathfrak{B} \subset \overline{B}^{1,*2}_{[A_0]}(\varepsilon)$ be the subset of points $[A]$ such that there exists a gauge transformation $u \in G^{k+1}_E$ satisfying the conclusions of Assertion (2) of Theorem 6.1; let $\mathfrak{B}^* \subset \overline{B}^{1,*2}_{[A_0]}(\varepsilon)$ be the subset of points $[A]$ such that there exists a gauge transformation $u \in G^{k+1}_E$ satisfying the conclusions of Assertion (1). As in the proof of Theorem 2.1 in [26], we apply the method of continuity to show that $\mathfrak{B}^* = \overline{B}^{1,*2}_{[A_0]}(\varepsilon)$ and $\mathfrak{B} = \overline{B}^{1,*2}_{[A_0]}(\varepsilon)$ for small enough $\varepsilon$. Not surprisingly, we have:

**Lemma 6.4.** The balls $\overline{B}^{1,*2}_{[A_0]}(\varepsilon)$ and $\overline{B}^{1,*2}_{[A_0]}(\varepsilon)$ are connected.

**Proof.** If $[A] \in \overline{B}^{1,*2}_{[A_0]}(\varepsilon)$, there is a gauge transformation $u \in G^{k+1}_E$ such that $\|u(A) - A_0\|_{L^{1,*2}_{A_0}} \leq \varepsilon$. Then $A_t = A_0 + t(u(A) - A_0)$, $t \in [0, 1]$, is a path in $A^4_E$ joining $A_0$ to $u(A)$ and $\|A_t - A_0\|_{L^{1,*2}_{A_0}} = t\|u(A) - A_0\|_{L^{1,*2}_{A_0}} \leq t\varepsilon$, so the path $[A_t]$ lies in $\overline{B}^{1,*2}_{[A_0]}(\varepsilon)$ and joins $[A_0]$ to $[A]$. Similarly for $\overline{B}^{1,*2}_{[A_0]}(\varepsilon)$. \hfill \Box

Our task then reduces to showing that $\mathfrak{B}^*$ is an open and closed subspace of $\overline{B}^{1,*2}_{[A_0]}(\varepsilon)$ and that $\mathfrak{B}$ is an open and closed subspace of $\overline{B}^{1,*2}_{[A_0]}(\varepsilon)$. First we consider closedness:
Lemma 6.5. The subspaces $\mathcal{B}^* \subset B_{[A_0]}^{1,\varepsilon,2}(\varepsilon)$ and $\mathcal{B} \subset B_{[A_0]}^{1,\varepsilon,2}(\varepsilon)$ are closed.

Proof. It suffices to consider the second assertion as the same argument yields the first. Suppose $[A_\alpha]$ is a sequence of points in $\mathcal{B}$ which converges in $\mathcal{B}^k_E$ to a point $[B_\infty]$. We may suppose, without loss of generality, that $A_\alpha \in A_E^k$ is the corresponding sequence of connections, representing the gauge-equivalence classes $[A_\alpha]$, which satisfy the defining conditions for $\mathcal{B}$:

$$d_{A_\alpha}^*(A_\alpha - A_0) = 0,$$

$$\|A_\alpha - A_0\|_{L^{2,4}} \leq cK_0 \text{dist}_{L^{2,4}}([A_\alpha], [A_0]),$$

$$\|A_\alpha - A_0\|_{L^{2,4}} \leq cK_0 \text{dist}_{L^{2,4}}([A_\alpha], [A_0]).$$

(6.4)

Since $[A_\alpha]$ converges in $\mathcal{B}^k_E$ to $[B_\infty]$, there is a sequence of gauge transformations $u_\alpha \in \mathcal{G}_E^{k+1}$ such that $B_\alpha := u_\alpha(A_\alpha)$ converges in $L^2_{E,A_0}$ to $B_\infty \in \mathcal{A}_E^k$. Since $B_\alpha = u_\alpha(A_\alpha)$ and $d_{A_\alpha}^*(A_\alpha - A_0) = 0$, we have

$$d_{A_\alpha} u_\alpha = u_\alpha(A_\alpha - A_0) - (B_\alpha - A_0) u_\alpha,$$

(6.5)

$$d_{A_\alpha}^* du_\alpha = -*(d_{A_\alpha} u_\alpha \wedge (A_\alpha - A_0)) - (d_{A_\alpha}^*(B_\alpha - A_0)) u_\alpha - *(B_\alpha - A_0) \wedge d_{A_\alpha} u_\alpha,$$

(6.6)

and so, as $\|u_\alpha\|_{C^\infty} \leq 1$,

$$\|d_{A_\alpha} u_\alpha\|_{L^{2,4}} \leq \|A_\alpha - A_0\|_{L^{2,4}} + \|B_\alpha - A_0\|_{L^{2,4}}$$

$$\|d_{A_\alpha}^* du_\alpha\|_{L^{2,4}} \leq \|d_{A_\alpha} u_\alpha\|_{L^{2,4}} + \|A_\alpha - A_0\|_{L^{2,4}} + \|d_{A_\alpha}^* B_\alpha - A_0\|_{L^{2,4}}$$

$$\|B_\alpha - A_0\|_{L^{2,4}} \|d_{A_\alpha} u_\alpha\|_{L^{2,4}}.$$

Therefore, the sequence $u_\alpha$ is bounded in $L^2_{E,A_0}(\mathfrak{gl}(E))$ and so, passing to a subsequence, we may suppose that $u_\alpha$ converges weakly in $L^2_{E,A_0}(\mathfrak{gl}(E))$ (and strongly in $L^p_{E,A_0}$ for any $p < 4$ via the compact embedding $L^2_{E,A_0} \subset L^p_{E,A_0}$) to a limit $u_\infty \in L^{\infty}_E \cap L^2_{E,A_0}(\mathfrak{gl}(E))$.

On the other hand, using $A_\alpha = u_\alpha^{-1}(B_\alpha)$, we have $\|F_{A_\alpha}\|_{L^2} = \|F_{B_\alpha}\|_{L^2}$ and the derivation of (6.2) gives

$$\|\nabla_{A_\alpha} F_{A_\alpha}\|_{L^2} \leq c(1 + \|u_\alpha\|_{L^4_{E,A_0}}) \|F_{B_\alpha}\|_{L^2_{E,A_0}}.$$

so the sequence $A_\alpha$ has curvature uniformly bounded in $L^2_{E,A_0}$. Thus, after passing to a subsequence we may assume by Proposition 6.2 that the sequence $A_\alpha$ converges weakly in $L^2_{E,A_0}$ and strongly in $L^p_{E,A_0}$, $2 \leq p < 4$, to a limit $A_\infty \in \mathcal{A}_E^k$.

Taking weak limits in (6.5) and (6.6) yields

$$d_{A_\alpha} u_\infty = u_\infty(A_\infty - A_0) - (B_\infty - A_0) u_\infty.$$

(6.7)

The equation (6.7) is first order, linear, elliptic in $u_\infty \in L^{\infty}_E \cap L^2_{E,A_0}$ with $L^2_{E,A_0}$ coefficients. Hence, $u_\infty$ is in $L^2_{E,A_0}(\mathfrak{gl}(E))$ and in particular, in $\mathcal{G}_E^{k}$, while
\[ B_\infty = u_\infty(A_\infty). \] From (6.7) we see that
\[ A_\infty - A_0 = u_\infty^{-1}(B_\infty - A_0)u_\infty + u_\infty^{-1}d_{A_0}u_\infty \]
and so, as \( d_{A_0}^*(A_\infty - A_0) = 0 \), we have
\[ (6.8) \quad d_{A_0}^*(u_\infty^{-1}d_{A_0}u_\infty + u_\infty^{-1}(B_\infty - A_0)u_\infty) = 0. \]
This is a second-order elliptic equation for \( u_\infty \in G_E^k \) with \( L^2 \) coefficients: In particular, since \( u_\infty \in L^2 \) for \( 2 \leq p \leq 4 \), a standard elliptic bootstrapping argument then implies that \( u_\infty \in L^2_{k+1} \) (see, for example, the proof of Proposition 3.3 in [6]) and therefore \( A_\infty = u_\infty^{-1}(B_\infty) \in \mathcal{A}_E^k \).

Now, taking weak limits in (6.4), we have
\[ d_{A_0}^*(A_\infty - A_0) = \lim_{\alpha \to \infty} d_{A_0}^*(A_\alpha - A_0) = 0, \]

\[ \|A_\infty - A_0\|_{L^{2,4}} = \lim_{\alpha \to \infty} \|A_\alpha - A_0\|_{L^{2,4}} \leq \lim_{\alpha \to \infty} cK_0 \text{dist}_{L^{2,2}}([A_\alpha], [A_0]), \]
\[ \|A_\alpha - A_0\|_{L^{2,2}}(\alpha) \leq \lim_{\alpha \to \infty} \|A_\alpha - A_0\|_{L^{1,2}}(\alpha) \leq \lim_{\alpha \to \infty} cK_0 \text{dist}_{L^{2,2}}([A_\alpha], [A_0]). \]
Moreover, as \( B_\infty = u_\infty(A_\infty) \) and \( u_\infty \in G_E^{k+1} \),
\[ \lim_{\alpha \to \infty} \text{dist}_{L^{2,2}}([A_\alpha], [A_0]) = \text{dist}_{L^{2,2}}([B_\infty], [A_0]) = \text{dist}_{L^{2,2}}([A_\infty], [A_0]), \]
\[ \lim_{\alpha \to \infty} \text{dist}_{L^{1,2}}([A_\alpha], [A_0]) = \text{dist}_{L^{1,2}}([B_\infty], [A_0]) = \text{dist}_{L^{1,2}}([A_\infty], [A_0]), \]
where the \( L^2 \) continuity of the distance functions is given by Lemma 6.3. Therefore, \( [B_\infty] = [A_\infty] \in \mathcal{B} \). Thus, \( \mathcal{B} \) is closed in \( \mathcal{B}_E^k \) and in particular, closed in \( \mathcal{B}_{[A_0]}^{1,2}(\varepsilon) \), as desired. □

6.3. Openness. We must first compare distances from the connection \( A_0 \) in the Coulomb slice through \( A_0 \) in \( \mathcal{A}_E^k \) and gauge-invariant distances in \( \mathcal{B}_{[A_0]}^{1,2}(\varepsilon) \) from the point \([A_0]:\)

**Lemma 6.6.** Let \((X, g)\) be a closed, smooth, Riemannian four-manifold. Then there are positive constants \( c, z \) with the following significance. Let \( K_0 = (1 + \nu_0[A_0]^{-1})(1 + \|F_{A_0}\|_{L^2}) \). If \( A \in \mathcal{A}_E^k \) satisfies
- \( d_{A_0}^*(A - A_0) = 0 \),
- \( \|A - A_0\|_{L^{2,4}} \leq zK_0^{-1} \),
then the following hold:

1. If \( \text{dist}_{L^{1,2}}([A], [A_0]) \leq zK_0^{-1} \), then
\[ \|A - A_0\|_{L^{2,4}} \leq cK_0 \text{dist}_{L^{2,2}}([A], [A_0]); \]

2. If \( \text{dist}_{L^{1,2}}([A], [A_0]) \leq zK_0^{-1} \), then
\[ \|A - A_0\|_{L^{2,4}} \leq cK_0 \text{dist}_{L^{1,2}}([A], [A_0]), \]
\[ \|A - A_0\|_{L^{1,2}} \leq cK_0 \text{dist}_{L^{1,2}}([A], [A_0]). \]
Proof. Recall that for either distance function, minimizing gauge transformations in \( G^3_E \) exist by Lemma 6.3; for convenience, we denote both by \( u \in G^3_E \) although they need not \textit{a priori} coincide. Since \( B := u(A) = A - (d_A u)u^{-1} \in A^2_E \), we have
\[
u(\lambda) - A_0 = u(A - A_0)u^{-1} - (d_A u)u^{-1}.
\]

Our task, in essence, is to estimate the second term on the right above. Rewriting this equality gives a first-order, linear elliptic equation in \( u \) with \( L^2 \) coefficients:
\begin{equation}
d_A u = u(A - A_0) - (B - A_0)u.
\end{equation}

Let \( u_0 \in L^2_2(\mathfrak{gl}(E)) \) be the \( L^2 \) orthogonal projection of \( u \in G^3_E \subset L^2(\mathfrak{gl}(E)) \) onto \( \text{Ker}(d_{A_0} | L^2_2) \), so \( u = u_0 + \gamma \), where \( \gamma \in \text{Ker} d_{A_0} \subset \Omega^0(\mathfrak{gl}(E)) \). Thus, as \( d^*_A(A - A_0) = 0 \) and \( d_A u = d_A u_0 \), we see that
\[
d^*_A d_A u_0 = - (d_A u \wedge (A - A_0)) + u(d^*_A(A - A_0)
- (d^*_A(B - A_0))u - *(B - A_0) \wedge d_A u)
= - (d_A u_0 \wedge (A - A_0)) - (d^*_A(B - A_0))u
- *(B - A_0) \wedge d_A u_0).
\]

Therefore, using the bound \( \|u\|_{C^0} \leq 1 \) for any \( u \in G^3_E \) (as the representation for \( G \) is orthogonal), we have
\[
\| \Delta^0_{A_0} u_0 \|_{L^1^2} \leq \|d_A u_0 \|_{L^2^4} \|A - A_0\|_{L^2^4} + \|d^*_A(B - A_0)\|_{L^1^2} \|u\|_{C^0}
+ \|B - A_0\|_{L^2^4} \|d_A u_0\|_{L^2^4}
\leq C \|A - A_0\|_{L^2^4} + \|B - A_0\|_{L^2^4} \|d^*_A d_A u_0\|_{L^1^2}
+ \|d^*_A(B - A_0)\|_{L^1^2},
\]
where \( C = cK_0 \). Now \( \|B - A_0\|_{L^2^4} \leq c \|B - A_0\|_{L^2^4} \) via the embedding \( L^2_1 \subset L^2^4 \) of Lemma 4.1. For either \( \text{dist}_{L^2^4}([A], [A_0]) \leq \frac{1}{4}C^{-1} \) or \( \text{dist}_{L^1^2}([A], [A_0]) \leq \frac{1}{4}C^{-1} \) and \( \|A - A_0\|_{L^2^4} \leq \frac{1}{4}C^{-1} \), rearrangement yields
\begin{equation}
\| \Delta^0_{A_0} u_0 \|_{L^1^2} \leq 2 \|d^*_A(B - A_0)\|_{L^1^2}.
\end{equation}

On the other hand, from Lemma 5.6 we have
\begin{align}
\|u_0\|_{L^2^2} &\leq C \|\Delta^0_{A_0} u_0\|_{L^2^2},
\|u_0\|_{L^1^2} &\leq C \|\Delta^0_{A_0} u_0\|_{L^1^2},
\end{align}
where \( C = cK_0 \) and the second bound follows from the embedding \( L^2_2 \subset L^4_1 \). So, combining (6.10) and (6.11) yields:

\[
\| u_0 \|_{L^2_2, A_0} \leq C\| d^*_A (B - A_0) \|_{L^2_2},
\]

(6.12)

\[
\| u_0 \|_{L^4_1, A_0} \leq C\| d^*_A (B - A_0) \|_{L^2_2}.
\]

From (6.14) and (6.12), we see that

\[
\| A - A_0 \|_{L^{2s,4}} \leq \| u^{-1} (B - A_0) u \|_{L^{2s,4}} + \| u^{-1} d_A u_0 \|_{L^{2s,4}},
\]

(6.13)

\[
\| A - A_0 \|_{L^2_1, A_0} \leq \| u^{-1} (B - A_0) u \|_{L^2_1, A_0} + \| u^{-1} d_A u_0 \|_{L^2_1, A_0}.
\]

Consequently, using \( d_A u = d_A u_0 \) and (6.9) rewritten in the form,

\[
u^{-1} (B - A_0) u = (A - A_0) - u^{-1} d_A u_0,
\]

we obtain

\[
\| A - A_0 \|_{L^{2s,4}} \leq \| B - A_0 \|_{L^{2s,4}} + \| d_A u_0 \|_{L^{2s,4}},
\]

(6.14)

\[
\| A - A_0 \|_{L^2_1, A_0} \leq \| u^{-1} (B - A_0) u \|_{L^2_1, A_0} + \| u^{-1} d_A u_0 \|_{L^2_1, A_0}.
\]

From (6.14) and (6.12), we see that

\[
\| A - A_0 \|_{L^{2s,4}} \leq \| B - A_0 \|_{L^{2s,4}} + \| d_A u_0 \|_{L^{2s,4}}
\]

\[
\leq \text{dist}_{L^{2s,2}} ([A], [A_0]) + C\| d^*_A (B - A_0) \|_{L^{2s,2}}
\]

\[
\leq (1 + C) \text{dist}_{L^{2s,2}} ([A], [A_0]),
\]

giving the desired \( L^{2s,4} \) estimate for \( A - A_0 \).

Considering the first term in (6.13), we have

\[
\nabla_A (u^{-1} (B - A_0) u) = -u^{-1} (\nabla_A u) u^{-1} \otimes (B - A_0) u
\]

\[
+ u^{-1} (\nabla_A (B - A_0) u) u + u^{-1} (B - A_0) \otimes \nabla_A u,
\]

and so applying (6.12), noting that \( \nabla_A u = \nabla_A u_0 \) and \( \| u \|_{C^0} \leq 1 \), we have

\[
\| \nabla_A (u^{-1} (B - A_0) u) \|_{L^2} \leq \| \nabla_A u_0 \|_{L^4} \| B - A_0 \|_{L^4} + \| \nabla_A (B - A_0) \|_{L^2}
\]

\[
\leq C \text{dist}_{L^{2s,2}} ([A], [A_0]) + \text{dist}_{L^{2s,2}} ([A], [A_0]).
\]

Thus, if \( \text{dist}_{L^{2s,2}} ([A], [A_0]) \leq \frac{1}{4} C^{-1} \), say, we obtain

\[
(6.16) \quad \| \nabla_A (u^{-1} (B - A_0) u) \|_{L^2} \leq 2 \text{dist}_{L^{2s,2}} ([A], [A_0]).
\]

Similarly, considering the second term in (6.13), we have

\[
\nabla_A (u^{-1} d_A u_0) = -u^{-1} (\nabla_A u) u^{-1} \otimes d_A u + u^{-1} \nabla_A d_A u
\]

and therefore, by (6.12), we see that

\[
\| \nabla_A (u^{-1} d_A u_0) \|_{L^2} \leq \| \nabla_A u_0 \|_{L^4}^2 + \| \nabla^2_A u_0 \|_{L^2}
\]

\[
\leq C \text{dist}_{L^{2s,2}} ([A], [A_0]) \left( 1 + C \text{dist}_{L^{2s,2}} ([A], [A_0]) \right).
\]
Taking the \( L^2 \) norm of (6.13) and applying (6.12) to estimate the second term gives
\[
\|A - A_0\|_{L^2} \leq \|B - A_0\|_{L^2} + \|d_{A_0} u_0\|_{L^2}
\]
\[
\leq \text{dist}_{L^1,A_0}^{L^2} ([A], [A_0]) + C \|d_{A_0}^* (B - A_0)\|_{L^1,A_0},
\]
and so
\[
\|A - A_0\|_{L^2} \leq (1 + C) \text{dist}_{L^1,A_0}^{L^2} ([A], [A_0]).
\]
Combining the estimates (6.15), (6.16), (6.17), and (6.18) yields
\[
\|A - A_0\|_{L^1,A_0} \leq 2(1 + C) \text{dist}_{L^1,A_0}^{L^2} ([A], [A_0]),
\]
giving the desired \( L^2_{1,A_0} \) estimate for \( A - A_0 \).

\[\square\]

Naturally, a comparison—going in the other direction—of distances from \( A_0 \) in the Coulomb slice in \( A^k_E \) through \( A_0 \) and gauge-invariant distances in \( B^k_E \) from the point \([A_0]\) is elementary: If \( A \in S_{A_0} \) and \( \|A - A_0\|_{L^2_{k,A_0}} < \delta \), say, then Lemma 4.1 implies that
\[
\text{dist}_{L^1,A_0}^{L^2} ([A], [A_0]) \leq C \|A - A_0\|_{L^2_{k,A_0}} < c\delta,
\]
where \( c \) is positive constant \( c(X,g,k) \). The observation is used in concluding
that \( \mathcal{B}^* \) and \( \mathcal{B} \) are open subspaces of \( B^{1,\frac{1}{2},2}_{[A_0]}(\varepsilon_1) \) and \( B^{1,\frac{1}{2},2}_{[A_0]}(\varepsilon_1) \), respectively:

**Lemma 6.7.** Let \( (X,g) \) be a closed, smooth, Riemannian four-manifold and let \( G \) be a compact Lie group. Then there is a positive constant \( z \) with the following significance. Let \( K_0 = (1 + \nu_0[A_0]^{-1})(1 + \|F_{A_0}\|_{L^2}) \). If \( \varepsilon_1 < zK_0^{-2}(1 + \nu_0[A_0]^{-1/2})^{-1} \), then:

- \( \mathcal{B}^* \) is an open subspace of \( B^{1,\frac{1}{2},2}_{[A_0]}(\varepsilon_1) \);
- \( \mathcal{B} \) is an open subspace of \( B^{1,\frac{1}{2},2}_{[A_0]}(\varepsilon_1) \).

**Proof.** It suffices to consider the second assertion, as the argument for the first is identical. Suppose \([A] \in \mathcal{B}\) and that \( A \in A^k_E \) is a representative satisfying the defining conditions for \( \mathcal{B} \). Then \( A \) satisfies \( d^*_{A_0} (A - A_0) = 0 \) and the estimates
\[
\|A - A_0\|_{L^{2,4}} \leq c_0 K_0 \text{dist}_{L^1,A_0}^{L^2} ([A], [A_0]), \leq c_0 K_0 \varepsilon_1
\]
\[
\|A - A_0\|_{L^1,A_0} \leq c_0 K_0 \text{dist}_{L^1,A_0}^{L^2} ([A], [A_0]) \leq c_0 K_0 \varepsilon_1,
\]
while \(\|A - A_0\|_{L^{2,4}} \leq c_1\|A - A_0\|_{L^2_{1,A_0}}\) via the Sobolev embedding \(L^2_{1,A_0} \subset L^{2,4}\) and Kato’s inequality. Consequently, if \(c_1c_0K_0\varepsilon_1 \leq \frac{1}{2}\varepsilon_0\), then \(A \in \mathcal{B}^4_{A_0}(\varepsilon_1) \subset \mathcal{A}_E^k\) and we see that

\[
\bar{B}^{1,*2}_{[A_0]}(\varepsilon_1) \subset \pi\left(\mathcal{B}^4_{[A_0]}(\varepsilon_0)\right).
\]

Lemma 3.6 implies that the map \(\pi : \mathcal{B}^4_{[A_0]}(\varepsilon_0)/\text{Stab}_{A_0} \rightarrow \mathcal{B}^k_E\) given by \(A' \mapsto [A']\) is a local homeomorphism onto its image \(\pi(\mathcal{B}^4_{[A_0]}(\varepsilon_0))\) for any \(\varepsilon_0 < z(1 + v_0[A_0]^{-1/2})^{-1}\). In particular, if \(A' \in \mathcal{B}^4_{[A_0]}(\varepsilon_0)\) and \(\|A' - A\|_{L^2_{k,A_0}} < \delta\), then \(A' \in \mathcal{B}^{1,*2}_{[A_0]}(\varepsilon_1) \subset \bar{B}^{1,*2}_{[A_0]}(\varepsilon_1)\) for small enough \(\delta\).

The embedding \(L^4_1 \subset L^{2,4}\) and Lemma 6.6 imply that if \(\|A' - A_0\|_{L^4_{1,A_0}} \leq zK_0^{-1}\) and \(\text{dist}_{L^4_{1,A_0}}([A'], [A_0]) \leq zK_0^{-1}\), then

\[
\|A' - A_0\|_{L^{2,4}} \leq cK_0 \text{dist}_{L^4_{1,A_0}}([A'], [A_0]) \leq cK_0\varepsilon_1,
\]

\[
\|A' - A_0\|_{L^4_{1,A_0}} \leq cK_0 \text{dist}_{L^4_{1,A_0}}([A'], [A_0]) \leq cK_0\varepsilon_1.
\]

These inequalities are satisfied by \(A\); moreover \(\text{dist}_{L^4_{1,A_0}}([A], [A_0]) \leq \varepsilon_1\) and \(\|A - A_0\|_{L^4_{1,A_0}} \leq c_0K\varepsilon_1\). Require that \(\varepsilon_1 \leq \frac{1}{2}zK_0^{-1}\) and \(c_0K\varepsilon_1 \leq \frac{1}{2}zK_0^{-1}\), so \(\varepsilon_1 \leq \frac{1}{2}\varepsilon_0\min\{1, c_0\}K_0^{-2}\). Hence, if \(A'\) is \(L^2_{k,A_0}\)-close enough to \(A\) (where \(k \geq 2\)), we can ensure \([A']\) obeys the last two defining conditions for \(\mathfrak{B}\) and so \([A'] \in \mathfrak{B}\). Thus, \(\mathfrak{B} \subset \bar{B}^{1,*2}_{[A_0]}(\varepsilon_1)\) is open, as desired. \(\square\)

We can now conclude the proof of Theorem 6.1:

**Proof of Theorem 6.1.** Lemmas 6.5 and 6.7 imply that \(\mathfrak{B}\) is an open and closed subset of the connected space \(\bar{B}^{1,*2}_{[A_0]}(\varepsilon_1)\), so \(\mathfrak{B} = \bar{B}^{1,*2}_{[A_0]}(\varepsilon_1)\). Similarly for \(\mathfrak{B}^*\) and \(\bar{B}^{1,*2}_{[A_0]}(\varepsilon_1)\) and hence the result follows. \(\square\)

Similarly, we conclude the proofs of our main theorems:

**Proof of Theorem 1.1.** Given Theorem 6.1, the only assertion left unaccounted for is the uniqueness of the gauge transformation \(u \in \mathcal{G}_{E}^{k+1}\), modulo \(\text{Stab}_{A_0}\). But this follows from Lemma 3.7, just as in the paragraph immediately following the proof of that lemma. \(\square\)

**Proof of Theorem 1.2.** For the proof of Assertion (1), see the first paragraph of Section 1.3. The first inclusion in Assertion (2), namely \(B^{1,*2}_{[A_0]}(\varepsilon_1) \subset \bar{B}^{1,*2}_{[A_0]}(c_1\varepsilon_1)\), follows from the definition (1.3) of the two distance functions defining the balls (1.4) and the Sobolev embedding \(L^2_1 \subset L^{2,4}\) in Lemma 4.1. The second inclusion in Assertion (2), namely \(B^{1,*2}_{[A_0]}(c_1\varepsilon_1) \subset \pi(\mathcal{B}^4_{A_0}(c_2K_0\varepsilon_1))\),
follows from the definition (1.4) of these balls and Assertion (1) in Theorem 1.1.

7. Critical-exponent Sobolev norms and the group of gauge transformations.

We now define an $L_2^{\sharp,2}$ space of gauge transformations, by analogy with the definition of $G_{k+1}^k$ when $k \geq 2$, and set

$$G_{k+1,\sharp}^k \subseteq \{ u \in L_2^{\sharp,2}(g(E)) : u \in G \text{ a.e.} \} \subset L_k^2(\mathfrak{g}(E)).$$

It is not entirely clear a priori that $G_{k+1,\sharp}^k$ is a Banach Lie group. In the case of its counterpart, $G_{k+1}^k$, the manifold structure follows from the fact that the exponential map

$$\text{Exp} : T_{\text{id}_E} G_E = \Omega^0(\mathfrak{g}_E) \to G_E, \quad \zeta \mapsto \text{Exp} \zeta,$$

extends to a smooth map $\text{Exp} : L_{k+1}^2(\mathfrak{g}_E) \to L_k^2(\mathfrak{g}_E)$ and defines a system of smooth coordinate charts for $G_{k+1}^k$. Here, $\text{Exp}$ is defined pointwise at $u \in G_E$ for $\zeta \in T_{\text{id}_E} G_E$ by setting

$$(\text{Exp}_u \zeta)(x) := \exp_{u(x)}(\zeta(x)), \quad x \in X,$$

where $\exp : \mathfrak{g} \to G$ is the usual, $C^\infty$ exponential map for the Lie group $G$ on the right-hand side [10, Appendix A].

To verify that $G_{k+1,\sharp}^k$ is in fact a Banach Lie group we will need estimates for the covariant derivatives of the exponential map. The estimates below follow by reworking the usual proof of the Sobolev lemma for left composition of Sobolev sections by smooth vector bundle maps [15, Lemma 9.9]. The difference here is that we keep track of the dependence of the constants on the geometric data: This precision is required for the implicit function argument in the next section in order to complete the proof of our slice theorem.

For $\chi, \zeta, \xi \in \Omega^0(\mathfrak{g}_E)$, the differentials

$$(D \text{Exp})_\chi : \Omega^0(\mathfrak{g}_E) \to T_{\text{Exp} \chi} G_E, \quad \zeta \mapsto (D \text{Exp})_\chi \zeta,$$

$$(D^2 \text{Exp})_{\chi,\zeta} : \Omega^0(\mathfrak{g}_E) \to T_{\text{Exp} \chi} G_E, \quad \xi \mapsto (D^2 \text{Exp})_{\chi,\zeta} \xi,$$

are defined pointwise by setting

$$(D \text{Exp})_\chi \zeta|_x = (D \exp)_\chi(x) \zeta(x),$$

$$(D^2 \text{Exp})_{\chi,\zeta} \xi|_x = (D \exp)_{\chi(x),\zeta(x)} \xi(x),$$

for any $x \in X$. When writing the differential $(D^2 \text{Exp})_{\chi,\zeta}$ above, we have identified $T_{(D \exp)_\chi}(T_{\text{Exp} \chi} G_E)$ with $T_{\text{Exp} \chi} G_E$. 

The maps $\left(D\operatorname{Exp}\right)_\chi : \Omega^0(\mathfrak{g}_E) \to \Omega^0(\mathfrak{g}_E)$ and $\left(D^2\operatorname{Exp}\right)_{\chi,\zeta} : \Omega^0(\mathfrak{g}_E) \to \Omega^0(\mathfrak{g}_E)$ extend linearly to maps

\[
\left(D\operatorname{Exp}\right)_\chi : C^\infty(\otimes^\ell (T^*X) \otimes \mathfrak{g}_E) \to C^\infty(\otimes^\ell (T^*X) \otimes \mathfrak{g}_E),
\]

\[
\left(D^2\operatorname{Exp}\right)_\chi : C^\infty(\otimes^\ell (T^*X) \otimes \mathfrak{g}_E) \to C^\infty(\otimes^\ell (T^*X) \otimes \mathfrak{g}_E),
\]

for $\ell \geq 1$, by setting

\[
\left(D\operatorname{Exp}\right)_\chi(\theta \otimes \zeta) = \theta \otimes \left(D\operatorname{Exp}\right)_\chi \zeta,
\]

\[
\left(D^2\operatorname{Exp}\right)_{\chi,\zeta}(\theta \otimes \zeta) = \theta \otimes \left(D^2\operatorname{Exp}\right)_{\chi,\zeta} \zeta,
\]

for $\theta \in \otimes^\ell (T^*X)$ and $\zeta \in \Omega^0(\mathfrak{g}_E)$. As usual, we embed $\mathcal{G}_E \subset \Omega^0(\mathfrak{g}_E)$ in order to compute the covariant derivatives of sections $u \in \mathcal{G}_E$.

**Lemma 7.1.** Let $G$ be a compact Lie group. Then there is a positive constant $c(G)$ with the following significance. Let $X$ be a closed, smooth, Riemannian four-manifold. If $A$ is a $C^\infty$ connection on a $G$ bundle $E$, and $\chi \in \Omega^0(\mathfrak{g}_E)$, then we have pointwise bounds:

\[(1) \quad |\nabla_A e^\chi| \leq |\nabla_A \chi| + c|\chi||\nabla_A \chi|,
\]

\[(2) \quad |\nabla_A^2 e^\chi| \leq c(|\chi| + |\nabla_A \chi|)|\nabla_A \chi| + c(1 + |\chi|)|\nabla_A^2 \chi|,
\]

\[(3) \quad |\nabla_A^* \nabla_A e^\chi| \leq c(|\chi| + |\nabla_A \chi|)|\nabla_A \chi| + c(1 + |\chi|)|\nabla_A^* \nabla_A \chi|.
\]

**Proof.** We have

\[\nabla_A e^\chi = \nabla_A (\operatorname{Exp} \chi) = \left(D\operatorname{Exp}\right)_\chi \circ dA \chi \in \Omega^1(\mathfrak{g}_E).\]

Since $(D\operatorname{exp})_0 = \operatorname{id}_E$ and $\operatorname{exp} : \mathfrak{g} \to G$ is analytic, we have the pointwise bound $|(D\operatorname{exp})_\chi(x) - \operatorname{id}_E| \leq c(G)|\chi(x)|$ and thus a pointwise bound

\[|(D\operatorname{Exp})_\chi - \operatorname{id}_E| \leq c|\chi|,
\]

noting that $(D\operatorname{Exp})_0 = \operatorname{id}_E$. Therefore, we have

\[|\nabla_A e^\chi| \leq |\nabla_A \chi| + c|\chi||\nabla_A \chi|,
\]

which gives the first assertion.

Define $\Phi(\chi, \zeta) = \left(D\operatorname{Exp}\right)_\chi(\zeta) \in \Omega^1(\mathfrak{g}_E)$, for $\chi \in \Omega^0(\mathfrak{g}_E)$ and $\zeta \in \Omega^1(\mathfrak{g}_E)$, noting that $\Phi$ is nonlinear in $\chi$, but linear in $\zeta$. Thus,

\[\nabla_A^2 u = (D_1 \Phi)_{\chi, \nabla_A \chi} (\nabla_A \chi) + (D_2 \Phi)_{\chi} (\nabla_A^2 \chi),\]

where $D_i \Phi, i = 1, 2$, denote the partial derivatives of $\Phi$ with respect to first and second variables. Since $(D \Phi)_{\chi, 0} = (D^2 \operatorname{Exp})_{\chi, 0} = \operatorname{id}_E$, as $(D^2 \operatorname{exp})_{\chi, 0} = \operatorname{id}_E$ and $\operatorname{exp} : \mathfrak{g} \to G$ is analytic we have the pointwise bound

\[|\nabla_A^2 u| \leq c(|\chi| + |\nabla_A \chi|)|\nabla_A \chi| + c(1 + |\chi|)|\nabla_A^2 \chi|,
\]

giving the second assertion. Similarly, as $* \Phi(\chi, \zeta) = \Phi(\chi, * \zeta)$ and $\nabla_A^* \nabla_A u = - * \nabla_A * \nabla_A u$, we have

\[|\nabla_A^* \nabla_A u| \leq c(|\chi| + |\nabla_A \chi|)|\nabla_A \chi| + c(1 + |\chi|)|\nabla_A^* \nabla_A \chi|,
\]
giving the third assertion.

The preceding pointwise bounds for $\nabla_A u$, $\nabla^2_A u$, and $\nabla_A \nabla_A u$ yield the following estimates for the exponential map:

**Lemma 7.2.** Continue the hypotheses of Lemma 7.1. If $k \geq 2$ is an integer (so $L^k_{k+1} < C^0$), $A$ is an $L^k_2$ connection on a $G$ bundle $E$, and $\chi \in L^2_{k+1}(\mathfrak{g}_E)$, then $e^\chi \in G^k_{E+1}$ satisfies

\[
\begin{align*}
(1) & \quad \|\nabla_A e^\chi\|_{L^2(X)} \leq \|\nabla_A \chi\|_{L^2(X)} + c\|\chi\|_{C^0(X)}\|\nabla_A \chi\|_{L^2(X)}, \\
(2) & \quad \|\nabla^2_A e^\chi\|_{L^2(X)} \leq \|\nabla_A \chi\|_{L^2(X)} + c\|\chi\|_{C^0(X)}\|\nabla_A \chi\|_{L^2(X)}, \\
(3) & \quad \|\nabla_A \nabla_A e^\chi\|_{L^2(X)} \leq c\|\chi\|_{C^0(X)}\|\nabla_A \chi\|_{L^2(X)} + \|\nabla_A \chi\|_{L^2(X)}^2 \\
& \quad + c(1 + \|\chi\|_{C^0(X)}) \|\nabla^2_A \chi\|_{L^2(X)}, \\
(4) & \quad \|\nabla_A \nabla_A e^\chi\|_{L^2(X)} \leq c\|\chi\|_{C^0(X)}\|\nabla_A \chi\|_{L^2(X)} + \|\nabla_A \chi\|_{L^2(X)}^2 \\
& \quad + c(1 + \|\chi\|_{C^0(X)}) \|\nabla^2_A \chi\|_{L^2(X)}.
\end{align*}
\]

The bounds (1)-(4) continue to hold for $\chi \in L^2_{\frac{5}{2}}(\mathfrak{g}_E) \subset C^0(\mathfrak{g}_E)$, with $A$ an $L^2_{\frac{5}{2}}$ connection on $E$, and $\text{Exp} : \Omega^0(\mathfrak{g}_E) \rightarrow G^k_E$ extends to a continuous map $\text{Exp} : L^2_{\frac{5}{2}}(\mathfrak{g}_E) \rightarrow G^k_{E+1}$.

Let $A^1_{\frac{5}{2}} = A_0 + L^2_{\frac{5}{2}}(\Lambda^1 \otimes \mathfrak{g}_E)$, for any $C^\infty$ reference connection $A_0$ on $E$. Recall that we have an embedding $L^2_{\frac{5}{2}}(\mathfrak{g}_E) \subset C^0(\mathfrak{g}_E)$ and that the space $L^2_{\frac{5}{2}}(\mathfrak{g}_E)$ is an algebra, while $L^2_{\frac{5}{2}}(\Lambda^1 \otimes \mathfrak{g}_E)$ and $L^2_{\frac{5}{2}}(\Lambda^1 \otimes \mathfrak{g}_E)$ are $L^2_{\frac{5}{2}}(\mathfrak{g}_E)$-modules. Therefore, the proofs of Propositions (A.2) and (A.3) in [10] extend to give the following analogue for $G^k_{E+1}$ in place of $G^k_{E+1}$:

**Lemma 7.3.** Let $X$ be a closed Riemannian four-manifold and let $E$ be a Hermitian vector bundle over $X$. Then the following hold.

\[
\begin{align*}
(1) & \quad \text{The space } G^k_{E+1} \text{ is a Banach Lie group with Lie algebra } T_{id_E} G^k_{E+1} = L^2_{\frac{5}{2}}(\mathfrak{g}_E); \\
(2) & \quad \text{The action of } G^k_{E+1} \text{ on } A^1_{\frac{5}{2}} \text{ and } A^1_{\frac{5}{2}} \text{ is smooth;} \\
(3) & \quad \text{For } A \in A^1_{\frac{5}{2}}, \text{ the differential, at the identity } id_E \in G^k_{E+1}, \text{ of the map } G^k_{E+1} \rightarrow A^1_{\frac{5}{2}} \text{ given by } u \mapsto u(A) = A - (d_A u) u^{-1} \text{ is } \zeta \mapsto -d_A \zeta \text{ as a map } L^2_{\frac{5}{2}}(\mathfrak{g}_E) \rightarrow L^1_{\frac{5}{2}}(\Lambda^1 \otimes \mathfrak{g}_E), \text{ and similarly for } A \in A^1_{\frac{5}{2}}.
\end{align*}
\]

8. Existence of gauge transformations via the inverse function theorem.

Our goal in this section is to give an alternative, “direct” proof of Theorem 6.1 via the inverse function theorem. A direct argument—due to our overarching constraint that the constants given there ultimately depend at most on the $L^2$ norm of the curvature and the least positive eigenvalue
Lemma 8.1. Let \( X \) be a closed, Riemannian four-manifold. Then there is a constant \( \varepsilon \) with the following significance. Let \( G \) be compact Lie group and let \( k \geq 2 \) be an integer. Suppose that \( A_0 \) is an \( L^2_0 \) connection on a \( G \) bundle \( E \), that \( a \in L^2_k(\Lambda^1 \otimes g_E) \) and \( \chi \in L^2_2(g_E) \), and that \( u = e^\chi \) is a solution to
\[
d^{\ast}_{A_0}(d_{A_0}u)u^{-1} - uu^{-1} = 0.
\]
If \( \|d_{A_0}u\|_{L^4} < \varepsilon \) then \( \chi \in L^2_{k+1}(g_E) \) and \( u = e^\chi \in \mathcal{G}^{k+1}_E \).

Proof. Differentiation and right multiplication by \( u \) yields
\[
(8.1) \quad d^{\ast}_{A_0}d_{A_0}u + *(d_{A_0}u \wedge u^{-1}d_{A_0}u) + *(d_{A_0}u \wedge a) + u(d^{\ast}_{A_0}a
\]
\[
+ *(ua \wedge u^{-1}d_{A_0}u) = 0.
\]

From Lemma 4.2 we know that \( u \in C^0 \cap L^2 \) and so the last four terms in (8.1) are in \( L^2 \). Hence, \( d^{\ast}_{A_0}d_{A_0}u \) is in \( L^2 \) and so \( u \in L^2_p \) by elliptic regularity for \( d^{\ast}_{A_0}d_{A_0} \). The Sobolev embedding \( L^2_p \subset L^4 \) and multiplication \( L^4 \times L^q \to L^p \) for \( 2 \leq p < 4 \) and \( 1/p = 1/4 + 1/q \) (so \( 4 \leq q < \infty \)) now show that the last three terms in (8.1) are in \( L^p \), so the equation takes the simpler form
\[
(8.2) \quad d^{\ast}_{A_0}d_{A_0}u + *(d_{A_0}u \wedge u^{-1}d_{A_0}u) = v,
\]
where \( v \in L^p(g_E) \) is the tautologically defined right-hand side and \( u \in L^\infty \cap L^2 \). Setting \( b = d_{A_0}u \) and noting that \( d_{A_0}b = F_{A_0}u \), with \( F_{A_0} \in L^2_{k-1}(\Lambda^2 \otimes g_E) \subset L^2_4(\Lambda^2 \otimes g_E) \) and \( F_{A_0}u \in L^2_2(\Lambda^2 \otimes g_E) \). Thus, we may conveniently rewrite (8.2) as a first-order elliptic equation in \( b \in L^2_4(\Lambda^1 \otimes g_E) \),
\[
(8.3) \quad (d^{\ast}_{A_0} + d_{A_0})b + *(d_{A_0}b \wedge u^{-1}b) = v' \in L^p(g_E) \oplus L^p(\Lambda^2 \otimes g_E),
\]
where \( 2 < p < 4 \) and \( v' = F_{A_0}u + v \). Finally, (8.3) can be rewritten as a local equation by writing \( A_0 = \Gamma + a_0 \), where \( \Gamma \) is the product connection in a local trivialization for \( E \) over a small ball \( U \subset X \). Thus, the operator \( d^{\ast}_{A_0} + d_{A_0} \) is replaced by \( d^{\ast} + d \) in (8.2) and the additional terms are absorbed into the \( L^p \) inhomogeneous term \( v' \) to give:
\[
(8.4) \quad (d^{\ast} + d)b + *(d_{A_0}b \wedge u^{-1}b) = v'' \in L^p(U, g_E) \oplus L^p(U, \Lambda^2 \otimes g_E).
\]
This is a first-order, elliptic equation with a quadratic non-linearity and Proposition 3.10 in \([6]\) implies that the solution \(b = d_{A_0} u \in L^2(U, \Lambda^1 \otimes \mathfrak{g}_E)\) is necessarily in \(L^p(U') \otimes \mathfrak{g}_E\) for \(U' \subset U\), provided \(\|b\|_{L^p(U)} < \varepsilon(g, p, U)\), and so \(u \in L^p(U', \mathfrak{g}_E)\). In particular, we find that \(b \in L^p_1(X, \Lambda^1 \otimes \mathfrak{g}_E)\) and \(u \in L^p_0(X, \mathfrak{g}_E)\) for any \(2 < p < 4\), provided \(\|d_{A_0} u\|_{L^4} < \varepsilon(g, p, X)\). The bootstrapping argument of Proposition 3.3 in \([6]\) now implies that \(d_{A_0} u \in L^2_1(X, \Lambda^1 \otimes \mathfrak{g}_E)\). Thus \(u \in \mathcal{G}^{k+1}_E\) and \(\chi \in L^2_{k+1}(X, \mathfrak{g}_E)\), as desired. \(\square\)

We can now proceed to the main argument:

**Theorem 8.2.** Let \(X\) be a closed, Riemannian four-manifold and let \(G\) be compact Lie group. Then there are positive constants \(c, z\) with the following significance. Let \(E\) be a \(G\) bundle over \(X\) and suppose that \(A_0 \in \mathcal{A}^2_E\), let \(K_0[A_0] = (1 + \nu_0[A_0]^{-1})(1 + \|F_{A_0}\|_{L^2})\) and let \(\varepsilon_1\) be a constant satisfying

\[
0 < \varepsilon_1 \leq zK_0^{-2}.
\]

If \(A \in \mathcal{A}^2_E\) obeys \(\|A - A_0\|_{L^1, A_0} < \varepsilon_1\) then \(u \in \mathcal{G}^3_E\) exists such that

- \(d^*_{A_0}(u(A) - A_0) = 0\);
- \(\|u(A) - A_0\|_{L^2, A_0} \leq cK_0\|A - A_0\|_{L^1, A_0}\);
- \(\|u - \text{id}_E\|_{L^2, A_0} < cK_0\|A - A_0\|_{L^1, A_0}\).

**Proof.** The argument is broadly similar to that of Lemma 3.6, except that we can show \(\Psi\) is a diffeomorphism directly—rather than just a local diffeomorphism—using the slightly stronger norms now at our disposal. Moreover, on this occasion we seek precise bounds on the solutions so we keep track of the dependence of constants on the curvature \(F_{A_0}\) and the least positive eigenvalue \(\nu_0 = \nu_0[A_0]\) of the Laplacian \(\Delta_{A_0} = d^*_{A_0} d_{A_0}\).

Write \(A = A_0 + a\) and observe that

\[
u(A) - A_0 = A - A_0 - (d_A u)u^{-1} = uau^{-1} - (d_{A_0} u)u^{-1}.
\]

Recall that we have an \(L^2\)-orthogonal decomposition

\[
\Omega^0(\mathfrak{g}_E) = (\text{Ker} d_{A_0})^\perp \oplus \text{Ker} d_{A_0} = \text{Im} d^*_{A_0} \oplus \text{Ker} d_{A_0},
\]

and that \(d^*_{A_0} : L^2_1(\Lambda^1 \otimes \mathfrak{g}_E) \to L^2(\mathfrak{g}_E)\) has closed range; this gives

\[
L^2_{2, A_0}(\mathfrak{g}_E) = \left(\text{Ker} d_{A_0} |_{L^2_{2, A_0}}\right)^\perp \oplus \text{Ker} d_{A_0} |_{L^2_{2, A_0}} = \left(\text{Ker} d^*_{A_0} |_{L^2_{2, A_0}}\right)^\perp \oplus \left(\text{Im} d^*_{A_0} |_{L^2_{1, A_0}}\right).
\]

We have a similar \(L^2\)-orthogonal decomposition

\[
\Omega^1(\mathfrak{g}_E) = \text{Im} d_{A_0} \oplus \text{Ker} d^*_{A_0} = (\text{Ker} d^*_{A_0})^\perp \oplus \text{Ker} d^*_{A_0},
\]
and $d_{A_0} : L^2_1(\mathfrak{g}_E) \to L^2(\Lambda^1 \otimes \mathfrak{g}_E)$ has closed range; this leads to the $L^2$-orthogonal decomposition

$$L^{\sharp,2}_{1,A_0}(\Lambda^1 \otimes \mathfrak{g}_E) = \left( \text{Im} \ d_{A_0} \big|_{L^2_{1,A_0}} \right) \oplus \left( \text{Ker} \ d_{A_0}^* \big|_{L^{\sharp,2}_{1,A_0}} \right),$$

$$= \left( \text{Ker} \ d_{A_0}^* \big|_{L^{\sharp,2}_{1,A_0}} \right) \perp \left( \text{Ker} \ d_{A_0} \big|_{L^{\sharp,2}_{1,A_0}} \right).$$

We now define a map

$$(8.5) \quad \Psi : \left( \text{Ker} \left( d_{A_0} \big|_{L^{\sharp,2}_{1,A_0}} \right) \right) \perp \text{Ker} \left( d_{A_0}^* \big|_{L^{\sharp,2}_{1,A_0}} \right) \to L^{\sharp,2}_{1,A_0}(\Lambda^1 \otimes \mathfrak{g}_E),$$

$$(\chi, a) \mapsto uau^{-1} - (d_{A_0}u)u^{-1},$$

where $u = e^x$ and the differential at $(\chi, a)$ given by

$$(8.6) \quad (D\Psi)_{(\chi, a)} : \left( \text{Ker} \left( d_{A_0} \big|_{L^{\sharp,2}_{1,A_0}} \right) \right) \perp \text{Ker} \left( d_{A_0}^* \big|_{L^{\sharp,2}_{1,A_0}} \right) \to L^{\sharp,2}_{1,A_0}(\Lambda^1 \otimes \mathfrak{g}_E),$$

$$(\zeta, b) \mapsto u(-d_A \otimes i)u^{-1}(\zeta, b) = u(-d_A\zeta + b)u^{-1},$$

since $(D\Psi)_{(0,a)}(\zeta, b) = -d_A\zeta + b$ and $\Psi$ is $G_E$-equivariant. Moreover, we have

$$(8.7) \quad (D^2\Psi)_{(\chi, a)}((\zeta, b), (\eta, \alpha)) = u[\eta, -d_A\zeta + b]u^{-1} + u[\alpha, \zeta]u^{-1},$$

for $(\zeta, b), (\eta, \alpha) \in \left( \text{Ker} \left( d_{A_0} \big|_{L^{\sharp,2}_{1,A_0}} \right) \right) \perp \text{Ker} \left( d_{A_0}^* \big|_{L^{\sharp,2}_{1,A_0}} \right)$.

We now verify that the conditions of the inverse function theorem (Theorem 3.2) hold for suitable constants $K$ and $\delta$. The operator

$$d_{A_0} : \left( \text{Ker} \left( d_{A_0} \big|_{L^{\sharp,2}_{1,A_0}} \right) \right) \perp \left( \text{Ker} \left( d_{A_0}^* \big|_{L^{\sharp,2}_{1,A_0}} \right) \right) \perp$$

has a two-sided inverse

$$G_{A_0}^0 d_{A_0}^* : \left( \text{Ker} \left( d_{A_0}^* \big|_{L^{\sharp,2}_{1,A_0}} \right) \right) \perp \left( \text{Ker} \left( d_{A_0} \big|_{L^{\sharp,2}_{1,A_0}} \right) \right) \perp.$$

Indeed, for $b \in \left( \text{Ker} \left( d_{A_0}^* \big|_{L^{\sharp,2}_{1,A_0}} \right) \right) \perp$, we have

$$\|G_{A_0}^0 d_{A_0}^* b\|_{L^{\sharp,2}_{1,A_0}} \leq c_0 K_0 \|d_{A_0}^* b\|_{L^{\sharp,2}_{1,A_0}} \leq c_0 K_0 \|b\|_{L^{\sharp,2}_{1,A_0}},$$

and so $G_{A_0}^0 d_{A_0}^*$ has Hom$(L^{\sharp,2}_{1,A_0}, L^{\sharp,2}_{1,A_0})$ operator norm bound

$$\|G_{A_0}^0 d_{A_0}^*\| \leq c_0 K_0.$$

In particular, we see that $(D\Psi)_{(0,0)}^{-1} = G_{A_0}^0 d_{A_0}^* \oplus \text{id}$ satisfies

$$(8.8) \quad \|(D\Psi)_{(0,0)}^{-1}\| \leq c_0 K_0$$

the first of the conditions we need to verify for $(D\Psi)_{(0,0)}$ in order to apply the inverse function theorem.
It remains to compare \((D \Psi)_{(x,a)}\) and \((D \Psi)_{(0,0)}\) using the mean value theorem,

\begin{equation}
(D \Psi)_{(x,a)}(\zeta, b) - (D \Psi)_{(0,0)}(\zeta, b) = \int_0^1 (D^2 \Psi)_{(tx,ta)}((\zeta, b), (\chi, a)) \, dt.
\end{equation}

Thus, we need an estimate for \(D^2 \Psi\):

**Claim 8.3.** There is a universal polynomial function \(f(x, y)\), depending only on \((X, g)\) and \(G\), with \(f(0, 0) = 0\), such that the following holds. For any \(t \in [0, 1]\) we have:

\[
\|(D^2 \Psi)_{(tx,ta)}((\zeta, b), (\chi, a))\|_{L^1_{A_0}}
\leq f \left( \|\chi\|_{L^2_{A_0}}, \|a\|_{L^2_{A_0}}, \|\zeta\|_{L^2_{A_0}}, \|b\|_{L^2_{A_0}} \right).
\]

**Proof.** From (8.7) we have the \(L^2\) estimate

\[
\|(D^2 \Psi)_{(tx,ta)}((\zeta, b), (\chi, a))\|_{L^2}
\leq c \|\chi\|_{C^6} (\|d_{A_0} \zeta\|_{L^2} + \|a\|_{L^2} \|\zeta\|_{C^6} + \|b\|_{L^2} + \|a\|_{L^2} \|\zeta\|_{C^6}),
\]

and thus:

\begin{equation}
|(D^2 \Psi)_{(tx,ta)}((\zeta, b), (\chi, a))|_{L^2}
\leq c \left( \|\chi\|_{L^2_{A_0}} + \|a\|_{L^2} \|\chi\|_{L^2_{A_0}} + \|a\|_{L^2} \right) \left( \|\zeta\|_{L^2_{A_0}} + \|b\|_{L^2} \right).
\end{equation}

The \(L^2\) estimate of \(\nabla A_0 (D^2 \Psi)_{(tx,ta)}((\zeta, b), (\chi, a))\) is given by

\[
\|\nabla A_0(D^2 \Psi)_{(tx,ta)}((\zeta, b), (\chi, a))\|_{L^2}
\leq c \left( \|\nabla A_0 u\|_{L^4} \|\chi\|_{C^6} + \|\nabla A_0 \chi\|_{L^4} \right) (\|d_{A_0} \zeta\|_{L^4} + \|a\|_{L^4} \|\zeta\|_{C^6} + \|b\|_{L^4})
\]

\[
+ c \|\chi\|_{C^6} (\|\nabla^2 A_0 \zeta\|_{L^2} + \|\nabla A_0 a\|_{L^2} \|\zeta\|_{C^6} + \|a\|_{L^4} \|\nabla A_0 \zeta\|_{L^4} + \|\nabla A_0 b\|_{L^2})
\]

\[
+ c \|\nabla A_0 a\|_{L^4} \|a\|_{L^2} \|\zeta\|_{L^4} + c \|\nabla A_0 a\|_{L^2} \|\zeta\|_{C^6} + c \|a\|_{L^4} \|\nabla A_0 \zeta\|_{L^4},
\]

and hence, using Lemma 7.2 to estimate \(u = e^x\) in terms of \(\chi\),

\begin{equation}
\|\nabla A_0(D^2 \Psi)_{(tx,ta)}((\zeta, b), (\chi, a))\|_{L^2}
\leq f_1 \left( \|\chi\|_{L^2_{A_0}}, \|a\|_{L^2_{A_0}}, \|\zeta\|_{L^2_{A_0}}, \|b\|_{L^2_{A_0}} \right),
\end{equation}

where \(f_1(x, y)\) is a polynomial function with \(f_1(0, 0) = 0\).

Noting that \(d_{A_0} a = 0\), we have

\begin{equation}
d_{A_0} [a, \zeta] = d_{A_0} (a \zeta - \zeta a)
= (d_{A_0} a) \zeta - a \wedge d_{A_0} \zeta - *(d_{A_0} \zeta \wedge *a) - \zeta (d_{A_0} a)
= -a \wedge d_{A_0} \zeta - *(d_{A_0} \zeta \wedge *a),
\end{equation}
and similarly for \( d^*_A \) since \( d^*_b = 0 \). For any \( \beta \in L^2_1(\Lambda^1 \otimes g_E) \) we have
\[
(8.13) \quad d^*_A(u \beta u^{-1}) = - * d_A(u(\beta)u^{-1}) = - * (d_A u \wedge \beta u^{-1}) + u(d_A \beta)u^{-1} - u((\beta) \wedge u(d_A u)u^{-1}).
\]

Therefore, Equations (8.7), (8.12), and (8.13) and the estimates for \( u = e^x \) in Lemma 7.2 yield
\[
(8.14) \quad \|d^*_A(D^2\Psi)_{(\chi,a)}((\xi, b), (\chi, a))\|_{L^2} \leq \|d^*_A(u[\chi, -d_A \xi + b]u^{-1} + u[a, \xi]u^{-1})\|_{L^2} \\
\leq f_2 \left( \|\chi\|_{L^2_{2,A}}, \|a\|_{L^2_{1,A}} \right) \left( \|\xi\|_{L^2_{2,A}} + \|b\|_{L^2_{2,A}} \right),
\]
where \( f_2(x, y) \) is a polynomial function with \( f_2(0, 0) = 0 \). The claim now follows by combining (8.10), (8.11), and (8.14).

Therefore, from Claim 8.3 and (8.9) we have
\[
(8.15) \quad \|(D\Psi)_{(\chi,a)}(\xi, b) - (D\Psi)_{(0,0)}(\xi, b)\|_{L^1_{1,A}} \leq \int \left( \|\chi\|_{L^2_{2,A}}, \|a\|_{L^2_{1,A}} \right) \left( \|\xi\|_{L^2_{2,A}} + \|b\|_{L^2_{1,A}} \right).
\]

Consequently, with respect to the Hom \( \left( L^2_{2,A}, L^2_{1,A} \right) \) operator norm, (8.15) yields the bound
\[
(8.16) \quad \|(D\Psi)_{(\chi,a)} - (D\Psi)_{(0,0)}\| \leq \frac{1}{2} c_0^{-1} K_0^{-1},
\]
where \( c_0 K_0 = K \) is the constant of (8.8), provided \( (\chi, a) \) satisfies the constraint
\[
(8.17) \quad \|\chi\|_{L^2_{2,A}} + \|a\|_{L^2_{1,A}} \leq c_1 K_0^{-1} = \delta.
\]

Define balls centered at the origins in \( \left( \text{Ker} \left( d_A \right) \right)^\perp \) and \( \left( \text{Ker} \left( d_A^* \right) \right) \) by setting
\[
B^1_{0,2,2}(\delta) = \left\{ \chi \in \left( \text{Ker} \left( d_A \right) \right)^\perp : \|\chi\|_{L^2_{2,A}} < \delta \right\},
\]
\[
B^1_{1,2,2}(\delta) = \left\{ a \in \text{Ker} \left( d_A^* \right) : \|a\|_{L^2_{1,A}} < \delta \right\}.
\]

Hence, Theorem 3.2 implies that the map
\[
\Psi : B^1_{0,2,2}(\delta) \times B^1_{0,2,2}(\delta) \rightarrow A^{1,2}_E
\]
is injective, its image is an open subset of \( A^{1,2}_E \) and contains the ball \( B^1_{2}(\delta/(2K)) \), the inverse map \( \Psi^{-1} \) is a diffeomorphism from \( B^1_{2}(\delta/(2K)) \)
onto its image, and if \((\chi_1, A_1), (\chi_2, A_2)\) are points in \(B_0^{1,2,2}(\delta) \times B_0^{1,2,2}(\delta)\), then
\[
\|\chi_1 - \chi_2\|_{L^2, A_0}^{1,2} + \|A_1 - A_2\|_{L^1, A_0}^{1,2} \leq 2K\|u_1(A_1) - u_2(A_2)\|_{L^1, A_0}^{1,2},
\]
where \(u_i = e^{\chi_i}, i = 1, 2\). In particular, setting \((\chi_2, A_2 - A_0) = (0, 0)\), we see that if \(A\) is a point in \(A_E^{1,2,2}\) such that \(\|A - A_0\|_{L^1, A_0}^{1,2} < \delta/(2K)\), then there is a unique solution \((\chi, u^{-1}(A)) = \Psi^{-1}(A)\) in \(B_0^{1,2,2}(\delta) \times B_0^{1,2,2}(\delta)\). Here, \(u = e^\chi\) is a gauge transformation with \(\chi \in B_0^{1,2,2}(\delta)\) such that
\[
d_{A_0}^*(u^{-1}(A) - A_0) = 0, \tag{8.18}
\]
\[
\|\chi\|_{L^2, A_0}^{1,2} + \|u^{-1}(A) - A_0\|_{L^1, A_0}^{1,2} \leq 2K\|A - A_0\|_{L^1, A_0}^{1,2}.
\]
Lemma 7.2 implies that \(u = e^\chi\) satisfies
\[
\|u - \text{id}_E\|_{L^2, A_0}^{1,2} \leq f_3\left(\|\chi\|_{L^2, A_0}^{1,2}\right) \leq c\|\chi\|_{L^2, A_0}^{1,2} \leq c\delta, \tag{8.19}
\]
where \(f_3(x)\) is a polynomial with coefficients depending only on \((X, g)\) and \(G\) such that \(f_3(0) = 0\). Noting that \(K = c_0K_0, \delta = c_1K_0^{-1}\), and \(\delta/(2K) = \frac{1}{2}c_0c_1K_0^{-2}\), the desired estimates follow from (8.18) and (8.19). Finally, Lemma 8.1 implies that \(u \in G_E^{2}\) and this completes the proof of the theorem. \(\Box\)

While the \(L^2_1\) estimate of Theorem 8.2 suffices for most purposes, it is occasionally useful to have the weaker \(L^{2,4}\) bound at hand. Recall from Section 4 that we defined
\[
\|a\|_{L^{2,4}_1, A_0} = \|a\|_{L^{2,4}} + \|d_{A_0}^*a\|_{L^2, A_0}, \quad a \in \Omega^1(g_E).
\]
A slight modification of the proof of Theorem 8.2 yields:

**Theorem 8.4.** Continue the hypotheses of Theorem 8.2. Then for any \(A \in A_E^2\) such that \(\|A - A_0\|_{L^2, A_0}^{1,2} < \varepsilon_1\) there is a gauge transformation \(u \in G_E^2\) with the following properties:

- \(d_{A_0}^*(u(A) - A_0) = 0\);
- \(\|u(A) - A_0\|_{L^{2,4}} \leq cK_0\|A - A_0\|_{L^1, A_0}^{1,2}\);
- \(\|u - \text{id}_E\|_{L^2, A_0}^{1,2} < cK_0\|A - A_0\|_{L^1, A_0}^{1,2}\).

**Proof.** The first difference in the argument is that the map \(\Psi\) in (8.5) is replaced by
\[
\Psi : \left(\text{Ker}\left(d_{A_0}\right)_{L^2}\right) \oplus \text{Ker}\left(d_{A_0}^*\right)_{L^1} \to L^{2,2}_1, A_0(\Lambda^1 \otimes g_E), \quad (\chi, a) \mapsto uau^{-1} - (d_{A_0}u)u^{-1}. \tag{8.20}
\]
The second difference is that Claim 8.3 is replaced by:
Claim 8.5. There is a universal polynomial function \( f(x,y) \), depending only on \((X,g)\) and \(G\), with \( f(0,0) = 0 \), such that the following holds. For any \( t \in [0,1] \) we have:

\[
\| (D^2 \Psi)_{(t\chi,ta)}((\zeta, b), (\chi, a)) \|_{L^2_{1,2,0}} \leq f \left( \| \chi \|_{L^2_{2,0}} + \| a \|_{L^2_{2,4}} \right) \left( \| \zeta \|_{L^2_{2,0}} + \| b \|_{L^2_{2,4}} \right).
\]

Proof. From (8.7) we now have the \( L^2_{2,4} \) estimate

\[
\| (D^2 \Psi)_{(t\chi,ta)}((\zeta, b), (\chi, a)) \|_{L^2_{2,4}} \leq c \| \chi \|_{C^0} (\| d_{A_0} \zeta \|_{L^2_{2,4}} + \| a \|_{L^2_{2,4}} \| \zeta \|_{C^0} + \| b \|_{L^2_{2,4}}) + c \| a \|_{L^2_{2,4}} \| \zeta \|_{C^0},
\]

and thus:

(8.21) \[
\| (D^2 \Psi)_{(t\chi,ta)}((\zeta, b), (\chi, a)) \|_{L^2_{2,4}} \leq c \left( \| \chi \|_{L^2_{2,0}} + \| a \|_{L^2_{2,4}} \| \chi \|_{L^2_{2,0}} + \| a \|_{L^2_{2,4}} \right)
\times \left( \| \zeta \|_{L^2_{2,0}} + \| b \|_{L^2_{2,4}} \right).
\]

Combining (8.14) and (8.21) yields the claim. \( \square \)

The rest of the argument is just as before. This completes the proof of the theorem. \( \square \)

We now have our second proof of Theorem 6.1 via Theorems 8.2 and 8.4:

Proof of Theorem 6.1. From the hypotheses we have \( A_0 \in \mathcal{A}_E^k \) and \( [A] \in \mathcal{B}_E^k \) with \( k \geq 2 \). According to Lemma 6.3, there is gauge transformation \( w \in \mathcal{G}_E^3 \) such that

\[
\text{dist}_{L^1_{1,2,0}}([A], [A_0]) = \| w(A) - A_0 \|_{L^1_{1,2,0}},
\]

where \( A \in \mathcal{A}_E^k \), so Theorems 8.2 and the argument of 8.4 imply that there is a gauge transformation \( v \in \mathcal{G}_E^2 \) so that \( u(A) \) satisfies the conclusions of Assertion (2) with \( u = vw \in \mathcal{G}_E^3 \). Since \( d_{A_0}^*(u(A) - A_0) = 0 \) and \( u \in \mathcal{G}_E^3 \) and \( A, A_0 \in \mathcal{A}_E^k \), a standard bootstrapping argument implies that \( u \in \mathcal{G}_E^{k+1} \).

Similarly, by Lemma 6.3, there is gauge transformation \( w \in \mathcal{G}_E^3 \) such that

\[
\text{dist}_{L^1_{1,2,0}}([A], [A_0]) = \| w(A) - A_0 \|_{L^1_{1,2,0}},
\]

so Assertion (1) follows from Theorem 8.4 in the same manner. \( \square \)
References


Received July 29, 1999 and revised August 10, 2000. The author was supported in part by an NSF Mathematical Sciences Postdoctoral Fellowship under grant DMS 9306061 and by NSF grant DMS 9704174.

Rutgers University
Piscataway, NJ 08854-8019
E-mail address: feehan@math.rutgers.edu

University of Dublin
Trinity College
Dublin 2
Ireland
TECHNIQUES FOR APPROACHING THE DUAL RAMSEY PROPERTY IN THE PROJECTIVE HIERARCHY

LORENZ HALBEISEN AND BENEDIKT LÖWE

We define the dualizations of objects and concepts which are essential for investigating the Ramsey property in the first levels of the projective hierarchy, prove a forcing equivalence theorem for dual Mathias forcing and dual Laver forcing, and show that the Harrington-Kechris techniques for proving the Ramsey property from determinacy work in the dualized case as well.

1. Introduction.

Set theory of the reals is a subfield of Mathematical Logic mainly concerned with the interplay between forcing and Descriptive Set Theory. One of the motivations behind Descriptive Set Theory is the strong intuition that simple sets of real numbers should not display irregular behaviour, or, in other words, they should be topologically and measure theoretically nice.

In order to fill this statement with mathematical content, we should make clear what we mean by “simple” and what we mean by “nice”. Both questions have a conventional and well known answer:

- The measure of simplicity with which we categorize our sets of reals is the projective hierarchy, in other words, the number of quantifiers necessary to define the sets with a formula in first order analysis (or second order arithmetic).
- A set should be considered “nice” or “regular” if it has the Baire property in all naturally occurring topologies on the real numbers and is a member of all conceivably natural $\sigma$-algebras.

Set theory teaches us that the axioms of ZFC do not entail a formal version of these intuitions: It is consistent with ZFC that there are irregular sets already at the first level of the projective hierarchy.\(^1\) Thus the focus shifts from proving that all simple sets are nice to investigating the situations under which our intuitions are met by the facts.

\(^1\)In Gödel’s Constructible Universe $L$, there is a $\Delta^1_2$ set which is not Lebesgue measurable and which does not have the Baire property. Worse still, there is an uncountable $\Pi^1_1$ set with no perfect subset and a $\Pi^1_1$ set which is not Martin measurable.
A whole array of research in this direction is dealing with the second level of the projective hierarchy. Solovay provided us with the prototype of a characterization theorem for the second level:

**Theorem 1.1.** The following are equivalent:

(i) Every $\Sigma^1_2$ set of reals has the Baire property.

(ii) For every real $a \in [\omega]^\omega$ the set of Cohen generic reals over the model $L[a]$ is comeager in the standard topology on the real numbers.

One could call a theorem like this a “transcendence principle over the constructible universe”. These principles connect the theory of forcing and the topological properties of the reals. Comparable theorems have been proved in [JuSh89] (for the $\Delta^1_2$ level) and in [BrLö99] (for different topologies and $\sigma$-algebras).

A particularly interesting instance of niceness in the above sense is the **Ramsey property**, a topological property which is deeply connected to Ramsey theory and infinitary combinatorics. The Ramsey property is linked to a forcing notion called **Mathias forcing**, introduced by Mathias in [Mat77], and Judah and Shelah were able to obtain the following Solovay–type characterization for it (cf. [JuSh89, Theorem 2.7 & Theorem 2.8]):

**Theorem 1.2.** The following are equivalent:

(i) Every $\Sigma^1_2$ set of reals has the Ramsey property.

(ii) Every $\Delta^1_2$ set of reals has the Ramsey property.

(iii) For every real $a \in [\omega]^\omega$ the set \( \{ r \in [\omega]^\omega : r \text{ is Ramsey over } L[a][\mathcal{F}] \} \)

is comeager in the Ellentuck topology.\(^2\)

One connection to Mathias forcing is given by the following result (cf. [HalbJu96, Theorem 4.1]):

**Proposition 1.3.** If $N$ is any model of ZFC, then the following are equivalent:

(i) $N$ is a model in which every $\Sigma^1_2$ set is Ramsey, and

(ii) $N$ is $\Sigma^1_3$-Mathias-absolute.\(^3\)

As the Ramsey property talks about infinite subsets of the natural numbers, it is easily dualized by something we shall call the **dual Ramsey property**, talking about infinite partitions of the natural numbers.\(^4\) This

---

\(^2\)A real $r$ is Ramsey over $L[a][\mathcal{F}]$ if and only if $\mathcal{F} := D_r \cap L[a][D_r]$ forms an ultrafilter in $L[a][\mathcal{F}]$, where $D_r := \{ r^* \in [\omega]^\omega : r^* \subseteq r \}$, and $r$ is $L[\mathcal{F}]$-generic over $L[a][\mathcal{F}]$, where $L[\mathcal{F}]$ is Laver forcing restricted to $\mathcal{F}$.

\(^3\)Similar characterizations also exist for some other properties, e.g., for Lebesgue measurability and Baire property (cf. [BaJu95, Theorem 9.3.8]).

\(^4\)Infinite subsets can be seen as images of injective functions and infinite partitions can be seen as preimages of surjective functions, so the move from infinite subsets to infinite partitions actually is a dualization process.
property has been introduced by Carlson and Simpson in [CaSi84] and further investigated in [Halb98-2] and [Halb98-1].

One thing that is striking about the relationship between the Ramsey property and the dual Ramsey property are the distinctive symmetries and asymmetries. This paper can be understood as a catalogue of some of the similarities; in fact, one could see parts of this paper as an attempt to reach the obvious dualization of Theorem 1.2:

**Conjecture 1.4.** The following are equivalent:

(i) Every $\Sigma^1_2$ set of reals has the dual Ramsey property.

(ii) Every $\Delta^1_2$ set of reals has the dual Ramsey property.

(iii) For every real $a \in [\omega]^\omega$ the set \{ $R : R$ is dual Ramsey over \(L[a][D^R]\)$ \} is comeager in the dual Ellentuck topology.

In order to approach this conjecture and to give an idea what “dual Ramsey over \(L[a][D^R]\)” could mean, several of the techniques of [JuSh89] and [Mat77] have to be adapted to the new environment:

Mathias forcing has a characteristic product form $M = \mathcal{P}(\omega)/\text{fin} \ast M_U$ where $U$ is the canonical name for the generic ultrafilter added by $\mathcal{P}(\omega)/\text{fin}$. This ultrafilter is in fact a Ramsey ultrafilter, and Judah and Shelah show in their [JuSh89] that Mathias forcing relative to an ultrafilter is forcing-equivalent to Laver forcing relative to the same ultrafilter, provided that the ultrafilter is Ramsey (cf. [JuSh89, Theorem 1.20 (i)]):

**Theorem 1.5.** Let $\mathcal{F}$ be a Ramsey ultrafilter. Then the forcing notions $L_{\mathcal{F}}$ and $\mathbb{M}_{\mathcal{F}}$ are equivalent.

This theorem was our motivation to search for a dual version of Laver forcing and the dualization of Ramsey ultrafilters to work towards a dualization of Theorem 1.2.

In our dualized situation there are many things to be done to make sense of the dualized versions: One has to find a dualized version of $\mathcal{P}(\omega)/\text{fin}$ and to prove the corresponding product form of dual Mathias forcing (already done in [Halb98-1]), one has to find a dualized version of Ramsey ultrafilters, and one has to make explicit what Laver forcing in this context is supposed to mean.

Section 2 of this paper defines all the dualized notions needed for the technical work on the dual Ramsey property. In Section 3, the reader will find a couple of facts about a dualization of Ramsey ultrafilters; their connection to the game filters from [Halb98-1] is given in the appendix. Section 4

---

5 A set $\mathcal{F} \subseteq [\omega]^\omega$ is a Ramsey filter if $\mathcal{F}$ is a filter and for any colouring $\tau : [\omega]^n \rightarrow r + 1$ (with $n, r \in \omega$) there is an $x \in \mathcal{F}$ such that $\tau|[x]^n$ is constant. Notice that every Ramsey filter is an ultrafilter.
moves on to discuss dual Laver forcing and proves the dualized version of Theorem 1.5.

In Sections 5 and 6 we investigate the extent of sets with the dual Ramsey property in the projective hierarchy. In Section 5 we prove a couple of consistency results for the first three levels of the projective hierarchy. After that, Section 6 looks at the dual Ramsey property from a completely different angle: If we assume an appropriate amount of determinacy, we know that a large collection of sets has the Ramsey property. This result is not at all immediate from the Banach-Mazur game for the topology associated with the Ramsey property. However, in that section we note that the Harrington-Kechris technique of proving the Ramsey property from standard determinacy (cf. [HarKe81]) alone works for the dualized case as well.

It should be mentioned that the technicalities of the dualization process are not always as easy as they seem in retrospect. Finding the correct and natural dualizations for the interesting notions from the classical case is the most challenging part in this project. After the right dualizations are at hand, in most cases one can follow the classical proofs. So, the merits of this paper lie mainly in the definitions that make the proofs nice and easy and give a proper and firmly rooted understanding of the symmetries. This is also the reason for the unproportional size of Section 2 compared to the other sections.

2. Definitions and notations.

2.0. Set-theoretic notation. Most of our set-theoretic notation is standard and can be found in textbooks like [Je78], [Ku83] or [BaJu95]. For the definitions and some basic facts concerning the projective hierarchy we refer the reader to [Kan94, §12].

We shall consider the set $\omega^\omega$ as the set of real numbers. For the Turing join of two reals $x$ and $y$ (i.e., coding two reals into one), we use the standard notation $x \oplus y$.

2.1. Partitions. A set $P \subseteq \mathcal{P}(S)$ is a partition of the set $S$ if $\emptyset \notin A$, $\bigcup P = S$ and for all distinct $p_1, p_2 \in P$ we have $p_1 \cap p_2 = \emptyset$. An element of a partition $P$ is also called a block of $P$ and $\text{dom}(P) := \bigcup P$ is called the domain of $P$. A partition $P$ is called infinite, if $|P|$ is infinite, where $|P|$ denotes the cardinality of the set $P$. The equivalence relation on $S$ uniquely determined by a partition $P$ is denoted by $\sim_P$.

The obstacle is that playing basic open sets in this topology cannot be coded by natural numbers. So the Banach-Mazur games essentially needs determinacy for games with real moves, e.g., $\text{PD}_R$. This is connected to the famous open question whether $\text{AD}$ implies that every set has the Ramsey property (cf. [Kan94, Question 27.18]).
Let $P$ and $Q$ be two arbitrary partitions. We say that $P$ is \textbf{coarser} than $Q$ (or that $Q$ is \textbf{finer} than $P$) and write $P \subseteq Q$, if for all blocks $p \in P$, the set $p \cap \text{dom}(Q)$ is the union of some sets $q_i \cap \text{dom}(P)$, where each $q_i$ is a block of $Q$. Let $P \cap Q$ be the finest partition which is coarser than $P$ and $Q$ with $\text{dom}(P \cap Q) = \text{dom}(P) \cup \text{dom}(Q)$. We say that $P$ is \textbf{almost coarser} than $Q$ and write $P \sqsubseteq^* Q$ if there is a partition $R$ such that $\text{dom}(R)$ is finite and $R \cap P \subseteq Q$. If $P \sqsubseteq^* Q$ and $Q \sqsubseteq^* P$, then we write $P \equiv^* Q$.\footnote{We choose this notation because the properties of $\subseteq$ and $\sqsubseteq$ are similar to those of $\subseteq$ and $\cap$.}

Let $P$ and $Q$ be two partitions. If for each $p \in P$ there is a $q \in Q$ such that $p = q \cap \text{dom}(P)$, we write $P \preceq Q$. Note that $P \preceq Q$ implies $\text{dom}(P) \subseteq \text{dom}(Q)$.

For $x \subseteq \omega$ let $\min(x) := \bigcap x$. If $P$ is a partition with $\text{dom}(P) \subseteq \omega$, then $\text{Min}(P) := \{ \min(p) : p \in P \}$; and for $n \in \omega$, $P(n)$ denotes the unique block $p \in P$ such that $|\min(p) \cap \text{Min}(P)| = n + 1$.

The set of all infinite partitions of $\omega$ is denoted by $(\omega)^{\omega}$; and the set of all partitions $s$ with $\text{dom}(s) \in \omega$ is denoted by $(\mathbb{N})$.

For $s \in (\mathbb{N})$, let $s^*$ denote the partition $s \cup \{\{\text{dom}(s)\}\}$. Notice that $|s^*| = |s| + 1$.

For a natural number $n$, let $(\omega)^{n*}$ denote the set of all $u \in (\mathbb{N})$ such that $|u| = n$. Further, for $n \in \omega$ and $X \in (\omega)^{\omega}$ let

$$(X)^{n*} := \{ u \in (\mathbb{N}) : |u| = n \wedge u^* \subseteq X \},$$

and for $s \in (\mathbb{N})$ such that $|s| \leq n$ and $s \subseteq X$, let

$$(s, X)^{n*} := \{ u \in (\mathbb{N}) : |u| = n \wedge s \preceq u \wedge u^* \subseteq X \}.$$ 

It will be convenient to consider $\omega$ as the partition which contains only singletons, and therefore, for $s \in (\mathbb{N})$, $(s, \omega)^{n*} := \{ u \in (\mathbb{N}) : |u| = n \wedge s \preceq u \}$.

A family $\mathcal{F} \subseteq (\omega)^{\omega}$ is called a \textbf{filter} if

\begin{itemize}
  \item [(\alpha)] $\emptyset \notin \mathcal{F}$;
  \item [(\beta)] If $X \in \mathcal{F}$ and $X \subseteq Y$, then $Y \in \mathcal{F}$;
  \item [(\gamma)] If $X$ and $Y$ belong to $\mathcal{F}$, then $X \cap Y \in \mathcal{F}$.
\end{itemize}

Further, we call $\mathcal{F} \subseteq (\omega)^{\omega}$ an \textbf{ultrafilter} if $\mathcal{F}$ is a filter which is not properly contained in any filter. Notice that if $X \in \mathcal{F}$ and $\mathcal{F} \subseteq (\omega)^{\omega}$ is an ultrafilter, then each $Y \in (\omega)^{\omega}$ with $Y \not\subseteq X$ belongs to $\mathcal{F}$, too.

\subsection*{2.2. The dual Ellentuck topology and the dual Ramsey property.}

Let $X \in (\omega)^{\omega}$ and $s \in (\mathbb{N})$ be such that $s \subseteq X$. Then

$$(s, X)^{\omega} := \{ Y \in (\omega)^{\omega} : s \preceq Y \subseteq X \}$$

and

$$(X)^{\omega} := (\emptyset, X)^{\omega} = \{ Y \in (\omega)^{\omega} : Y \subseteq X \}.$$
Obviously, this definition depends on the model we are working in, so, if this should become important, we denote by \((s,X)\) the corresponding set interpreted in the model \(N\).

Let the basic open sets on \((\omega)^\omega\) be \(\emptyset\) and the sets \((s,X)\), where \(s\) and \(X\) are as above. These sets are called the \textbf{dual Ellentuck neighbourhoods}. The topology induced by the dual Ellentuck neighbourhoods is called the \textbf{dual Ellentuck topology} (cf. [CaSi84]).

A family \(A \subseteq (\omega)^\omega\) has the \textbf{dual Ramsey property} (or just \textbf{is dual Ramsey}) if and only if there is a partition \(X \in (\omega)^\omega\) such that either \((X)^\omega \subseteq A\) or \((X)^\omega \cap A = \emptyset\).

Closely related (but stronger) is the notion of a completely dual Ramsey set: A set \(A \subseteq (\omega)^\omega\) is said to be \textbf{completely dual Ramsey} if and only if for each dual Ellentuck neighbourhood \((s,X)\) there is a \(Y \in (s,X)^\omega\) such that \((s,Y)^\omega \subseteq A\) or \((s,Y)^\omega \cap A = \emptyset\). If we are always in the latter case, then \(A\) is called \textbf{completely dual Ramsey-null}. It is not clear if “every projective set is completely dual Ramsey” is really stronger than just “every projective set is dual Ramsey”, because we cannot simply translate Lemma 2.1 of [BrLo99], where it is shown among other things that “every projective set is Ramsey” and “every projective set is completely Ramsey” are equivalent.

Carlson and Simpson proved in [CaSi84] that a set \(A\) is completely dual Ramsey if and only if \(A\) has the Baire property with respect to the dual Ellentuck topology and \(A\) is completely dual Ramsey-null if and only if \(A\) is meager with respect to the dual Ellentuck topology.\(^8\) As a matter of fact we like to mention that in the dual Ellentuck topology every meager set is nowhere dense and hence, the dual Ellentuck topology is a Baire topology (i.e., no open set is meager). This corresponds to the similar facts about “being completely Ramsey” and the Ellentuck topology.\(^9\)

\subsection*{2.3. Dual Mathias forcing.} The conditions of the \textbf{dual Mathias forcing} \(\mathbb{M}^* = (M^*, \leq)\) are the pairs \((s,X)\) such that \((s,X)^\omega\) is a non-empty dual Ellentuck neighbourhood, and the partial order is defined by

\[\langle s, X \rangle \leq \langle t, Y \rangle \iff (s,X)^\omega \subseteq (t,Y)^\omega.\]

If \((s,X)\) is an \(\mathbb{M}^*\)-condition, then we call \(s\) the \textbf{stem} of the condition.

If \(G\) is \(\mathbb{M}^*\)-generic over \(N\), then \(G\) induces in a canonical way an infinite partition \(X_G \in (\omega)^\omega\) such that \(N[G] = N[X_G]\), and therefore we consider

\(^8\)A set \(S\) has the Baire property if there is an Borel set \(B\) such that the symmetric difference \(S \triangle B\) is meager, where a meager set is the union of countably many nowhere dense sets.

\(^9\)Cf. [El74] & [Ke95, §19.D].
the partition $X_G$ as the generic object. We can reconstruct the original $G$
from $X_G$ by observing that

$$\langle s, X \rangle \in G \iff X_G \in (s, X)_{N(G)}^\omega.$$ 

Since the dual Ellentuck topology is innately connected with dual Mathias
forcing, we choose the following notation for meager and comeager sets in
the dual Ellentuck topology:

$$A \in (m_0^*) \iff A \text{ is dual Ellentuck meager, and}$$

$$A \in (m_1^*) \iff A \text{ is dual Ellentuck comeager,}$$

i.e., $(\omega)^\omega \setminus A$ is dual Ellentuck meager.

Since $A$ is dual Ellentuck meager if and only if $A$ is completely dual
Ramsey-null, $(m_0^*) \subseteq \mathcal{P}((\omega)^\omega)$ is also the ideal of completely dual Ramsey-null sets.

The following fact gives two properties of dual Mathias forcing which also
hold for Mathias forcing.

**Fact 2.1.** If $X_G$ is $M^*$-generic and $Y \in (X_G)^\omega$, then $Y$ is $M^*$-generic as
well (we will call this property the **homogeneity property**); and therefore,
dual Mathias forcing is proper. Moreover, for any sentence $\Phi$ of the forcing
language $M^*$ and for any $M^*$-condition $\langle s, X \rangle$, there is an $M^*$-condition
$\langle s, Y \rangle \leq \langle s, X \rangle$ such that $\langle s, Y \rangle \Vdash_{M^*} \Phi$ or $\langle s, Y \rangle \Vdash_{M^*} \neg \Phi$ (this property is
called **pure decision**).

**Proof.** For a proof, cf. [CaSi84, Theorem 5.5 & Theorem 5.2].

As an immediate consequence we get that the set of dual Mathias generic
partitions over every model $N$ is either empty or a non-meager set which is
completely dual Ramsey.

Like Mathias forcing, dual Mathias forcing has also a characteristic product
form.

Let $U^* = ((\omega)^\omega, \leq)$ be the partial order defined as follows:

$$X \leq Y \iff X \sqsupseteq^* Y.$$ 

$U^*$ is the natural dualization of $\mathcal{P}(\omega)/\text{fin}$.

For a family $\mathcal{E} \subseteq (\omega)^\omega$ we define the **restricted dual Mathias forcing**
$M^*_\mathcal{E}$ as follows. The conditions of $M^*_\mathcal{E} = (M^*_\mathcal{E}, \leq)$ are the $M^*$-conditions
$\langle s, X \rangle$ such that $X \in \mathcal{E}$.

Now we get

**Fact 2.2.** $M^* = U^* * M^*_G$, where $G$ is the canonical name for the $U^*$-generic
object.

**Proof.** For a proof, cf. [Halb98-1, Fact 2.5].
2.4. Restricted dual Laver forcing. In order to define the forcing notion which will be investigated later on, we first have to give some notations.

For $T \subseteq (\mathbb{N})$ and $t \in T$ we define the successor set of $t$ in $T$ as follows:

$$\text{succ}_T(t) := \{ u \in T : t \not\leq u \land |u| = |t| + 1 \}.$$ 

Let $\mathcal{E} \subseteq (\omega)^\omega$ be any non-empty family (later on we investigate only the case when $\mathcal{E}$ is an ultrafilter).

With respect to $\mathcal{E}$, we define the dual Laver forcing restricted to $\mathcal{E}$, denoted by $L^*_\mathcal{E} = \langle L^*_\mathcal{E}, \leq \rangle$, as follows:

- $(\alpha)$ $p \in L^*_\mathcal{E}$ if and only if $p \subseteq (\mathbb{N})$ with the property that there is an $s \in p$ (denoted $\text{stem}(p)$) such that for all $t \in p$ we have $s \not\leq t$.
- $(\beta)$ There exists a set $\{ X^p_t : t \in p \} \subseteq \mathcal{E}$ such that for $t \in p$ we have $t^* \subseteq X^p_t$ and

$$\text{succ}_p(t) = \{ u : u \in (t^*, X^p_t)(|t|+1)^* \}. $$

Further, for $t, u \in p$ with $t \not\leq u$ we have

$$(u, X^p_u)^\omega \subseteq (t, X^p_t)^\omega,$$
and if $\text{dom}(t) = \text{dom}(u)$ and $t \subseteq u$, then

$$X^p_t = X^p_u.$$

- $(\gamma)$ For two $L^*_\mathcal{E}$-conditions $p$ and $q$ we stipulate

$$p \leq q \iff p \subseteq q.$$ 

Notice that $p \leq q$ implies $\text{stem}(q) \not\leq \text{stem}(p)$ and hence, if $G \subseteq L^*_\mathcal{E}$ is $L^*_\mathcal{E}$-generic over some $\mathbb{N}$, then the set $\{ s : s = \text{stem}(p) \text{ for some } p \in G \}$ forms in a canonical way a partition $X_G \in (\omega)^\omega$. Moreover, $\mathbb{N}[G] = \mathbb{N}[X_G]$ and therefore we may consider also the partition $X_G$ as the $L^*_\mathcal{E}$-generic object.

For an $L^*_\mathcal{E}$-condition $p$ we call a partition $X \in (\omega)^\omega$ a branch of $p$ if each $t \in (\mathbb{N})$ with $t^* \subseteq X$ belongs to $p$.

**Fact 2.3.** If $X$ is a branch of the $L^*_\mathcal{E}$-condition $p$ where $\text{stem}(p) = s$ and $Y \in (s, X)^\omega$, then $Y$ is a branch of $p$, too.

**Proof.** This follows immediately from $(\beta)$. □

2.5. Special ultrafilters on $(\omega)^\omega$. A family $\mathcal{F}$ has the segment colouring property (or just scp) if for any $s \subseteq X \in \mathcal{F}$ with $|s| = n$ and for any colouring $\pi : (s, X)^{(n+k)^*} \to r$, where $r$ and $n + k$ are positive natural numbers, there is a $Y \in (s, X)^\omega \cap \mathcal{F}$ such that $(s, Y)^{(n+k)^*}$ is monochromatic.

A family $\mathcal{F} \subseteq (\omega)^\omega$ is an scp-filter if $\mathcal{F}$ is a filter which has the segment colouring property.

In Section 7 we shall introduce the notion of game filters (from [Halb98-1]) and show that game filters are scp-filters.

**Fact 2.4.** If $\mathcal{F} \subseteq (\omega)^\omega$ is an scp-filter, then $\mathcal{F}$ is an ultrafilter.
Proof. Let \( F \subseteq (\omega)^\omega \) be an scp-filter and assume that there exists an \( X \in (\omega)^\omega \) such that for every \( Y \in F \), \( X \cap Y \in (\omega)^\omega \). Let \( \pi : (\omega)^{n^*} \to 2 \) be such that \( \pi(s) = 0 \) if \( s \in (X)^{n^*} \), otherwise \( \pi(s) = 1 \). Because \( F \) has the segment colouring property, we find a \( Y \in F \) such that \( \pi|Y|^{n^*} \) is constant. If \( \pi|Y|^{n^*} = \{1\} \), then \( X \cap Y \notin (\omega)^\omega \) which contradicts the assumption. Thus, \( \pi|Y|^{n^*} = \{0\} \), which implies \( X \in F \) and hence, the filter \( F \) is maximal. \( \square \)

A family \( \mathcal{F} \) is diagonalizable if for any \( L^*_F \)-condition \( p \), there is a partition \( X \in \mathcal{F} \) such that \( X \) is a branch of \( p \). Notice that a diagonalizable family can also be characterized by a two player game, where the \( L^*_F \)-condition \( p \) can be considered as a strategy for player I.

A family \( \mathcal{F} \) is a Ramsey* filter if \( \mathcal{F} \) is a diagonalizable scp-filter.

In Footnote 5 we have defined Ramsey ultrafilters over \( \omega \) in terms of colourings. This definition corresponds to the definition of scp-filters. On the other hand, Galvin and Shelah proved that Ramsey ultrafilters can be characterized as well by a two player game without a winning strategy for player I, where a winning strategy for player I is in fact a restricted Laver-condition (cf. [BaJu95, Theorem 4.5.3]). This definition of Ramsey ultrafilters corresponds to diagonalizable filters. It is possible that the notions of “scp-filters” and “diagonalizable filters” are equivalent, but this is still open.

Beyond the dualization of the notion of a Ramsey ultrafilter, the dualization process leading from \( [\omega]^\omega \) to \( (\omega)^\omega \) has interesting consequences for the spaces of ultrafilters on these spaces. These consequences belong to the asymmetrical aspects of the relationship between \( [\omega]^\omega \) and \( (\omega)^\omega \) and are the point of focus in [HalbLö∞].

2.6. Switching between reals and partitions. We fix \( b : [\omega]^2 \to \omega \) to be any arithmetic bijection between the set of pairs of natural numbers and \( \omega \).

Let \( x \in [\omega]^\omega \); then the set \( \text{trans}(x) \subseteq \omega \) is defined by

\[
n \in \text{trans}(x) : \iff \exists s \in \omega^\omega \left( n = b(s(0), s(|s| - 1)) \text{ and } \forall k \in |s| - 1 (b(s(k), s(k + 1)) \in x) \right).
\]

As the name suggests, \( \text{trans}(x) \) is the set of codes of pairs in the transitive closure of the relation \( b(k, \ell) \in x \). A real \( x \) is called transitive if \( \text{trans}(x) = x \).

Note that in general \( \text{trans}(x) \subseteq x \) and that the relation \( R_x(k, \ell) : \iff b(k, \ell) \in \text{trans}(x) \)
is symmetric (by choice of the domain of \(\mathcal{b}\)) and transitive. Thus, if \(x \in [\omega]^\omega\),
we can consider \(x\) as a partition (by reflexivization of \(R_x\)) via
\[ n \sim_x m : \iff n = m \text{ or } \mathcal{b}(n, m) \in \text{trans}(x). \]

We call this partition the \textbf{corresponding partition} of \(x \in [\omega]^\omega\), and
denote it by \(\text{cp}(x)\). Note that \(\text{cp}(x) \in (\omega)^\omega\) if
\[ \forall k \exists n > k \forall m < n(\neg(n \sim_x m)) \]
and further if \(y \subseteq x\), then \(\text{cp}(y) \sqsubseteq \text{cp}(x)\).

We encode a partition \(X\) of \(\omega\) by a real \(\text{pc}(X)\) (the \textbf{partition code} of \(X\)) as follows.
\[ \text{pc}(X) := \{ k \in \omega : \exists n \exists m(k = \mathcal{b}(n, m) \land (n \sim_X m)) \}. \]

Note that if \(X \sqsubseteq Y\) then \(\text{pc}(X) \supseteq \text{pc}(Y)\).

Notice that both the function \(\text{pc}\) and the function \(\text{cp}\) are arithmetic, and
that they are in a sense inverse to each other:

\textbf{Observation 2.5.} For every \(X \in (\omega)^\omega\) and every \(x \in [\omega]^\omega\) the following hold:
(i) \(\text{cp}(\text{pc}(X)) = X\), and
(ii) if \(x\) is transitive, then \(\text{pc}(\text{cp}(x)) = x\).

Now, a set \(A \subseteq [\omega]^\omega\) has the \textbf{dual Ramsey property} (or just \textbf{is dual Ramsey}) if and only if the set \(\{X \in (\omega)^\omega : \exists x \in A(X = \text{cp}(x))\}\) has the
dual Ramsey property. By Observation 2.5, this is equivalent to saying that
the set \(\{X \in (\omega)^\omega : \text{pc}(X) \in A\}\) has the dual Ramsey property.

By the definition of the dual Ramsey property we have that every \(\Sigma^1_n\) set
is dual Ramsey if and only if every \(\Pi^1_n\) set is dual Ramsey. Furthermore, we
have by [Halb98-1, Lemma 7.2] that if every \(\Sigma^1_n\) set is dual Ramsey then
every \(\Sigma^1_n\) set has the classical Ramsey property.

As a matter of fact we like to mention the following

\textbf{Proposition 2.6.} If every \(\Delta^1_n\) set has the dual Ramsey property, then every
\(\Delta^1_n\) set has the Ramsey property.

\textit{Proof.} Suppose \(A\) is a \(\Delta^1_n\) set of reals. Let \(\varphi\) be a \(\Sigma^1_n\) formula and \(\psi\) be a
\(\Pi^1_n\) formula witnessing this, i.e.,
\[ x \in A \iff \varphi(x) \iff \psi(x). \]

To show that \(A\) is Ramsey we define a different \(\Delta^1_n\) set by formulae \(\varphi^*\) and
\(\psi^*\) as follows:
\[ \varphi^*(v) : \iff \exists w(w = \text{Min}(\text{cp}(v)) \land \varphi(w)), \]
\[ \psi^*(v) : \iff \forall w(w = \text{Min}(\text{cp}(v)) \rightarrow \psi(w)). \]
Obviously, $\varphi^*$ is $\Sigma^1_n$ and $\psi^*$ is $\Pi^1_n$, and since $\text{Min}(cp(v))$ is uniquely determined for each $v$, these two formulae are equivalent and hence define a $\Delta^1_n$ set $A^*$ of reals. The rest of the proof is exactly as in [Halb98-1, Lemma 7.2].

And as a corollary we get

**Corollary 2.7.** If every $\Delta^1_2$ set is dual Ramsey, then every $\Sigma^1_2$ set is Ramsey.

**Proof.** This follows immediately from Proposition 2.6 by Theorem 1.2. □

### 2.7. Descriptive Set Theory of the Cabal

For our results in Section 6 we shall need some basic notions of the Descriptive Set Theory of the Cabal Seminar. Everything we lay out here can be found in [Mo80], our account is just for the convenience of the more combinatorially oriented reader who might be unfamiliar with the language of the Cabal.

We shall presuppose basic knowledge with the standard notation for determinacy and the elementary results of the theory of perfect information games as outlined in [Kan94, §27].

Let $X$ be a set of reals and $\alpha \in \text{Ord}$. Any surjective function $\varphi : X \to \alpha$ is called a **norm on** $X$. The ordinal $\alpha$ is called the **length** of $\varphi$. A family $\Phi := \langle \varphi_n : n \in \omega \rangle$ of norms on $X$ is called a **scale on** $X$ if for every sequence $\langle x_i : i \in \omega \rangle \subseteq X$ and every $n \in \omega$ the following holds: If $\langle \varphi_n(x_i) : i \in \omega \rangle$ is eventually constant, say, equal to $\lambda_n$, then $x := \lim_{i \in \omega} x_i \in X$ and $\varphi_n(x) \leq \lambda_n$.  

Let $\Gamma$ be any pointclass, $\varphi$ any norm on $X$, and $\Phi$ any scale on $X$. We shall call $\varphi$ a $\Gamma$ **norm** if there are two relations $R$ and $R^*$ in $\Gamma$ such that:

$$y \in X \Rightarrow \forall x \left( (x \in X \land \varphi(x) \leq \varphi(y)) \iff R(x,y) \iff \neg R^*(x,y) \right).$$

We call a scale $\Phi$ a $\Gamma$ **scale** if all norms $\varphi_n$ occurring in $\Phi$ are $\Gamma$ norms, uniformly in $n$. We shall say that a set $X$ **admits a $\Gamma$ norm** (a $\Gamma$ scale) if there is a norm (a scale) on $X$ that is a $\Gamma$ norm (a $\Gamma$ scale).

The fundamental theorems connecting determinacy, norms and scales are the “Periodicity Theorems” of [AdMo68], [Mar68] and [Mo71]. In the following we shall need the first two Periodicity Theorems in special cases:

**First Periodicity Theorem 2.8.** Suppose that $\text{Det}(\Delta^1_{2n})$ holds and $x \in [\omega]^\omega$ is a real. Then every $\Pi^1_{2n+1}(x)$ set admits a $\Pi^1_{2n+1}(x)$ norm.

**Second Periodicity Theorem 2.9.** Suppose that $\text{Det}(\Delta^1_{2n})$ holds and $x \in [\omega]^\omega$ is a real. Then every $\Pi^1_{2n+1}(x)$ set admits a $\Pi^1_{2n+1}(x)$ scale.

---

10For the basic theory of scales, cf. [KeMo78].

11A more precise definition can be found in [Mo80, p. 228].
For proofs of these theorems (in a much more general formulation), we refer the reader to \([\text{Mo80}, 6B.1 & 6C.3]\).

We define (for every \(n \in \omega\)) the projective ordinals by
\[
\delta^1_n := \sup\{\| \leq \| : \leq \text{ is a } \Delta^1_n \text{ prewellordering on } [\omega]^\omega\},
\]
and note that for every \(\Pi^1_{2n+1}\) complete set the length of every \(\Pi^1_{2n+1}\) norm on it is exactly \(\delta^1_{2n+1}\) ([\text{Mo80}, 4C.14]).

Let \(P^x_{2n+1}\) be a \(\Pi^1_{2n+1}(x)\) complete set of reals. Assuming \(\text{Det}(\Delta^1_{2n})\) we get a \(\Pi^1_{2n+1}(x)\) scale \(\Phi^x = \{\varphi^x_m : m \in \omega\}\) for \(P^x_{2n+1}\) by Theorem 2.9.

For any real \(y \in P^x_{2n+1}\), we denote by \(\Phi^x(y)\) the sequence of ordinals determined by the scale, i.e., \(\Phi^x(y) = \langle \varphi^x_m(y) : m \in \omega \rangle\).

Now let
\[
T^x_{2n+1} := \{\langle y|m, \Phi^x(y)|m \rangle : y \in P^x_{2n+1}, m \in \omega\}
\]
be the tree associated to \(\Phi^x\). By the remark about the lengths of norms, it is a tree on \(\omega \times \delta^1_{2n+1}\). If \(x\) is any recursive real, we write \(T_{2n+1}\) instead of \(T^x_{2n+1}\).

The model \(L[T_{2n+1}]\) can be seen as an analogue of the constructible universe \(L\) in the odd projective levels: The (Shoenfield) \(\Pi^1_1\) scale for a \(\Pi^1_1\) complete set is in \(L\), hence \(L[T_1] = L\). Indeed, the reals of \(L[T_{2n+1}]\) are exactly the reals of \(M_{2n}\), the canonical iterable inner model with 2\(n\) Woodin cardinals.

Moreover, not just the reals of the models, but the models \(L[T_{2n+1}]\) themselves are independent of the choices of the particular \(\Pi^1_{2n+1}\) complete set and the scale on it, as has been shown by Becker and Kechris in [\text{BeKe84, Theorem 1 & 2}]:

**Theorem 2.10.** Assume PD and let \(x \in [\omega]^\omega\) be a real. If \(P\) and \(Q\) are \(\Pi^1_{2n+1}(x)\) complete sets, \(\Phi\) and \(\Psi\) are scales on \(P\) and \(Q\), respectively, and \(T\) and \(S\) are the trees associated to \(\Phi\) and \(\Psi\), respectively. Then \(L[T] = L[S]\).

Another consequence of determinacy which will be mentioned only briefly to simplify notation is the existence of largest countable sets of certain (lightface) complexity classes:

**Theorem 2.11.** Let \(x \in [\omega]^\omega\) be a real. Suppose that \(\text{Det}(\Delta^1_{2n}(x))\) holds. Then there is a largest countable \(\Sigma^1_{2n+2}(x)\) set which will be denoted by \(C_{2n+2}(x)\).

**Proof.** Cf. [\text{KeMo72, Theorem 2}]. \(\square\)

---

12For a proof, cf. [\text{KeMo78, 9C}].

13Combine [\text{St95, Corollary 4.9}] with [\text{HarKe81, Theorem 7.2.1}].

14As for the trees \(T_{2n+1}\), we shall omit the parameter \(x\) and write \(C_{2n+1}\) if the real \(x\) is recursive.
3. On Ramsey\* ultrafilters.

In this section we show that Ramsey\* ultrafilters exist if we assume CH and that in general both existence and non-existence of Ramsey\* ultrafilters are consistent with ZFC.

First we show that an scp-ultrafilter induces in a canonical way a Ramsey filter on $\omega$.

Fact 3.1. If $\mathcal{F} \subseteq (\omega)^{\omega}$ is an scp-ultrafilter, then $\{\text{Min}(X) : X \in \mathcal{F}\} \setminus \{0\}$ is a Ramsey filter on $\omega$.

Proof. For positive natural numbers $n$ and $r$ let $\tau : [\omega]^n \to r$ be any colouring. We define $\pi : (\omega)^{n+k} \to r$ by stipulating $\pi(s) := \tau(\text{Min}(s^*) \setminus \{0\})$. It is easy to see that if $\pi | (X)^{n+k}$ is constant for an $X \in \mathcal{F}$, then $\tau | [\text{Min}(X) \setminus \{0\}]^n$ is constant, too. \qed

Proposition 3.2. It is consistent with ZFC that there are no scp-ultrafilters.

Proof. Kunen proved (cf. [Je78, Theorem 91]) that it is consistent with ZFC that there are no Ramsey filters on $\omega$. Therefore, by Fact 3.1, in a model of ZFC in which there are no Ramsey filters, there are also no scp-ultrafilters. \qed

Let $U^* = (\omega, \leq)$ be the partial order defined as in Subsection 2.3. It is easy to see that the forcing notion $U^*$ is $\sigma$-closed (this is part of Fact 2.3 of [Halb98-1]).

Lemma 3.3. If $G$ is $U^*$-generic over $V$, then $G$ is an scp-ultrafilter in $V[G]$.

Proof. Let $s \in (\mathbb{N})$ and $k \in \omega$ with $|s| = n$ and $n+k > 0$. Further, let $\pi : (\omega)^{n+k} \to r$ be any colouring and for $s \subseteq X \in (\omega)^{\omega}$ let

$$H_{\pi(s,X)} := \{Y \in (s,X)^{\omega} : \pi(s,X)(s,Y)^{(n+k)*} \text{ is constant}\}.$$

By the main result of [Halb\$\infty$] and its proof, for every dual Ellentuck neighbourhood $(s,X)^{\omega}$ and for any colouring $\pi : (s,X)^{(n+k)*} \to r$, there is a $Y \in (s,X)^{\omega}$ such that $\pi | (s,Y)^{(n+k)*}$ is constant. Hence, for any dual Ellentuck neighbourhood $(s,X)^{\omega}$ and for any colouring $\pi : (\omega)^{(n+k)*} \to r$, the set $H_{\pi(s,X)}$ is dense below $X$. Because every such colouring $\pi$ can be encoded by a real and $U^*$ is $\sigma$-closed, the forcing notion $U^*$ does not add any colouring $\pi$, which implies, because $G$ meets each dense set, that $G$ is an scp-ultrafilter in $V[G]$. \qed

We can prove with similar arguments:

Lemma 3.4. If $G$ is $U^*$-generic over $V$, then $G$ is a diagonalizable ultrafilter in $V[G]$.
Proof. Let $\dot{p}$ be a $\mathbb{U}^\ast$-name such that

$$\models_{\mathbb{U}^\ast} \text{“$\dot{p}$ is an $\mathbb{L}_G^\ast$-condition”,}$$

where $G$ is the canonical name for the $\mathbb{U}^\ast$-generic object, and let $X$ be any $\mathbb{U}^\ast$-condition. Because $\dot{p}$ can be encoded by a real number and $\mathbb{U}^\ast$ is $\sigma$-closed, there is a $\mathbb{U}^\ast$-condition $Y \subseteq X$ and a real $p' \in V$ such that $Y \models_{\mathbb{U}^\ast} p' = \dot{p}$, which implies $Y \subseteq^* \text{succ}_p(t)$ for every $t \in p'$. By induction one can construct a $Z \subseteq^* Y$ such that $Z$ is a branch of $p'$ and therefore,$$

Z \models_{\mathbb{U}^\ast} \text{“there is a branch of $\dot{p}$ which belongs to $G^\ast$.”}
$$

Since $Z \subseteq X$, this completes the proof. \hfill \square

Proposition 3.5. Assume CH, then there is a Ramsey$^\ast$ ultrafilter.

Proof. Assume $V \models \text{CH}$. Let $\chi$ be large enough such that $\mathcal{P}(\omega^\omega) \subseteq H(\chi)$, i.e., the power set of $(\omega)^\omega$ (in $V$) is hereditarily of size $< \chi$. Let $N$ be an elementary submodel of $\langle H(\chi), \in \rangle$ containing all reals of $V$ with $|N| = 2^{\aleph_0}$. We consider the forcing notion $\mathbb{U}^\ast$ in the model $N$. Because $|N| = 2^{\aleph_0}$, in $V$ there is an enumeration $\{D_\alpha : \alpha < 2^{\aleph_0}\}$ of all dense sets of $\mathbb{U}^\ast$ which lie in $N$. Since $\mathbb{U}^\ast$ is $\sigma$-closed and because $V \models \text{CH}$, $\mathbb{U}^\ast$ is $2^{\aleph_0}$-closed in $V$ and therefore we can construct a descending sequence $\{p_\alpha : \alpha < 2^{\aleph_0}\}$ in $V$ such that $p_\alpha \in D_\alpha$ for each $\alpha < 2^{\aleph_0}$. Let $G := \{p \in (\omega)^\omega : p_\alpha \subseteq p \text{ for some } p_\alpha\}$, then $G$ is $\mathbb{U}^\ast$-generic over $N$. By Lemma 3.3 and Lemma 3.4 we have $N[G] \models \text{“there is a Ramsey$^\ast$ ultrafilter”,}$ and because $N$ contains all reals of $V$ and every function $f : (\omega)^n \to r$ (where $n, r \in \omega$) and every $\mathbb{L}_G^\ast$-condition $p$ can be encoded by a real number, the Ramsey$^\ast$ ultrafilter in $N[G]$ is also a Ramsey$^\ast$ ultrafilter in $V$, which completes the proof. \hfill \square

4. On $L_\delta^\ast$ and $M_\delta^\ast$ for Ramsey$^\ast$ filters $\mathcal{F}$.

In this section, $\mathcal{F} \subseteq (\omega)^\omega$ denotes always a Ramsey$^\ast$ ultrafilter.

We shall show that the forcing notions $L_\delta^\ast$ and $M_\delta^\ast$ are equivalent and that both forcing notions have pure decision and the homogeneity property (this means that coarsenings of generic objects remain generic, see Fact 2.1). We show first that $M_\delta^\ast$ has pure decision and the homogeneity property. To show this we will follow [Halb98-1, Section 4].

If $s \in (\mathbb{N})$ and $s \subseteq X \in \mathcal{F}$, then we call the dual Ellentuck neighbourhood $(s, X)^\omega$ an $\mathcal{F}$-dual Ellentuck neighbourhood and write $(s, X)^\omega_\mathcal{F}$ to emphasize that $X \in \mathcal{F}$. A set $O \subseteq (\omega)^\omega$ is called $\mathcal{F}$-open if $O$ can be written as the union of some $\mathcal{F}$-dual Ellentuck neighbourhoods.

For $s \in (\mathbb{N})$ remember that $s^* = s \cup \{\text{dom}(s)\}$.

Let $O \subseteq (\omega)^\omega$ be an $\mathcal{F}$-open set. Call $(s, X)^\omega_\mathcal{F}$ good (with respect to $O$), if for some $Y \in (s, X)^\omega_\mathcal{F}$, $(s, Y)^\omega_\mathcal{F} \subseteq O$; otherwise call it bad. Note that if $(s, X)^\omega_\mathcal{F}$ is bad and $Y \in (s, X)^\omega_\mathcal{F} \cap \mathcal{F}$, then $(s, Y)^\omega_\mathcal{F}$ is bad, too. We call $(s, X)^\omega_\mathcal{F}$
Lemma 4.1. Let \( \mathcal{F} \subseteq (\omega)^\omega \) be a Ramsey* ultrafilter and \( \mathcal{O} \subseteq (\omega)^\omega \) an \( \mathcal{F} \)-open set. If \( (s,X)_\mathcal{F}^\omega \) is bad (with respect to \( \mathcal{O} \)), then there is a \( Z \in (s,X)_\mathcal{F}^\omega \) such that \( (s,Z)_\mathcal{F}^\omega \) is ugly.

Proof. We begin by constructing an \( \mathbb{L}^*_\mathcal{F} \)-condition \( p \). Let \( s_0 \) be such that \( s \preceq s_0 \preceq X \) and \( |s| = |s_0| \), and put \( \text{stem}(p) := s_0 \). If there is an \( Y \in (s_0,X)^\mathcal{F} \subseteq \mathcal{F} \) such that \( (s_0,Y)_\mathcal{F} \) is bad, then \( s \preceq s_0 \preceq \emptyset \) and put \( \text{stem}(p) := s_0 \). Otherwise, there is a \( \emptyset \)-branch of \( X \) which is a branch of \( p \). Let \( s_n := s_0 \land \emptyset \) for all \( n \in \omega \), and let \( \{ t_i : i \leq h \} \) be an enumeration of all \( t \) such that \( s_0 \preceq t \preceq s_n \), \( |t| = |s| \), and \( \text{dom}(t) = \text{dom}(s_n) \). Further let \( Y^{-1} := X_{s_n} \). Now choose for each \( i \leq h \) a partition \( Y^i \) in \( \mathcal{F} \) such that \( Y^i \) is a Ramsey* ultrafilter and \( (s_i,Y^i)_\mathcal{F}^\omega \) is bad or \( (s_i,Y^i)_\mathcal{F}^\omega \) is ugly. We claim that \( (s_i,Y^i)_\mathcal{F}^\omega \) is bad or \( (s_i,Y^i)_\mathcal{F}^\omega \) is ugly.

Put \( p := \{ t \in (\mathbb{N}) : s_0 \preceq t \preceq s_n \) for some \( n \in \omega \} \). Since \( \mathcal{F} \) is diagonalizable, there is a partition \( Y \in \mathcal{F} \) which is a branch of \( p \). We may assume that \( s_0 \preceq Y \). Define \( S_Y := \{ t : s \preceq t \preceq Y \land |t| = |s| \} \); then, by the construction of \( p \), for all \( t \in S_Y \) we have either \( (t,Y)_\mathcal{F}^\omega \) is bad or \( (t,Y)_\mathcal{F}^\omega \) is bad. Now let \( B_0 := \{ t \in S_Y : (t,Y)_\mathcal{F}^\omega \) is bad \} and \( B_1 := \{ t \in S_Y : (t,Y)_\mathcal{F}^\omega \subseteq \mathcal{O} \} = S_Y \setminus B_0 \). Because \( \mathcal{F} \) is an scp-filter, there is a partition \( Z \in (s,Y)^\mathcal{F} \subseteq \mathcal{F} \) such that \( S_Z \subseteq B_0 \) or \( S_Z \subseteq B_1 \). If we are in the latter case, we have \( (s,Z)_\mathcal{F}^\omega \subseteq \mathcal{O} \), which is a contradiction to our assumption that \( (s,X)_\mathcal{F}^\omega \) is bad. So, we must have \( S_Z \subseteq B_0 \), which implies that \( (s,Z)_\mathcal{F}^\omega \) is ugly and completes the proof of the Lemma.

Lemma 4.2. If \( \mathcal{F} \) is a Ramsey* ultrafilter and \( \mathcal{O} \subseteq (\omega)^\omega \) is an \( \mathcal{F} \)-open set, then for every \( \mathcal{F} \)-dual Ellentuck neighbourhood \( (s,X)_\mathcal{F}^\omega \), there is a \( Y \in (s,X)_\mathcal{F}^\omega \cap \mathcal{F} \) such that \( (s,Y)_\mathcal{F}^\omega \subseteq \mathcal{O} \) or \( (s,Y)_\mathcal{F}^\omega \cap \mathcal{O} \cap \mathcal{F} = \emptyset \).

Proof. If \( (s,X)_\mathcal{F}^\omega \) is good, then we are done. Otherwise, we can construct an \( \mathbb{L}^* \)-condition \( p \) in a similar way as in Lemma 4.1, such that for any branch \( Y \) of \( p \) which belongs to \( \mathcal{F} \) we have the following: For each \( t \) with \( s \preceq t \preceq Y \), the set \( (t,Y)_\mathcal{F}^\omega \) is bad. We claim that \( (s,Y)_\mathcal{F}^\omega \cap \mathcal{O} \cap \mathcal{F} = \emptyset \). Take any \( Z \in (s,Y)_\mathcal{F}^\omega \cap \mathcal{O} \cap \mathcal{F} \). Because \( \mathcal{O} \) is \( \mathcal{F} \)-open we find a \( t \preceq Z \) such that \( (t,Z)_\mathcal{F}^\omega \subseteq \mathcal{O} \). Because \( s \preceq t \preceq Y \) we have by construction that \( (t,Y)_\mathcal{F}^\omega \) is bad. Hence, there is no \( Z \in (t,Y)_\mathcal{F}^\omega \) such that \( (t,Z)_\mathcal{F}^\omega \subseteq \mathcal{O} \). This completes the proof.

Now we can show that \( M_\mathcal{F}^* \) has pure decision and the homogeneity property.

Theorem 4.3. Let \( \mathcal{F} \) be a Ramsey* ultrafilter and let \( \Phi \) be a sentence of the forcing language. For any \( M_\mathcal{F}^* \)-condition \( \langle s,X \rangle \) there is a \( M_\mathcal{F}^* \)-condition \( \langle s,Y \rangle \leq \langle s,X \rangle \) such that \( \langle s,Y \rangle \not\Vdash_{M_\mathcal{F}^*} \Phi \) or \( \langle s,Y \rangle \not\Vdash_{M_\mathcal{F}^*} -\Phi \).
Proof. The proof is same as the proof of [Halb98-1, Theorem 4.3], using Lemma 4.2.

The next theorem shows in fact that if $\mathcal{F}$ is a Ramsey$^\star$ ultrafilter, then $M^\star_{\mathcal{F}}$ is proper.

**Theorem 4.4.** Let $\mathcal{F} \subseteq (\omega)^\omega$ be a Ramsey$^\star$ ultrafilter, then $M^\star_{\mathcal{F}}$ has the homogeneity property.

**Proof.** The proof is same as the proof of [Halb98-1, Theorem 4.4], using Lemma 4.2.

In order to show that $M^\star_{\mathcal{F}}$ and $L^\star_{\mathcal{F}}$ are equivalent if $\mathcal{F}$ is a Ramsey$^\star$ ultrafilter, we define first some special $L^\star_{\mathcal{F}}$-conditions.

An $L^\star_{\mathcal{F}}$-condition $p$ is called uniform if there is a partition $X \in F$ such that $(t,X)^\omega = (t,X^p_t)^\omega$ for every $t \in p$; this partition is denoted by $u(p)$. These conditions roughly correspond to the simple conditions of [JuSh89, Definition 1.10].

**Lemma 4.5.** If $\mathcal{F}$ is a Ramsey$^\star$ ultrafilter, then the set of all uniform $L^\star_{\mathcal{F}}$-conditions is dense and open in $L^\star_{\mathcal{F}}$.

**Proof.** Let $p \in L^\star_{\mathcal{F}}$ with $s = \text{stem}(p)$, then, since $\mathcal{F}$ is diagonalizable, there is an $X \in \mathcal{F}$ which is a branch of $p$. Let $q$ be the uniform condition with $u(q) = X$ and $\text{stem}(q) = s$. Note that $X$ is a branch of $q$. By Fact 2.3, each $Y \in (s,X)^\omega$ is also a branch of $p$, which implies that $q \leq p$. □

**Theorem 4.6.** If $\mathcal{F}$ is a Ramsey$^\star$ ultrafilter, then $M^\star_{\mathcal{F}}$ and $L^\star_{\mathcal{F}}$ are forcing equivalent.

**Proof.** Let $I := \{ p \in L^\star_{\mathcal{F}} : p \text{ is uniform}\}$ and define

$$
j : I \rightarrow M^\star_{\mathcal{F}}
p \mapsto \langle \text{stem}(p), u(p) \rangle,
$$

then it is easily checked that $j$ is a dense embedding and because (by Lemma 4.5) $I$ is dense open in $L^\star_{\mathcal{F}}$, this completes the proof. □

This is the promised dualization of Theorem 1.5 and possibly one step towards a proof of Conjecture 1.4.

5. The dual Ramsey property for simple pointclasses.

In the following we will show that it is consistent with ZFC that the sets in the first levels of the projective hierarchy are dual Ramsey. We begin with the analytic sets:

Because $M^\star$ has pure decision and the homogeneity property, one can show the pretty straightforward
Fact 5.1. Analytic sets are completely dual Ramsey.

Proof. Let $A$ be an arbitrary $\Sigma^1_1(a)$ set with parameter $a \in [\omega]^{\omega}$ and let $(s,Y)^{\omega}$ be any dual Ellentuck neighbourhood and $(s,Y)$ the corresponding $M^*$-condition. Take a countable model $N$ of a sufficiently large fragment of $\text{ZFC}$ which contains $Y$ and $a$. Let $X_G$ be the canonical name for the $M^*$-generic object. Because $M^*$ has pure decision we find an $M^*$-condition $\langle s,Z \rangle \leq \langle s,Y \rangle$ which decides “$X_G \in A$”. Since $N$ is countable, there is an $X \in (s,Z)^{\omega}$ which is $M^*$-generic over $N$ and because every $X' \in (s,X)^{\omega}$ is also $M^*$-generic we have

$$N[X] \models "(s,X)^\omega \subseteq A \text{ or } (s,X)^\omega \cap A = \emptyset".$$  

Because $A$ and $(s,Y)^{\omega}$ were arbitrary and $\Sigma^1_1$ sets are absolute between $V$ and $N$, we are done. \qed

Note that Fact 5.1 is verified without any reference to forcing by looking at the unfolded version of the Banach–Mazur game for the dual Ellentuck topology.\footnote{Cf. \cite{Kechris1995}, Theorem (21.8).}

Remember that according to Proposition 2.6 if every $\Delta^1_2$ set is dual Ramsey then every $\Delta^1_2$ set has the classical Ramsey property. Because it is not provable in $\text{ZFC}$ that every $\Delta^1_2$ set is Ramsey, it is also not provable in $\text{ZFC}$ that every $\Delta^1_2$ set is dual Ramsey (e.g., $L \models "\text{There is a } \Delta^1_2 \text{ set which is not dual Ramsey}"$). On the other hand we have

Fact 5.2. The following theories are equiconsistent:

(a) $\text{ZFC}$.
(b) $\text{ZFC} + \text{CH} + \text{every } \Sigma^1_2 \text{ set is dual Ramsey}$.
(c) $\text{ZFC} + 2^{\aleph_0} = \aleph_2 + \text{every } \Sigma^1_2 \text{ set is dual Ramsey}$.

Proof. We get both (b) and (c) by an iteration (of length $\omega_1$ and $\omega_2$ respectively) of dual Mathias forcing with countable support, starting from $L$ (cf. also \cite[Theorems 6.2 & 6.3]{Halbeisen1998}). \qed

Remember that $(m^*_0) = \{A \subseteq (\omega)^{\omega} : A \text{ is completely dual Ramsey-null}\}$ and let

$$\text{add}(m^*_0) := \min \left\{ |\mathcal{E}| : \mathcal{E} \subseteq (m^*_0) \land \bigcup \mathcal{E} \notin (m^*_0) \right\},$$  

and

$$\text{cov}(m^*_0) := \min \left\{ |\mathcal{E}| : \mathcal{E} \subseteq (m^*_0) \land \bigcup \mathcal{E} = (\omega)^{\omega} \right\}.$$  

In \cite{Halbeisen1998-2} it is shown that $\text{add}(m^*_0) = \text{cov}(m^*_0) = \mathfrak{h}$, where $\mathfrak{h}$ is the dual shattering cardinal. If $(m_0)$ denotes the ideal of classical completely Ramsey-null sets, then we get the analogous result, namely $\text{add}(m_0) = \text{cov}(m_0) = \mathfrak{h}$, where $\mathfrak{h}$ is the shattering cardinal (cf. \cite{Pillay1986}). Because every

\footnote{Cf. \cite{Kechris1995}, Theorem (21.8).}
$\Sigma_1^1$ set is the union of $\aleph_1$ Borel sets (cf. [Je78, Theorem 95]), it is easy to see that $\mathfrak{h} > \aleph_1$ implies that every $\Sigma_1^1$ set is even completely dual Ramsey (and the analogous result holds for the classical Ramsey property with respect to $h$). Now, an $\omega_2$-iteration with countable support of dual Mathias forcing starting from $L$ yields a model in which $\mathfrak{h} = \aleph_2$ (cf. [Halb98-2]). Thus, this provides another proof that “Every $\Sigma_1^1$ set is dual Ramsey” is consistent with ZFC. In Section 6 we shall provide a third proof as a byproduct of the analysis of scales under PD.

Concerning Martin’s Axiom MA, it is well-known that MA implies $h = 2^{\aleph_0}$. Hence, by the facts mentioned above, MA + $\neg$CH implies that all $\Sigma_1^1$ sets have the classical Ramsey property.

A similar argument for the dualized case does not work: Brendle has shown in [Br00-2], that MA + $2^{\aleph_0} > \mathfrak{h} = \aleph_1$ is consistent with ZFC. However, this result does not preclude that MA + $\neg$CH might imply that every $\Sigma_1^1$ set is dual Ramsey.

At the next level we like to mention the following

Fact 5.3. “Every $\Delta_1^3$ set is dual Ramsey” is consistent with ZFC.

Proof. An $\omega_1$-iteration (or also an $\omega_2$-iteration) with countable support of dual Mathias forcing starting from $L$ yields a model in which every $\Delta_1^3$ set is dual Ramsey. The proof is exactly the same as the proof of the corresponding result for the classical Ramsey property given in [JuSh93]; the reason for this is that they needed only that Mathias forcing is proper and has the homogeneity property, but these two properties hold also for dual Mathias forcing. $\square$

The $\Delta_1^3$ level is probably as far as we can get in ZFC without further assumptions. It is a famous open question whether the consistency of “Every $\Pi_3^1$ set is Ramsey” implies the existence of an inner model with an inaccessible cardinal (cf. [Kan94, Question 11.16] & [Rai84, p. 49]). Most likely, the dualized question is equally hard to conquer.

6. Determinacy and the dual Ramsey property.

We shall move on to arbitrary projective sets in this section. As we mentioned earlier, this means that we probably have to go beyond ZFC.

In [CaSi84, Section 5], the authors prove in fact that in the Solovay model constructed by collapsing an inaccessible cardinal to $\omega_1$ every projective set is dual Ramsey. As we remarked, it is unknown whether the inaccessible cardinal is necessary for that.

But there is another question connected to the dual Ramsey property of projective sets: As with the standard Ramsey property we can ask whether
an appropriate amount of determinacy implies the dual Ramsey property. As usually with regularity properties of sets of reals we would expect that Det($\Pi_n^1$) implies the dual Ramsey property for all $\Sigma_{n+1}^1$ sets. But a direct implication using determinacy is not as easy as with the more prominent regularity properties (as Lebesgue measurability and the Baire property) since the games connected to the dual Ramsey property (the Banach–Mazur games in the dual Ellentuck topology) cannot be played using natural numbers.

The same problem had been encountered with the standard Ramsey property and had been solved in [HarKe81] by making use of the scale property and the Periodicity Theorems 2.8 and 2.9:

**Theorem 6.1.** If Det($\Delta_{2n+2}^1$), then every $\Pi_{2n+2}^1$ set is Ramsey.

The main ingredient of this proof was an analysis of the models $L[T_{2n+1}]$ under Determinacy assumptions (Lemma 6.3). In the following we shall give a brief review of the result with sketches of an adaptation to our context.

**Lemma 6.2.** Let $\Gamma$ be any $\omega$-parametrized pointclass. If $U$ is $\omega$–universal for $\Gamma$, $A \in \Gamma$, $N$ any model, and $T \in N$ a tree such that $p[T] = U$. Then there is a tree $S \in N$ such that $p[S] = A$.

**Proof.** This is basically [Kan94, Proposition 13.13 (g)], apart from the assertion that $S \in N$. But this is clear since the reduction function reducing $A$ to $U$ is just the trivial function $x \mapsto \langle n_0, x \rangle$ (where $n_0$ is the index of $A$ in $U$) and hence in $N$. □

**Lemma 6.3.** Assume Det($\Delta_{2n+2}^1$). Then $[\omega]^\omega \cap L[T_{2n+1}] = C_{2n+2}$. In particular, this is a countable set.

**Proof Sketch.** We shall very roughly sketch the argument of [HarKe81, Theorem 7.2.1]:

First of all $[\omega]^\omega \cap L[T_{2n+1}]$ is easily seen to be $\Sigma_{2n+2}^1$. That every countable $\Sigma_{2n+2}^1$ set of reals is a member of $L[T_{2n+1}]$ follows directly from Mansfield’s Theorem (cf. [Kan94, Theorem 14.7]) and Lemma 6.2.

So, what is left to show is that $[\omega]^\omega \cap L[T_{2n+1}]$ actually is countable. The proof uses the following steps:

(i) Fix a $\Pi_{2n+1}^1$ norm $\varphi^* : P_{2n+1} \to \delta_{2n+1}^1$ which exists according to Theorem 2.8.

(ii) Using $\varphi^*$, code the tree $T_{2n+1}$ by some $A \subseteq \delta_{2n+1}^1$ and show that $L[T_{2n+1}] = L[A]$.

(iii) For arbitrary subsets $X \subseteq \delta_{2n+1}^1$, define $X^* := \{ z \in [\omega]^\omega : \varphi^*(z) \in X \}$.

Note that by Theorem 2.10 the choice of the complete set for the definition of the $L[T_{2n+1}]$ doesn’t matter.
(iv) Show: If \( X^* \in \Delta^1_{2n+2} \), then the set \( [\omega]^\omega \cap L[X] \) is contained in a countable \( \Sigma^1_{2n+2} \) set.

(v) Compute: \( A^* \in \Delta^1_{2n+2} \).

\[ \square \]

Harrington and Kechris used this result to receive results about projective sets from PD alone that formerly could only be derived from stronger hypotheses. The results for the classical Ramsey property follows Solovay’s argument for the \( \Sigma^1_2 \) case. We shall outline this argument in full generality and then apply it to the dual Ramsey property.

At first we need to relativize Lemma 6.3 in two different parameters:

**Lemma 6.4.** Assume \( \text{Det}(\Delta^1_{2n+2}) \). Let \( x, y \in [\omega]^\omega \) be any real numbers. Then \( [\omega]^\omega \cap L[T^y_{2n+1}, x] = C_{2n+2}(x \oplus y) \).

**Proof Sketch.** As an immediate relativization of Lemma 6.3 (for the pointclass \( \Pi^1_{2n+1} \)) instead of \( \Pi^1_{2n+1} \), we get:

\[ [\omega]^\omega \cap L[T^y_{2n+1}] = C_{2n+2}(y). \]

To show that \( [\omega]^\omega \cap L[T^y_{2n+1}, x] \) is a countable set, we have to relativize (iv) and (v) again. The obvious relativization of (iv) is (iv∗) If \( X^* \in \Delta^1_{2n+2}(x \oplus y) \), then the set \( [\omega]^\omega \cap L[X] \) is contained in a countable \( \Sigma^1_{2n+2}(x \oplus y) \) set.

Since \( L[T^y_{2n+1}] = L[A] \) for some \( A \subseteq \delta^1_{2n+1} \) such that \( A^* \) is \( \Delta^1_{2n+2}(y) \) according to (ii) and (v), we know that \( L[T^y_{2n+1}, x] = L[A, x] \). Thus we have to find a set \( B \subseteq \delta^1_{2n+1} \) such that \( L[B] = L[A, x] \) and \( B^* \in \Delta^1_{2n+1}(x \oplus y) \). This would prove the theorem.

The natural choice for \( B \) is:

\[ B := \{ \alpha \in \delta^1_{2n+1} : (\alpha \geq \omega \wedge \alpha \in A) \text{ or } (\alpha < \omega \wedge \exists n(\alpha = 2n \wedge n \in A)) \text{ or } (\alpha < \omega \wedge \exists n(\alpha = 2n + 1 \wedge n \in x)) \}. \]

Obviously, \( B \in L[A, x] \) and \( A \) and \( x \in L[B] \), so \( L[A, x] = L[B] \). Thus, what is left is to show that \( B^* \) is \( \Delta^1_{2n+2}(x \oplus y) \). But this is easy to see using the definition of \( B \), and the facts that \( \varphi^* \) was a \( \Pi^1_{2n+1} \) norm and that \( A^* \) was \( \Delta^1_{2n+2}(y) \).

The following lemma is an obvious generalization of Shoenfield’s Absoluteness Lemma (cf. also [Mo80, Theorem 8G.10]):

**Lemma 6.5.** \( \Sigma^1_{2n+2}(x) \) formulae are absolute for models containing \( T^y_{2n+1} \), i.e., if \( N \) is a model with \( T^y_{2n+1} \in N \) and \( \varphi \) is a \( \Sigma^1_{2n+2} \) formula, then

\[ \forall X \in N \ (N \models \varphi[X, x] \iff V \models \varphi[X, x]). \]
Proof. By Theorem 2.10, we can assume that $T^{x}_{2n+1}$ was constructed using an $\omega$-universal set for $\Pi^{1}_{2n+1}(x)$, enabling us to use Lemma 6.2.

Thus every $\Pi^{1}_{2n+1}(x)$ set is represented by a tree $S \in N$. We easily get a tree $S^*$ for each $\Sigma^{1}_{2n+2}(x)$ set (cf. [Kan94, Proposition 13.13 (d)]).

But now the theorem follows from standard absoluteness of illfoundedness as in Shoenfield’s proof (cf. [Kan94, Exercise 12.9 (a)]). □

Theorem 6.6. Let $x \in [\omega]^{\omega}$ be a real. Suppose that there is a dual Mathias generic partition over $L[T^{x}_{2n+1}]$. Then every $\Sigma^{1}_{2n+2}(x)$ set is dual Ramsey.

Proof. Let $A$ be a $\Sigma^{1}_{2n+2}(x)$ set and $\varphi$ a $\Sigma^{1}_{2n+2}(x)$ expression describing $A$, i.e.,

$$\forall y \ (y \in A \iff \varphi[y]).$$

By Fact 2.1 we find a $M^*$-condition $\langle \emptyset, X \rangle \in L[T^{x}_{2n+1}]$ such that

either $\langle \emptyset, X \rangle \Vdash \varphi(pc(X_G))$ or $\langle \emptyset, X \rangle \Vdash \neg \varphi(pc(X_G))$,

where $X_G$ is the name for a dual Mathias generic partition.

Without loss of generality, we assume the former. By our assumption, we actually have a generic partition $Z$ over $L[T^{x}_{2n+1}]$ with $\langle \emptyset, X \rangle \in G_Z$, where $G_Z$ is the filter associated to $Z$, i.e.,

$$\langle t, Y \rangle \in G_Z \iff Z \in (t, Y)^{\omega}.$$  

This means that $Z \in (X)^{\omega}$. Now by 2.1 (homogeneity of $M^*$) again, every element $Z^*$ of $(Z)^{\omega}$ is also $M^*$-generic over $L[T^{x}_{2n+1}]$. Since $G_{Z^*} \subseteq G_Z$, we still have $\langle \emptyset, X \rangle \in G_{Z^*}$. Consequently, we have $L[T^{x}_{2n+1}][Z^*] \models \varphi(pc(Z^*))$. But $\varphi$ was absolute for models containing $T^{x}_{2n+1}$ by Lemma 6.5, hence we have $V \models \varphi(pc(Z^*))$.

Summing up, we have found a partition $Z$ such that $\{pc(Z^*) : Z^* \subseteq Z\} \subseteq A$. This is exactly what we had to show by Observation 2.5. □

Note that this type of argument probably will not work if you replace “dual Ramsey” by “completely dual Ramsey”. What you would have to do is to relativize the argument to arbitrary partitions $W \in V$. But at least this does not work in the classical case: Brendle has shown in [Br00-1] that in any model containing one Mathias real over a ground model $N$, there is an Ellentuck neighbourhood that doesn’t contain any Mathias reals over $N$.

Another useful comment about Theorem 6.6 is that if you look at the case $n = 0$ you get a third proof of the consistency of “Every $\Sigma^1_2$ set is dual Ramsey”:

Corollary 6.7. Suppose that for each real $x \in [\omega]^{\omega}$ there is a dual Mathias generic partition over $L[x]$. Then every $\Sigma^1_2$ set is dual Ramsey.

Proof. Immediate from Theorem 6.6, keeping in mind that $T_1 \in L$ by [KeMo78, 9C], as mentioned in Subsection 2.7. □
This particularly generic version of proving the consistency of properties of \( \Sigma^1_2 \) sets should be compared to analogous results for Random forcing, Cohen forcing and Hechler forcing.\(^{17}\)

We now move on to use Lemma 6.4 and Theorem 6.6 to get that Determinacy implies the dual Ramsey property:

**Corollary 6.8.** Assume \( \text{Det}(\Delta^1_{2n+2}) \). Then every \( \Sigma^1_{2n+2} \) set is dual Ramsey.

**Proof.** By Lemma 6.4, we get that for any reals \( x, y \in [\omega]^\omega \), the set of reals in \( L[T^x_{2n+1}, y] \) is countable.

By the same argument that is used to show that \( \forall x (\aleph_1^{L[x]} < \aleph_1^{V}) \) implies that \( \aleph_1^V \) is strongly inaccessible in every \( L[x] \), we get that \( P(\omega^\omega) \cap L[T^x_{2n+1}, y] \) is countable for arbitrary choices of \( x \) and \( y \in [\omega]^\omega \).

Thus there are dual Mathias generic partitions over each \( L[T^x_{2n+1}, y] \), in particular over each \( L[T^x_{2n+1}] \), and we can use Theorem 6.6 to prove the claim. \( \square \)

7. Appendix: Game-filters have the segment-colouring-property.

Let \( \mathcal{F} \subseteq (\omega)^\omega \) be an ultrafilter. Associated with \( \mathcal{F} \) we define the game \( G_\mathcal{F} \) as follows. This type of game, which is the Choquet-game with respect to the dual Ellentuck topology (cf. [Ke95, 8.C]), was first suggested by Kastanas in [Kas83].

\[
\begin{align*}
\text{I} & \langle t_0, Y_0 \rangle & \langle t_1, Y_1 \rangle & \langle t_2, Y_2 \rangle & \ldots \\
\text{II} & \langle X_0 \rangle & \langle X_1 \rangle & \langle X_2 \rangle & \ldots 
\end{align*}
\]

All the moves \( X_n \) of player II must be elements of the ultrafilter \( \mathcal{F} \) and all the moves \( \langle t_n, Y_n \rangle \) of player I plays must be such that \( Y_n \in \mathcal{F} \) and \( (t_n, Y_n)^\omega \) is a dual Ellentuck neighbourhood. Further, the \( n \)th move \( X_n \) of player II is such that \( X_n \in (t_n^*, Y_n)^\omega \) and then player I plays \( t_{n+1} \) such that \( t_n^* \leq t_{n+1} \subseteq X_n \) and \( |t_{n+1}| = |t_n| + 1 = |t_0| + n + 1 \). Player I wins if and only if the unique \( Y \) with \( t_n \leq Y \) (for all \( n \)) is not in \( \mathcal{F} \).

An ultrafilter \( \mathcal{F} \) is a **game filter** if and only if player I has no winning strategy in the game \( G_\mathcal{F} \).\(^{18}\)

In the following we outline the proof that game filters are also scp-filters. The crucial point will be to show the Preliminary Lemma 7.1, which is in

\(^{17}\)Cf. [BrLö99]. Note that in most cases the existence of generics doesn’t give more than regularity at the \( \Delta^1_2 \) level, and something more than mere existence is needed for the \( \Sigma^1_2 \) level.

\(^{18}\)For the existence of game filters see [Halb98-1], where one can find also some results concerning dual Mathias forcing restricted to such filters.
fact Carlson’s Lemma (cf. [CaSi84, Lemma 2.4]) restricted to game filters. But first we have to give some notations.

Let $s,t \in \mathbb{N}$ be such that $s \subseteq t$, $|s| = n$ and $|t| = m$. For $k$ with $k \leq m - n$ let

$$(t)_s^k := \{ u \in \mathbb{N} : \text{dom}(u) = \text{dom}(t) \wedge (s \subseteq u \subseteq t \wedge |u| = |s| + k) \}.$$

For $s \nsubseteq t \subseteq X$, let

$$(t,X)_s^k := \{ u \in \mathbb{N} : t \nsubseteq u^* \subseteq X \wedge |u| = |s| + k \},$$

and let $(X)_s^{k*} := (s,X)^{(n+k)*}$.

We have chosen this notation following [CaSi84, Definition 2.1], where one can consider $s$ as an alphabet of cardinality $n$.

For the remainder of this section, let $\mathfrak{F}$ be an arbitrary but fixed game filter.

**Preliminary Lemma 7.1.** Let $s \nsubseteq X \in \mathfrak{F}$ and $\pi : (X)_s^{0*} \rightarrow l$, then there exists a $Y \in (s,X)^{\omega} \cap \mathfrak{F}$ such that $\pi|(Y)_s^{0*}$ is constant.

Following the ideas of the proof of Theorem 6.3 of [CaSi84], the proof of the Preliminary Lemma will be given in a sequence of lemmas. We start by stating the well-known Hales–Jewett Theorem in our notation.

**Hales-Jewett Theorem 7.2.** Let $s \in \mathbb{N}$. For all $d \in \omega$, there is an $h \in \omega$ such that for any $t \in \mathbb{N}$ with $s \nsubseteq t$ and $|t| = |s| + h$, and for any colouring $\tau : (t)_s^h \rightarrow d$, there is a $u \in (t)_s^h$ such that $(u)_s^0$ is monochromatic.

The number $h$ in the Hales–Jewett Theorem depends only on the number $d$ and the size of $|s|$. Let $\text{HJ}(d,|s|)$ denote the smallest number $h$ which verifies the Hales–Jewett Theorem.

Let $s,t \in \mathbb{N}$ and $X \in \mathfrak{F}$ be such that $s \nsubseteq t \subseteq X$. A set $K \subseteq \mathbb{N}$ is called dense in $(t,X)^{\omega}$, if for all $Y \in (t,X)^{\omega} \cap \mathfrak{F}$, there is a $u$ with $t \nsubseteq u^* \subseteq Y$ which belongs to $K$. A set $D \subseteq (\omega)_s^{k*}$ is called k-dense in $(t,X)_s^{k*}$, if for all $Y \in (t,X)^{\omega} \cap \mathfrak{F}$, we have $(t,Y)_s^{k*} \cap D \neq \emptyset$.

**Lemma 7.3.** Let $s \nsubseteq t \nsubseteq X \in \mathfrak{F}$ and assume that $D \subseteq (\omega)_s^{0*}$ is 0-dense in $(X)_s^{0*}$. Further assume that $K = \{ u : t \nsubseteq u \wedge (u)_s^0 \cap D \neq \emptyset \}$ is dense in some $(t,Z)^{\omega}$, where $Z \in (t,X)^{\omega} \cap \mathfrak{F}$. Then there is an $s^* \subseteq Z$ with $t \nsubseteq s$ such that for all $v^* \subseteq Z$ with $s \nsubseteq v$ we have $(v)_s^0 \cap D \neq \emptyset$.

**Proof.** We shall define a strategy for player I in the game $G_\mathfrak{F}$, such that player I can follow this strategy just in the case when Lemma 7.3 fails. This means that for every $s^* \subseteq Z$ with $t \nsubseteq s$ there is a $v^* \subseteq Z$ with $s \nsubseteq v$ such that $(v)_s^0 \cap D = \emptyset$.

Let $t_0^s \subseteq Z$ be such that $t \nsubseteq t_0$, $|t| = |t_0|$ and $(t_0)_s^0 \cap D = \emptyset$. Further put $Y_0 = t_0 \cap Z$ and player I plays $\langle t_0,Y_0 \rangle$. Assume $(t_m,Y_m)$ is the $m$th move of player I and player II replies with $(X_m)$. If the lemma fails with $s = t_m^*$,
player I can play \( (t_{m+1}, Y_{m+1}) \), according to the rules of the game, such that \((t_{m+1})^0_s \cap D = \emptyset \).

Since \( \mathcal{F} \) is a game filter, the strategy of player I is not a winning strategy and the unique \( Y \in (Z)^\omega \) such that \( t_m \preceq Y \) (for all \( m \in \omega \)) belongs to \( \mathcal{F} \). Take an arbitrary \( u \) with \( t \preceq u^* \subsetneq Y \). For such a \( u \) we find a \( t_n \preceq Y \) such that \( u \subsetneq t_n \) and \( \text{dom}(u) = \text{dom}(t_n) \). By the strategy of player I we have \((t_n)^0_s \cap D = \emptyset \) and therefore \((u)^0_s \cap D = \emptyset \). But this is a contradiction to the assumption that \( K \) is dense in \((t,Z)^\omega \). Hence, player I cannot follow this strategy, which completes the proof. \( \square \)

**Lemma 7.4.** Suppose \( s \preceq t \preceq X \in \mathcal{F} \) and \( D \) is 0-dense in \((X)^0_s \) and \( K = \{u : t \preceq u \wedge (u)^0_s \cap D \neq \emptyset \} \) is dense in \((t,Z)^\omega \), where \( Z \in (t,X)^\omega \cap \mathcal{F} \). Then there is a \( t^* \subsetneq Z \) with \( t \preceq t^* \) and \( |t^*| = |t| + 1 \) such that \((t^*)^0_s \subsetneq D \).

**Proof.** Let \( s \) be as in the Lemma 7.3, and let \( d := (\langle s \rangle^0_s) \). By the Hales–Jewett Theorem, let \( h := HJ(d, |s|) \). Pick \( v \in (N) \) such that \( s \preceq v^* \subsetneq X \) and \( |v| = |s| + h \). Let \( \{s_i : s_i \in (\langle s \rangle^0_s) \wedge i \in d \} \) be an enumeration of the elements of \((\langle s \rangle^0_s) \). We colour \((v)^0_s \) by stipulating \( \tau(u) = i \) if and only if \( s_i \preceq u \wedge u \in D \). By the choice of \( n \), there are \( t \in (v)^1_s \) such that \((\tilde{t})^0_s \) is monochromatic, and therefore, \((\tilde{t})^0_s \subsetneq D \). Thus, we have found a \( t \) with \( t \preceq \tilde{t} \) and \( |\tilde{t}| = |t| + 1 \) such that \((\tilde{t})^0_s \subsetneq D \). \( \square \)

**Lemma 7.5.** Suppose \( s \preceq X \in \mathcal{F} \) and \( D \) is 0-dense in \((X)^0_s \). Then there are \( t \in (X)^1_s \) and \( Y \in (t,X)^\omega \cap \mathcal{F} \) such that \( \{u : t \preceq u \wedge (u)^0_s \subsetneq D \} \) is 1-dense in \((t,Y)^1_s \).

**Proof.** In a similar way as above we can define a strategy for player I in the game \( G_{\mathcal{F}} \), such that player I can follow this strategy only if Lemma 7.4 fails. But if Lemma 7.4 is wrong, this would yield – because \( \mathcal{F} \) is a game filter – a contradiction (cf. [CaSi84, Lemma 6.5]). \( \square \)

Notice that in Lemma 7.5 we did not require that the \( r \in (N) \) for which we have \( r^* \preceq t \in (X)^1_s \) belongs to \( D \). This we do in

**Lemma 7.6.** Suppose \( s \preceq X \in \mathcal{F} \) and \( D \) is 0-dense in \((X)^0_s \). Then there are \( r^* \preceq t \in (X)^1_s \) and \( Y \in (t,X)^\omega \cap \mathcal{F} \) such that \( \{u : t \preceq u \wedge (u)^0_s \subsetneq D \} \) is 1-dense in \((t,Y)^1_s \) and \( r \in D \).

**Proof.** Let \( Y_0 \in (t_0,X)^\omega \cap \mathcal{F} \) be as in the conclusion of Lemma 7.5. Thus \( D_0 := \{u : t_0 \preceq u \wedge (u)^0_s \subsetneq D \} \) is 1-dense in \((t_0,Y_0)^1_s \). Using Lemma 7.5, player I can play \( \langle t_m, Y_m \rangle \) at the \( m \)th move such that \( D_m := \{u : t_m \preceq u \wedge (u)^0_s \subsetneq D \} \) is \((m + 1)\)-dense in \((t_m,Y_m)^{(m+1)*}_s \).

Because player I has no winning strategy, the unique \( Y \in (s,X)^\omega \) such that \( t_m \preceq Y \) (for all \( m \)) belongs to \( \mathcal{F} \), and because \( D \) is 0-dense in \((X)^0_s \), there is an \( r \in (Y)^1_s \) which belongs to \( D \). Let \( t \in (t_m)^1_s \) be such that \( r^* \preceq \tilde{t} \). Since \((\tilde{t})^0_s \subsetneq (t_m)^0_s \) and because \( D_m \) is \((m + 1)\)-dense in \((t_m,Y_m)^{(m+1)*}_s \) we
get \( \{ u : \bar{t} \preceq u \land (u)_s^0 \subseteq D \} \) is 1-dense in \( (\bar{t}, Y_m)_s^{1*} \). Hence, we have found an \( r^* \) such that \( \{ u : r^* \preceq u \land (u)_s^0 \subseteq D \} \) is 1-dense in \( (r^*, Y)_s^{1*} \) and \( r \in D \). \( \square \)

Now we can go back to the Proof of the Preliminary Lemma. Let \( s \preceq X \in \mathfrak{F} \). We have to show that for any colouring \( \pi : (X)_s^{0*} \to \bar{t} \), there is a \( Y \in (s, X)^{\omega} \cap \mathfrak{F} \) such that \( (Y)_s^{0*} \) is monochromatic.

It is easy to see that at least one of the colours is 0-dense in \( (X)_s^{0*} \), say \( j \) and let \( D := \{ t \in (X)_s^{0*} : \pi(t) = j \} \). Now we can prove the Preliminary Lemma in almost the same way as Lemma 7.6, the only difference is that player I uses now Lemma 7.6 to construct the \( m \)th move, instead of Lemma 7.5. \( \square \)

Finally we get the main result of this section.

**Proposition 7.7.** Each game filter is also an scp-filter.

**Proof.** We have to show that for any colouring \( \pi : (s, \omega)^{(|s|+k)^*} \to r \), where \( r \) and \( k \) are positive natural numbers and \( s \in (N) \), there is an \( X \in \mathfrak{F} \) such that \( s \preceq X \) and \( (s, X)^{(|s|+k)^*} \) is monochromatic.

Following the proof of [Halb\( \infty \), THEOREM] and using the Preliminary Lemma, it is not hard to define a strategy for player I in such a way that if player I follows this strategy, then for the resulting partition \( X \) – which must belong to \( \mathfrak{F} \), since \( \mathfrak{F} \) is a game filter – we get \( s \preceq X \) and \( (s, X)^{(|s|+k)^*} \) is monochromatic. \( \square \)

We have seen that every game filter has the segment-colouring property. It seems that the reverse implication is unlikely, since a strategy for player I cannot be encoded by a real number, which makes it hard (if not impossible) to prove that CH implies the existence of game filters. But on the other hand we know that scp-filters always exist if we assume CH.

**References**


Received September 14, 1999. The first author wishes to thank the Swiss National Science Foundation for its support during the time at U.C. Berkeley in which the research for this paper has been done. The second author wishes to thank the Studienstiftung des deutschen Volkes for the funding of his travels to Boise ID and Las Vegas NV, and in particular Gunter Fuchs (Berlin).

Queen’s University Belfast
Belfast BT7 1NN
Northern Ireland
E-mail address: halbeis@qub.ac.uk

Mathematisches Institut
Rheinische Friedrich-Wilhelms-Universität Bonn
Beringstrasse 6
53115 Bonn
Germany
E-mail address: loewe@math.uni-bonn.de
ELLIPITC SURFACES AND AMPLE VECTOR BUNDLES

ANTONIO LANTERI AND HIDETOSHI MAEDA

Let $E$ be an ample vector bundle of rank $n - 2 \geq 2$ on a complex projective manifold $X$ of dimension $n$ having a section whose zero locus is a smooth surface $Z$. We determine the structure of pairs $(X, E)$ as above under the assumption that $Z$ is a properly elliptic surface. This generalizes known results on threefolds containing an elliptic surface as a smooth ample divisor. Among the applications we prove a conjecture relating the Kodaira dimension of $X$ to that of $Z$, and we show that if $0 \leq \kappa(Z) \leq 1$, then $p_g(Z) > 0$ unless $X$ is a $\mathbb{P}^{n-2}$-bundle over a smooth surface $S$ with $p_g(S) = 0$.

Introduction.

Consider the following set-up.

(*) $X$ is a smooth complex projective variety of dimension $n$ and $E$ is an ample vector bundle of rank $n - 2 \geq 2$ on $X$ such that there exists a section $s \in \Gamma(E)$ whose zero locus $Z := (s)_0$ is a smooth surface.

According to a general philosophy coming from the study of ample divisors $[S1]$, if $Z$ is special, then $X$ has to be special as well. By relying on the study of the nefness of the adjoint bundle $K_X + \text{det } E [M]$, the structure of pairs $(X, E)$ as above is now well understood when $\kappa(Z) \leq 0$ in view of $[LM1]$ and $[L1]$. In this paper we investigate the structure of $(X, E)$ when $Z$ is a properly elliptic surface.

Our result is as follows.

**Theorem.** Let $X$, $E$ and $Z$ be as in (*) and assume further that $Z$ is an elliptic surface with $\kappa(Z) = 1$. Then one of the following conditions holds.

(a) $X = \mathbb{P}_S(F)$, where $F$ is an ample vector bundle of rank $n - 1$ over a smooth surface $S$, $E = \pi^* V \otimes H$, where $H = H(F)$ is the tautological line bundle on $X$, $V$ is a vector bundle of rank $n - 2$ on $S$, $\pi : X \to S$ is the bundle projection and its restriction to $Z$, $\pi|_Z : Z \to S$ is a birational morphism, but not an isomorphism.

(b) There exist a birational morphism $f : X \to X'$ expressing $X$ as a projective manifold $X'$ blown-up at a finite set $B$ of points (possibly empty) and an ample vector bundle $E'$ of rank $n - 2$ on $X'$ such that $E = f^* E' \otimes [-f^{-1}(B)]$ and that $K_{X'} + \text{det } E'$ is nef. The triplet
\((X', \mathcal{E}', Z' := f(Z))\) satisfies \((*)\) and \(\kappa(Z') = 1\). Moreover \(X'\) is endowed with a morphism \(\varphi : X' \to Y\) onto a smooth curve \(Y\) of genus \(g(Y) = h^{1,0}(Z)\), whose general fibre \(F\) is a projective manifold of dimension \(n - 1\) satisfying the condition \(K_F + \det \mathcal{E}'_F = \mathcal{O}_F\); \(\varphi\) induces on \(Z'\) the elliptic fibration, and \(\varphi|_{Z'} : Z' \to Y\) has no multiple fibres.

Pairs \((F, \mathcal{E}'_F)\) of this kind have been classified in [PSW].

In the 80’s Sommese investigated the structure of projective threefolds containing an elliptic surface as a hyperplane section, or more generally as an ample divisor [S2, Theorem 3.1]. Our Theorem can be viewed as a complete generalization of what Sommese and Shepherd-Barron proved in that setting (see [D, Theorem 0.7]). In the threefold case Sommese proved that the elliptic fibration cannot have multiple fibres [S2, Claim 3.1.4] (see also [S3, Lemma 0.5.1]), by relying on a formula he obtained for the plurigenera of the surface. Our proof of the fact that \(\varphi|_{Z'} : Z' \to Y\) has no multiple fibres is conceptually much easier and we derive from this fact a formula expressing both \(h^0(m(K_{X'} + \det \mathcal{E}'_F))\) and the \(m\)-th genus \(P_m(Z)\) for \(m > 0\) as a linear polynomial in \(m\) (Corollary (1.3)). Moreover we show that in case (b) the surface \(Z\) must have positive geometric genus, which was not explicitly noted even in the setting of ample divisors (Corollary (2.2)). In particular, if \(p_g(Z) = 0\) then \((X, \mathcal{E})\) is as in case (a). We show that this holds also when \(\kappa(Z) = 0\), which gives some more evidence for a conjecture stated in [L2]. As a consequence of our Theorem we also see that a product of an elliptic curve with a smooth curve of positive genus can never occur as the zero locus of a section of an ample vector bundle (Corollary (1.4)).

As another application of our Theorem we prove that if \(X, \mathcal{E}\) and \(Z\) are as in \((*)\) and \(\kappa(Z) < \dim Z\), then \(\kappa(X) = -\infty\) (Corollary (2.1)). This fact, which was conjectured in [LM1], also comes from a recent, more general, theory of normal pairs with \(\mathbb{Q}\)-effective normal sheaf, developed by Peternell, Schneider and Sommese [PSS].

The present paper was prepared in the framework of the National Research Project “Algebraic Geometry, Commutative Algebra and Computational Aspects”, supported by the M.U.R.S.T. of the Italian Government. The second author was also partially supported by Grant-in-Aid for Encouragement of Young Scientists (No. 11740027), the Ministry of Education, Science, Sports and Culture, Japan, as well as Waseda University Grant for Special Research Projects (No. 99A-192). Both authors would like to thank the M.U.R.S.T. and the University of Milan for making this collaboration possible.

0. Preliminaries.

In this paper we will work over the complex number field \(\mathbb{C}\). We use the standard notation from algebraic geometry. We make no distinction between
vector bundles and locally free sheaves. The tensor products of line bundles are denoted additively, while we use multiplicative notation for intersection products in Chow rings. The pull-back $i^* \mathcal{E}$ of a vector bundle $\mathcal{E}$ on $X$ by an embedding $i : Y \hookrightarrow X$ is denoted by $\mathcal{E}_Y$. If $X$ is a smooth projective variety, the canonical bundle of $X$ is denoted by $K_X$ and $\kappa(X)$ stands for the Kodaira dimension of $X$. For any $m \geq 1$ the $m$-th plurigenus $h^0(mK_X)$ is denoted by $P_m(X)$. In particular, for $m = 1$, we denote by $p_g(X) = P_1(X)$ the geometric genus of $X$. Moreover, if $X$ is a smooth projective curve, the genus of $X$ is denoted by $g(X)$ instead of $p_g(X)$.

Let $X, \mathcal{E}$ and $Z$ be as in $(\ast)$. Since $Z$ is smooth of the expected dimension, we have $N_{Z/X} \cong E_Z$, where $N_{Z/X}$ denotes the normal bundle of $Z$ in $X$. Hence, by adjunction, we get

$$K_Z = (K_X)_Z + \det N_{Z/X} = (K_X + \det \mathcal{E})_Z.$$ 

We will use this fact over and over. The following fact follows easily from [M, Theorem].

**Lemma 0.1.** Let $X, \mathcal{E}$ and $Z$ be as in $(\ast)$ and assume that $\kappa(Z) \geq 0$. Then either $K_X + \det \mathcal{E}$ is nef, or one of the following cases occurs:

(i) $X$ contains an effective divisor $E$ such that

$$(E, \mathcal{O}_E(E), \mathcal{E}_E) \cong (\mathbb{P}^{n-1}, \mathcal{O}_{\mathbb{P}^{n-1}}, \mathcal{O}_{\mathbb{P}^{n-1}}(1)^{\oplus (n-2)})$$;

(ii) $X = \mathbb{P}_S(\mathcal{F})$ for some ample vector bundle $\mathcal{F}$ of rank $n-1$ on a smooth surface $S$ and $\mathcal{E} = \pi^* \mathcal{V} \otimes H$ for some vector bundle $\mathcal{V}$ of rank $n-2$ on $S$, where $H$ is the tautological line bundle on $X$ and $\pi : X \to S$ is the bundle projection.

We recall the following fact.

**Lemma 0.2** ([LM1, Lemma 5.1]). Let $\mathcal{E}$ be an ample vector bundle of rank $r \geq 2$ on a smooth projective variety $X$ of dimension $n \geq 2$ and assume that $X$ contains a smooth divisor $E$ such that

$$(E, \mathcal{O}_E(E), \mathcal{E}_E) \cong (\mathbb{P}^{n-1}, \mathcal{O}_{\mathbb{P}^{n-1}}, \mathcal{O}_{\mathbb{P}^{n-1}}(1)^{\oplus r}).$$

Let $f : X \to X'$ be the blow-down of $E$ to a point $x \in X'$. Then there exists an ample vector bundle $\mathcal{E}'$ of rank $r$ on $X'$ such that

$$\mathcal{E} = f^* \mathcal{E}' \otimes \mathcal{O}_X(-E).$$

(0.2.1)

In connection with this we prove the following lemma, inspired by [L1, Lemma 2.2].

**Lemma 0.3.** Let $(X, \mathcal{E}), E, f, x$ and $(X', \mathcal{E}')$ be as above. Then $X'$ cannot contain any submanifold $F$ of positive dimension, such that $x \in F$ and

$$(F, \mathcal{E}'_F) \cong (\mathbb{P}^s, \mathcal{O}_{\mathbb{P}^s}(1)^{\oplus r}).$$
Proof. By contradiction, assume that $X'$ contains such an $F$, and let $\tilde{F}$ be the proper transform of $F$ via $f$. Note that $\tilde{F}$ is a $\mathbb{P}^1$-bundle over $\mathbb{P}(N^*_x/F) = \mathbb{P}^{s-1}$, whose fibres are the proper transforms of the lines on $F$ passing through $x$. Moreover $E \cap \tilde{F}$ is a section of this $\mathbb{P}^1$-bundle. So, if $l$ is a fibre of $\tilde{F}$, then $O_X(E)l = O_{\tilde{F}}(E)l = 1$. Since $E'_F \cong O_F(1)^{\oplus r}$ we thus get from (0.2.1)
\[ E_i = (f^*E' \otimes O_X(-E))_l = O_l^{\oplus r}, \]
contradicting the ampleness of $E$.

Remark 0.4. In particular Lemma 0.3 shows that for the pair $(X', E')$ obtained after the contraction $f$ above,

1. $X'$ cannot be a $\mathbb{P}^s$-bundle, $s \geq 1$, over a smooth projective variety $W$, $\dim W \geq 0$, with $E'_F = O_F(1)^{\oplus r}$ for any fibre $F$ of the projection $X' \to W$, and

2. if $X'$ contains an exceptional divisor $E'$ such that

\[ (E', O_{E'}(E'), E'_E) \cong (\mathbb{P}^{n-1}, O_{\mathbb{P}}(-1), O_{\mathbb{P}}(1)^{\oplus r}), \]

then $E$ and $f^{-1}(E')$ are two disjoint exceptional divisors in $X$.

The following lemma will be useful in Section 1 to add something more in case (b) of our Theorem.

Lemma 0.5. Let $Z$ be a minimal elliptic surface with $κ(Z) = 1$, and let $ψ : Z \to Y$ denote the elliptic fibration over a smooth projective curve $Y$. If $ψ$ has no multiple fibres and if $K_Z = ψ^*M$ for some ample line bundle $M$ on $Y$, then the $m$-th plurigenus of $Z$ is given by the formula

\[ P_m(Z) = m(χ(O_Z) + 2g(Y) - 2) + 1 - g(Y) \quad \text{for every} \quad m \geq 2. \]

Moreover, (0.5.1) is valid even for $m = 1$, unless $M = K_Y$.

Proof. Assume first that $m \geq 2$. We note that $χ(mK_Z) = χ(O_Z)$, because $K_Z^2 = 0$. Moreover, by Serre duality we have $h^2(mK_Z) = h^0(-K_Z - (m-1)K_Z)$. On the other hand, $K_ZA > 0$ for any ample line bundle $A$ on $Z$, since $κ(Z) = 1$. So $h^2(mK_Z) = 0$. Consequently

\[ P_m(Z) = h^0(mK_Z) - h^1(mK_Z) + h^2(mK_Z) + h^1(mK_Z) - h^2(mK_Z) \]
\[ = χ(mK_Z) + h^1(mK_Z) - h^2(mK_Z) \]
\[ = χ(O_Z) + h^1(mK_Z). \]

From now on, let us calculate $h^1(mK_Z)$. Since $ψ$ has no multiple fibres, the canonical bundle formula for elliptic fibrations [F, Theorem 15, p. 176] tells us that $K_Z = ψ^*(K_Y + L)$, where $L$ is the dual of the line bundle $R^1ψ_*O_Z$. We note that $\deg L = χ(O_Z) \geq 0$ [F, Lemmas 13 and 14, p. 176]. Furthermore, since $ψ_*O_Z = O_Y$, the homomorphism $ψ^* : \text{Pic}(Y) \to \text{Pic}(Z)$
is injective. Therefore, by assumption, \( M = K_Y + L \). Now we consider the exact sequence

\[
0 \to H^1(\psi_*(mK_Z)) \to H^1(mK_Z) \to H^0(R^1\psi_*(mK_Z)) \to 0
\]

induced by the Leray spectral sequence

\[
E_2^{p,q} = H^p(R^q\psi_*(mK_Z)) \implies E^{p+q} = H^{p+q}(mK_Z).
\]

This gives

\[
h^1(mK_Z) = h^1(\psi_*(mK_Z)) + h^0(R^1\psi_*(mK_Z)).
\]

As to the first summand, by the projection formula, we see that \( \psi_*(mK_Z) = \psi_*\psi^*(mM) = mM \). Moreover, we can write \( mM = K_Y + (m - 1)M + L \). Since \( \deg((m - 1)M + L) > 0 \), we get \( h^1(\psi_*(mK_Z)) = h^1(mM) = 0 \). So it suffices to calculate \( h^0(R^1\psi_*(mK_Z)) \). From the projection formula, we obtain

\[
R^1\psi_*(mK_Z) = R^1\psi_*\psi^*(mM) = R^1\psi_*(mK_Z) = mM - L = K_Y + (m - 1)M.
\]

Since \( \deg((m - 1)M) > 0 \), the Riemann-Roch theorem for \( K_Y + (m - 1)M \) gives

\[
(0.5.3)
\]

\[
h^0(R^1\psi_*(mK_Z)) = h^0(K_Y + (m - 1)M)
= 2g(Y) - 2 + (m - 1)\deg M + 1 - g(Y)
= 2g(Y) - 2 + (m - 1)(2g(Y) - 2 + \chi(O_Z)) + 1 - g(Y).
\]

Combining (0.5.2) with (0.5.3), we get the desired formula.

When \( m = 1 \), an easy calculation shows that the right-hand side of (0.5.1) is \( P_1(Z) + g(Y) - h^1(O_Z) \). If \( M \neq K_Y \), i.e., \( L \) is not trivial, then it follows from [F, Lemma 14, p. 176] that \( g(Y) = h^1(O_Z) \). Therefore (0.5.1) is still true unless \( M = K_Y \). \( \square \)

1. The structure of \((X, \mathcal{E})\).

1.1. Proof of the Theorem. Assume first that \( K_X + \det \mathcal{E} \) is not nef. Then, since \( \kappa(Z) = 1 \), \((X, \mathcal{E})\) is either of type (i) or of type (ii) in Lemma 0.1.

Assume that \((X, \mathcal{E})\) is of type (ii). Let \( s_F \) denote the restriction of \( s \) to any fibre \( F \) of \( \pi \). Then, since \( s_F \in H^0(O_{\pi^{-1}(1)}^{\oplus(n-2)}) \), \( Z \cap F = (s_F)_0 \) is a linear subspace of dimension \( \geq 0 \) in \( F \). Hence the restriction \( \pi|_Z : Z \to S \) of \( \pi \) to \( Z \) is surjective. We thus conclude that \( \pi|_Z \) is birational. However, if \( \pi|_Z \) is an isomorphism, then \( \text{Pic}(X) \cong \text{Pic}(Z) \times Z \), which contradicts the fact that the restriction homomorphism \( \text{Pic}(X) \to \text{Pic}(Z) \) is injective [LM2, (1.1.6)]. Thus we are in case (a) of the Theorem.
Now assume that \((X, \mathcal{E})\) is of type (i). Set \(l = Z \cap E\). Then the same argument as that in the proof of [LM1, Corollary 5.2] shows that \(l\) is a \((-1)\)-curve in \(Z\). Let \(f_1 : X \to X_1\) be the blow-down of \(E\). Then \(Z_1 := f_1(Z)\) is a smooth projective surface in \(X_1\) because \(f_1|_Z : Z \to Z_1\) is nothing but the contraction of \(l\). In particular \(\kappa(Z_1) = 1\). Moreover, by Lemma 0.2 there exists an ample vector bundle \(\mathcal{E}_1\) of rank \(n - 2\) on \(X_1\) such that \(\mathcal{E} = f_1^*\mathcal{E}_1 \otimes \mathcal{O}_X(-E)\). If we let \(s_1 \in \Gamma(\mathcal{E}_1)\) denote the section corresponding to \(s\), then \((s_1)_0 = Z_1\), so that \(X_1, \mathcal{E}_1\) and \(Z_1\) also satisfy condition \((\ast)\) in the Introduction. However, note that \((X_1, \mathcal{E}_1)\) is never of type (ii) by virtue of Remark 0.4 (1). Therefore \((X_1, \mathcal{E}_1)\) must be of type (i), if \(K_{X_1} + \det \mathcal{E}_1\) is not nef. For this reason, when \(K_{X_1} + \det \mathcal{E}_1\) is not yet nef, we can apply the same argument as above to \(X_1, \mathcal{E}_1\) and \(Z_1\), and continue in this manner. It should be emphasized that if \(X_1\) contains \(E_1\) such that
\[
(E_1, \mathcal{O}_{E_1}(E_1), \mathcal{E}_{1E_1}) \cong (\mathbb{P}^{n-1}, \mathcal{O}_{\mathbb{P}}(-1), \mathcal{O}_{\mathbb{P}}(1)^{\oplus(n-2)}),
\]
then \(E \cap f_1^{-1}(E_1) = \emptyset\) by Remark 0.4 (2), so that we can contract \(E\) and \(f_1^{-1}(E_1)\) at the same time. This procedure must come to an end after a finite number of repetitions. Therefore we conclude that there exist a birational morphism \(f : X \to X'\) expressing \(X\) as a projective manifold \(X'\) blown-up at a finite set \(B\) of points and an ample vector bundle \(\mathcal{E}'\) of rank \(n - 2\) on \(X'\) satisfying condition \((\ast)\), such that \(\mathcal{E} = f^*\mathcal{E}' \otimes [-f^{-1}(B)]\) and that \(K_{X'} + \det \mathcal{E}'\) is nef unless \((X, \mathcal{E})\) is as in case (a) of the Theorem.

From this, in order to complete our analysis, it suffices to investigate the structure of \((X, \mathcal{E})\) when the adjoint bundle \(K_X + \det \mathcal{E}\) is nef. So in what follows, we assume that \(K_X + \det \mathcal{E}\) is nef. Then, by [KMM, Theorem 3-2-1] there exists a morphism \(\varphi : X \to Y\) with connected fibres from \(X\) onto a normal projective variety \(Y\) such that \(K_X + \det \mathcal{E} = \varphi^*M\) for some ample line bundle \(M\) on \(Y\).

If \(\dim Y = 0\), then \(K_X + \det \mathcal{E} = \mathcal{O}_X\). Recalling that \(K_Z = (K_X + \det \mathcal{E})_Z\), we have \(K_Z = \mathcal{O}_Z\), which contradicts our assumption on \(\kappa(Z)\). Hence \(\dim Y \geq 1\). Moreover, since \(K_X + \det \mathcal{E}\) is nef in the present case, \(K_Z\) is nef, i.e., \(Z\) is a minimal surface with \(\kappa(Z) = 1\). In particular we have \(K_Z^2 = 0\).

We first claim that \(\dim Y = 1\). To see this, suppose to the contrary that \(\dim Y \geq 2\). Then we can find effective divisors \(D_1, D_2 \in |m(K_X + \det \mathcal{E})|\) for some \(m > 0\) such that every irreducible component of \(D_1 \cap D_2\) has dimension \(n - 2\). Applying [BG, Theorem 2.5] to \(\mathcal{E}, D_1\) and \(D_2\), we have
\[
m^2K_Z^2 = m^2(K_X + \det \mathcal{E})^2Z = (D_1 \cap D_2)c_{n-2}(\mathcal{E}) > 0,
\]
which gives a contradiction. Hence \(\dim Y = 1\). As a direct consequence of this, \(Y\) is a smooth curve and \(\varphi\) is flat. Take a general fibre \(F\) of \(\varphi\). Then, since \(K_F + \det \mathcal{E}_F = \mathcal{O}_F\), \(F\) is a Fano manifold of dimension \(n - 1\) and \(\mathcal{E}_F\) is an ample vector bundle of rank \(n - 2\) on \(F\).
Next we claim that $\varphi|_Z : Z \to Y$ is the elliptic fibration. Indeed, $\varphi|_Z$ is clearly surjective, and satisfies $K_Z = \varphi^*_Z M$. Set $f := Z \cap F$ for a general fibre $F$ of $\varphi$. Then $f$ is a 1-equidimensional smooth fibre of $\varphi|_Z$. Moreover, $f$ is the zero locus of the section $s_F \in \Gamma(E_F)$. Therefore $H^0(f, Z) \cong H^0(F, Z) = Z$ by [LM2, (1.1.1)] (note that its proof is valid without assuming the connectedness of $f$). So $f$ is a smooth curve in $Z$. Combining this with the fact that $K_Z = \varphi^*_Z M$, we see that $f$ is an elliptic curve. Since $\kappa(Z) = 1$, $Z$ admits a unique elliptic fibration, and so the result is proved.

Next we claim that $Y$ has genus

$$
(1.1.1) \quad g(Y) = h^1(O_Z).
$$

To see this, take a general fibre $F$ of $\varphi$. Then $h^i(O_F) = 0$ for all $i > 0$, because $F$ is Fano. Hence $R^i\varphi_* O_X = 0$ for all $i > 0$. Since $\varphi_* O_X = O_Y$, we have

$$
H^1(O_X) = H^1(\varphi_* O_X) = H^1(O_Y).
$$

On the other hand, it follows from [LM2, (1.1.3)] that $H^1(O_X) \cong H^1(O_Z)$. Thus $h^1(O_Y) = h^1(O_Z)$, as required.

Finally we show that $\varphi|_Z$ has no multiple fibres. First note that $m(K_X + \det E) = \varphi^*(mM)$ for every $m \geq 1$. As we have seen, $R^i\varphi_* O_X = 0$ for all $i > 0$. Combining this with the projection formula, we have $R^i\varphi_* \varphi^*(mM) = 0$ for all $i > 0$. Moreover, since $\varphi_* O_X = O_Y$, the projection formula also tells us that $\varphi_* \varphi^*(mM) = mM$. Therefore

$$
(1.1.2) \quad h^0(m(K_X + \det E)) = h^0(\varphi^*(mM)) = h^0(mM) \quad \text{for every} \quad m \geq 1.
$$

Furthermore, by the Kodaira vanishing theorem,

$$
(1.1.3) \quad 0 = h^1(m(K_X + \det E)) = h^1(mM) \quad \text{for each} \quad m \geq 1,
$$

since $m(K_X + \det E) = K_X + ((m-1)(K_X + \det E) + \det E)$ and $K_X + \det E$ is nef. In particular, from (1.1.3) we get

$$
(1.1.4) \quad h^0(mM) = m \deg M + 1 - g(Y) \quad \text{for every} \quad m \geq 1.
$$

Now we consider the exact sequence

$$
0 \to m(K_X + \det E) \otimes I_Z \to m(K_X + \det E) \to mK_Z \to 0,
$$

where $I_Z$ is the ideal sheaf of $Z$. We know that $\varphi|_Z$ is surjective. In other words, $Z$ is not contained in a finite union of fibres of $\varphi$. This implies that $h^0(m(K_X + \det E) \otimes I_Z) = 0$ for every $m \geq 1$, because $m(K_X + \det E) = \varphi^*(mM)$ and every global section of $m(K_X + \det E)$ is the pullback of a global section of $mM$ by (1.1.2). Thus we have from the exact sequence above

$$
(1.1.5) \quad h^0(m(K_X + \det E)) \leq h^0(mK_Z) \quad \text{for all} \quad m \geq 1.
$$
In particular, combining (1.1.2) with (1.1.5) for \( m = 1 \), we get
\[(1.1.6) \quad h^0(M) \leq h^0(K_Z) + h^2(O_Z).\]

Hence, by recalling (1.1.1),
\[(1.1.7) \quad h^0(M) \leq h^2(O_Z) - h^1(O_Z) + h^0(O_Z) + h^1(O_Z) - h^0(O_Z) = \chi(O_Z) + g(Y) - 1.\]

Furthermore, from (1.1.7) and (1.1.4) for \( m = 1 \),
\[(1.1.8) \quad \deg M \leq \chi(O_Z) + 2(g(Y) - 1).\]

On the other hand, as mentioned previously, \( K_Z = \varphi^*_Z M \). So, by (1.1.8) and the canonical bundle formula for elliptic fibrations, we conclude that \( \varphi|_Z \) has no multiple fibres. This completes the proof. \( \square \)

**Remarks 1.2.** (i) As we have shown in the proof of the Theorem, \( \varphi|_Z \) has no multiple fibres. Hence the equality holds in (1.1.8). This directly implies that the equalities also hold in (1.1.7) and (1.1.6). In particular, it follows from (1.1.6) that
\[(1.2.1) \quad h^0(M) = p_g(Z) \quad \text{if} \quad K_X + \det \mathcal{E} \quad \text{is nef.}\]

(ii) The morphism \( \varphi : X \rightarrow Y \) considered above comes from the Stein factorization of the morphism associated with \( |m(K_X + \det \mathcal{E})| \) for \( m \) large enough, when \( K_X + \det \mathcal{E} \) is nef. Since \( K_Z^2 = 0 \), we see that \( K_X + \det \mathcal{E} \) is not ample. At this point we could refer to a result of Andreatta and Mella listing the possible structures of \((X, \mathcal{E})\), depending on whether \( K_X + \det \mathcal{E} \) is big or not \([\text{AM, Theorem 5.1, parts 2) and 3})\) respectively\]. Checking which of these structures is compatible with condition (*) and our assumption that \( Z \) is a minimal surface with \( \kappa(Z) = 1 \) leads to subcase (ii) in \([\text{AM, Theorem 5.1, part 2})\]. This gives an alternate proof to our first claim that \( \dim Y = 1 \), with the general fibre \( F \) of \( \varphi \) satisfying the condition \( K_F + \det \mathcal{E}_F = \mathcal{O}_F \).

**Corollary 1.3.** Let \( X' \) and \( \mathcal{E}' \) be as in case (b) of the Theorem. Then
\[h^0(m(K_{X'} + \det \mathcal{E}')) = P_m(Z) = m(\chi(O_Z) + 2g(Y) - 2) + 1 - g(Y)\]
for all \( m \geq 1 \).

**Proof.** Since \( P_m(Z) \) and \( \chi(O_Z) \) are birational invariants, it is enough to prove the formula for \( Z' \) instead of \( Z \). Since \( K_{X'} + \det \mathcal{E}' \) is nef, combining (1.1.2) with (1.1.4) gives
\[h^0(m(K_{X'} + \det \mathcal{E}')) = m \deg M + 1 - g(Y)\]
for \( m \geq 1 \). Moreover, by Remark 1.2 (i) we know that the equality holds in (1.1.8). In other words, \( \deg M = \chi(O_{Z'}) + 2g(Y) - 2 \). Therefore
\[h^0(m(K_{X'} + \det \mathcal{E}')) = m(\chi(O_{Z'}) + 2g(Y) - 2) + 1 - g(Y)\]
for all \( m \geq 1 \). Next we consider \( P_m(Z') \). As we showed in 1.1, \( Z' \) is a minimal surface with \( \kappa(Z') = 1 \), the elliptic fibration \( \varphi|_{Z'} : Z' \rightarrow Y \) has
no multiple fibres, and $K_{Z'} = \varphi_*^* M$ for some ample line bundle $M$ on $Y$. Hence, by Lemma 0.5

$$P_m(Z') = m(\chi(\mathcal{O}_{Z'}) + 2g(Y) - 2) + 1 - g(Y)$$

for every $m \geq 2$. Now assume that $M = K_Y$. Then $h^1(M) = 1$. However, this contradicts (1.1.3). Thus (1.3.1) is also valid for $m = 1$ by Lemma 0.5 again. □

**Corollary 1.4.** Let $X, E$ and $Z$ be as in (*). Then $Z$ cannot be the product of an elliptic curve $E$ and a smooth curve $Y$ of genus $g(Y) \geq 1$.

*Proof.* By contradiction, assume that $Z = E \times Y$. If $g(Y) = 1$, then $Z$ is an abelian surface, which cannot occur in view of [L1, Corollary 1.5]. If $g(Y) \geq 2$, then $K_Z = p^* K_Y$, where $p : Z \to Y$ denotes the projection, hence $Z$ is a minimal surface with $\kappa(Z) = 1$ and our Theorem applies. Case (a) cannot hold, since $Z$ is minimal. On the other hand, in case (b) we have $h^1(\mathcal{O}_Z) = g(Y)$ by (1.1.1). But this gives a contradiction, since $h^1(\mathcal{O}_{E \times Y}) = g(E) + g(Y) = 1 + g(Y)$. □

This generalizes a well-known fact in the setting of ample divisors [S1, Proposition IV]. We would like to point out that by the same argument as in the proof of Corollary 1.4 we can also exclude the case where $Z$ is an elliptic fibre bundle over a smooth curve of positive genus.

## 2. Kodaira dimensions.

Let things be as in the Introduction. In [L1] we conjectured that if $Z$ has Kodaira dimension $\kappa(Z) < 2$ then $\kappa(X) = -\infty$. The fact that $\kappa(X) = -\infty$ was proved when $\kappa(Z) = -\infty$ in [LM1, Corollary 5.2] and when $\kappa(Z) = 0$ in [L1, Corollary 4.3]. At present our conjecture follows from a more general result of Peternell, Schneider and Sommese on the Kodaira dimension of subvarieties, obtained in connection with the theory of normal pairs $(X,A)$ with $\mathbb{Q}$-effective normal sheaf $\mathcal{N}_{A/X}$ [PSS, Theorem 4.13]. Here we apply our Theorem to give an elementary proof of the conjecture above.

**Corollary 2.1.** Let $X, E$ and $Z$ be as in (*) and assume that $\kappa(Z) < 2$. Then $\kappa(X) = -\infty$.

*Proof.* By what we mentioned before we can assume that $\kappa(Z) = 1$. Then $(X,E)$ is as in the Theorem. In case (a), $X$ is a $\mathbb{P}^{n-2}$ bundle over a smooth surface $S$. Hence by the easy addition theorem we get $\kappa(X) \leq \kappa(\mathbb{P}^{n-2}) + \dim S = -\infty$. In case (b), up to the birational morphism $f$, which does not affect the Kodaira dimension, we can suppose that $X$ fibres over a smooth curve $Y$, with the general fibre being a smooth projective variety $F$ such that $K_F + \det E_F = \mathcal{O}_F$. In particular $F$ is Fano and so $\kappa(F) = -\infty$. Then the easy addition theorem again gives $\kappa(X) \leq \kappa(F) + \dim Y = -\infty$. □
Another consequence of Section 1 is the following:

**Corollary 2.2.** Let $X$, $E$, $Z$ be as in case (b) of the Theorem. Then $p_g(Z) > 0$.

**Proof.** In view of the birational invariance of $p_g$ we can assume that $Z = Z'$. Thus, we know by 1.1 that the elliptic fibration of $Z$ is given by $\varphi_{|Z} : Z \to Y$ with $K_Z = \varphi_{|Z}^*M$ for some ample $M \in \text{Pic}(Y)$. In particular, $\deg M > 0$. Moreover, by (1.2.1) we have $p_g(Z) = h^0(M)$. Since $\kappa(Z) = 1$, we know by the Castelnuovo–de Franchis theorem that $\chi(O_Z) \geq 0$. On the other hand, by (1.1.1) we have $g(Y) = h^{1,0}(Z)$. So, if $p_g(Z) = 0$, combining this equality with the inequality above implies $1 - g(Y) \geq 0$. But then, by (1.1.4) for $m = 1$, we get

$$p_g(Z) = h^0(M) = \deg M + 1 - g(Y) \geq 1,$$

a contradiction. $\square$

Note that the same property expressed by Corollary 2.2 still holds if $Z$ is a surface of Kodaira dimension zero. Actually, in this case, if $X$, $E$ and $Z$ are not as in case (a) (case (1) in [L1, Theorem]), after contracting a finite number of $(-1)$-hyperplanes we get a new triplet $(X', E', Z')$ satisfying (*), with $K_{X'} + \det E' \text{ nef}$ and $Z'$ a minimal surface with $\kappa(Z') = 0$. In particular, $12K_{Z'} = O_{Z'}$. The injectivity of the restriction homomorphism $\text{Pic}(X') \to \text{Pic}(Z')$ [LM2, (1.1.6)] combined with the fact that $(K_{X'} + \det E')_{Z'} = K_{Z'}$ implies that $12(K_{X'} + \det E') = O_{X'}$. Hence $X'$ is Fano, but since Fano manifolds have no torsion in their Picard groups, we conclude that $K_{X'} + \det E' = O_{X'}$. Hence, by adjunction, $K_{Z'} = O_{Z'}$, which gives $p_g(Z) = p_g(Z'') = 1$.

In [L2, (4.2)] the first author formulated the following conjecture: Let $X$, $E$ and $Z$ be as in (*). If $p_g(Z) = 0$ and $\kappa(Z) \geq 0$, then $(X, E)$ is as in case (a) of the Theorem. Corollary 2.2 and the discussion above show that this conjecture is true when $\kappa(Z) = 0$ or 1.

**References**


Received October 21, 1999.

DIPARTIMENTO DI MATEMATICA “F. ENRIQUES”
UNIVERSITÀ DEGLI STUDI DI MILANO
VIA C. SALDINI, 50
I-20133 MILANO
ITALY
E-mail address: lanteri@mat.unimi.it

DEPARTMENT OF MATHEMATICAL SCIENCES
SCHOOL OF SCIENCE AND ENGINEERING
WASEDA UNIVERSITY
3-4-1 OHKUBO, SHINJUKU
TOKYO 169-8555
JAPAN
E-mail address: hmaeda@mse.waseda.ac.jp
QUASIHOMEOMORPHISMS AND UNIVALENT HARMONIC MAPPINGS ONTO PUNCTURED BOUNDED CONVEX DOMAINS

Abdallah Lyzzaik

Dedicated to the memory of Professor Ahmad Shamsuddin

This paper deals with univalent harmonic mappings of annuli onto punctured bounded convex domains. Several aspects of these mappings are investigated; in particular, boundary functions, existence and uniqueness questions, and the geometry of their analytic and (co-analytic) parts. The paper also considers univalence criteria for Dirichlet solutions in annuli of boundary functions that are a generalized type of homeomorphisms, called quasihomeomorphisms, on one boundary component and constants on the other.

1. Introduction.

A harmonic mapping $f$ of a region $D$ is a complex-valued function of the form $f = h + \overline{g}$, where $h$ and $g$ are analytic functions in $D$, unique up to an additive constant, that are single-valued if $D$ is simply connected and possibly multiple-valued otherwise. We call $h$ and $g$ the analytic and co-analytic parts of $f$, respectively. If $f$ is (locally) injective, then $f$ is called (locally) univalent. Note that every conformal and anti-conformal function is a univalent harmonic mapping. The Jacobian and second complex dilatation of $f$ are given by the functions $J(z) = |h'(z)|^2 - |g'(z)|^2$ and $\omega(z) = g'(z)/h'(z)$, $z \in D$, respectively. Note that $\omega$ is either a nonconstant meromorphic function or a (possibly infinite) constant. A result of Lewy [13] states that if $f$ is a locally univalent mapping, then its Jacobian $J$ is never zero; namely, for $z \in D$, either $J(z) > 0$ or $J(z) < 0$. In the first case $|\omega(z)| < 1$ and $f$ is sense-preserving, and in the second $|\omega(z)| > 1$ and $f$ is sense-reversing.

A ring domain is a doubly-connected open subset of the plane. Denote by $\mathbb{A}(\rho, 1)$ the annulus $\{z : \rho < |z| < 1\}$, $0 \leq \rho < 1$. It seems that Nitsche [16] was the first to consider univalent harmonic mappings of $\mathbb{A}(\rho, 1)$ onto $\mathbb{A}(R, 1)$. Indeed, Nitsche observed that, unlike with conformal mappings, $R$ can possibly be zero as with the harmonic mapping

$$f(z) = (z - \rho^2/\overline{z})/(1 - \rho^2)$$

159
which can be easily shown to map $A(\rho,1)$ univalently onto the punctured disc $A(0,1)$. Later, Nitsche \[17\] §879 posed the following question.

**Question (Nitsche).** All univalent harmonic mappings from $A(\rho,1)$ onto $A(0,1)$, up to a rotation, are of form (1.1).

A negative answer to this question was given by Hengartner and Schober \[9\]. In their paper, the authors also investigated existence and uniqueness theorems for univalent harmonic mappings with given dilatations between annuli. Subsequently, Hengartner and Szynal \[10\] and Bshouty and Hengartner \[1\] gave a representation for harmonic mappings $f$ defined on an annulus $A(\rho,1)$ and constant on the inner circle as follows.

**Theorem A.** Let $f$ be a harmonic function of $A(\rho,1)$, $0 < \rho < 1$, that extends continuously across $|z| = \rho$ with $f$ identically $\zeta$ there. Then there exists a constant $c$ and a function $h$ analytic in $A(\rho^2,1)$ such that

$$f(z) = h(z) - h(\rho^2/z) + \zeta + 2c \log(|z|/\rho).$$

Further, if $f$ extends continuously across $|z| = 1$, and $f^*$ is the restriction of $f$ on $|z| = 1$, then $c = 0$ if and only if $\zeta$ equals

$$\zeta_0 = \frac{1}{2\pi} \int_0^{2\pi} f^*(e^{it}) \, dt.$$  

Using Theorem A, Bshouty and Hengartner \[1\] proved the following result.

**Theorem B.** Suppose that the following are true:

(i) $G$ is a bounded convex domain.

(ii) $f^*$ is a sense-preserving homeomorphism between the unit circle and $\partial G$, and the constant $\zeta_0 \in G$ given by Equation (1.3) on $|z| = \rho$.  

(iii) $f$ is the Dirichlet solution of $f^*$ in $A(\rho,1)$.

Then $f : A(\rho,1) \to G \setminus \{\zeta_0\}$ is a homeomorphism.

The author \[14, Theorem 2\] observed that Theorem B remains true under the weaker condition $f(A(\rho,1)) \subset G$ rather than the convexity of $G$.

In this paper, we investigate univalent harmonic mappings of ring domains onto bounded punctured convex domains. Throughout the paper we shall use the following notation: $\mathbb{C}$ for the complex plane, $\hat{\mathbb{C}}$ for the extended complex plane, $\mathbb{D}$ for the open unit disc $\{z \in \mathbb{C} : |z| < 1\}$, $T$ for the unit circle $\{z \in \mathbb{C} : |z| = 1\}$, $0 < \rho < 1$, $T_\rho$ for the circle $\{z \in \mathbb{C} : |z| = \rho\}$, $A(\rho,1)$ for the annulus $\{z \in \mathbb{C} : \rho < |z| < 1\}$, $G$ for a bounded convex domain. Also, for a subset $S \subset \mathbb{C}$, we denote by $\partial S$ and $\overline{S}$ the boundary and closure of $S$ in $\mathbb{C}$, respectively.

The paper is organized as follows. In Section 2, we describe the boundary functions, called quasihomoeomorphisms, of univalent harmonic mappings
onto punctured convex domains, and extend Theorem B to sense-preserving quasihomeomorphisms. Section 3 is devoted to investigate the geometry of the analytic parts of univalent harmonic mappings in $A(\rho, 1)$ onto punctured convex domains. One result of this section asserts that these (analytic parts) have nonvanishing derivatives on $T_\rho$, and that they map $T_\rho$ univalently onto Jordan convex curves. Another concludes that these can be written as univalent close-to-convex functions of homeomorphisms in $A(\rho, 1)$. In Section 4, we study univalence criteria for Dirichlet solutions in $A(\rho, 1)$ of boundary functions that are sense-preserving quasihomeomorphisms between $T$ and $\partial G$ and constants on $T_\rho$—a study which was motivated by Hengartner [2, Problem 15]. In Section 5, we prove a uniqueness result implying that the function $f$ defined by (1.1) is the only univalent harmonic mapping, up to rotation, of $A(\rho, 1)$ onto $A(0, 1)$ having zero as an average value on $T$ and with analytic part that extends analytically throughout $D$. This somehow corrects Nitsche’s question above and sheds light on Nitsche’s insight in that direction.

2. Quasihomeomorphisms and Univalent Harmonic Mappings.

The purpose of this section is to characterize the boundary functions of univalent harmonic mappings, and to extend Theorem B to “quasihomeomorphisms”. For this purpose, we need the following definition.

Definition 2.1. Let $f$ be a function of $T$ into a Jordan curve $C$ of $\mathbb{C}$. We say $f$ is a sense-preserving quasihomeomorphism of $T$ into $C$ if it is a pointwise limit of a sequence of sense-preserving homeomorphisms of $T$ onto $C$. If in addition, $f$ is a continuous function onto $C$, then $f$ is called a sense-preserving weak homeomorphism.

The definition is based on Bshouty, Hengartner and Naghibi-Beidokhti [3, Definitions 3.1, 3.2]. Sense-preserving quasihomeomorphisms and sense-preserving weak homeomorphisms are characterized as follows.

Proposition 2.1. Let $f$ be a function of $T$ into a Jordan curve $C$, and let $F$ be a sense-preserving homeomorphism of $T$ onto $C$.

(i) If $f$ is a sense-preserving quasihomeomorphism of $T$ onto $C$, then there is a real-valued nondecreasing function $\varphi$ on $\mathbb{R}$ such that $\varphi(t + 2\pi) = \varphi(t) + 2\pi$ and $f(e^{it}) = F(e^{i\varphi(t)})$.

(ii) If $f(e^{it}) = F(e^{i\varphi(t)})$, where $\varphi$ is a real-valued nondecreasing function on $\mathbb{R}$ such that $\varphi(t + 2\pi) = \varphi(t) + 2\pi$, and if $E$ is the countable set of points $e^{i\varphi(t)}$ where $\varphi$ is discontinuous, then $f$ coincides on $T \setminus E$ with a sense-preserving quasihomeomorphism of $T$. In this case, $f$ is the pointwise limit in $T \setminus E$ of a sequence of sense-preserving homeomorphisms $f_n(e^{it}) = F(e^{i\varphi_n(t)})$ of $T$ onto $C$, where each $\varphi_n$ is a
real-valued infinite differentiable function on \( \mathbb{R} \) such that \( \varphi_n(t + 2\pi) = \varphi_n(t) + 2\pi \) and \( \varphi'_n(t) \) is always positive.

(iii) \( f \) is a sense-preserving weak homeomorphism of \( \mathbb{T} \) onto \( C \) if and only if there is a real-valued continuous nondecreasing function \( \varphi \) on \( \mathbb{R} \) such that \( \varphi(t + 2\pi) = \varphi(t) + 2\pi \) and \( f(e^{it}) = F(e^{i\varphi(t)}) \). In this case, \( f \) is the uniform limit of a sequence of sense-preserving homeomorphisms \( f_n(e^{it}) = F(e^{i\varphi_n(t)}) \) of \( \mathbb{T} \) onto \( C \), where each \( \{\varphi_n\} \) is a real-valued infinite differentiable function on \( \mathbb{R} \) such that \( \varphi_n(t + 2\pi) = \varphi_n(t) + 2\pi \) and \( \varphi'_n(t) \) is always positive.

Proof. (i) There is a sequence \( \{f_n\} \) of sense-preserving homeomorphisms of \( \mathbb{T} \) onto \( C \) that converges pointwise to \( f \). We can write \( f_n(e^{it}) = F(e^{i\varphi_n(t)}) \), where each \( \{\varphi_n\} \) is a real-valued increasing function on \( \mathbb{R} \) such that \( 0 < \varphi_n(0) < 2\pi \) and \( \varphi_n(t + 2\pi) = \varphi_n(t) + 2\pi \). Then, by Helly’s selection theorem, there is a real-valued nondecreasing function \( \varphi \) on \( \mathbb{R} \) such that \( \varphi(t + 2\pi) = \varphi(t) + 2\pi \) and \( \varphi_n \to \varphi \) pointwise in \( \mathbb{R} \). Therefore, \( f(e^{it}) = F(e^{i\varphi(t)}) \) and (i) follows.

(ii) The function \( \varphi(t) - t \) is bounded, a.e. differentiable, and has period \( 2\pi \). For fixed \( n = 1, 2, \ldots \), define the function

\[
\varphi_n(t) = t + \frac{1}{2\pi} \int_0^{2\pi} P(1 - 1/n, t - \theta)[\varphi(\theta) - t] d\theta,
\]

where \( P(r, \theta) \) is the Poisson kernel. Then \( \varphi_n \) is an infinite differentiable function such that \( \varphi_n(t + 2\pi) = \varphi_n(t) + 2\pi \). Also,

\[
\varphi'_n(t) = \frac{1}{2\pi} \int_0^{2\pi} P(1 - 1/n, t - \theta) d\varphi(\theta) > \frac{1}{2n - 1}
\]

since \( P(1 - 1/n, t - \theta) < 1/(2n - 1) \), \( -\infty \leq t, \theta \leq \infty \). Denote by \( E \) the set of points of \( \mathbb{T} \) where \( e^{i\varphi(t)} \) is discontinuous; then \( E \) is countable since \( \varphi \) is a nondecreasing function. But by a Schwarz’s theorem, \( \varphi_n \to \varphi \) pointwise in the set of continuity of \( \varphi \). Therefore, \( f_n \to f \) pointwise in \( \mathbb{T} \setminus E \).

(iii) If \( f \) is a sense-preserving weak homeomorphism of \( \mathbb{T} \) onto \( C \), then, by (i), \( f(e^{it}) = F(e^{i\varphi(t)}) \) where \( \varphi(t) \) is a real-valued nondecreasing function on \( \mathbb{R} \) such that \( \varphi(t + 2\pi) = \varphi(t) + 2\pi \). Since \( F : \mathbb{T} \to C \) is a homeomorphism and \( f \) is continuous, \( e^{i\varphi(t)} = F^{-1} \circ f(e^{it}) \) is also continuous in \( \mathbb{R} \). This, together with the nonconstancy of \( f \), implies that \( \varphi \) is also continuous in \( \mathbb{R} \).

Suppose now that \( f(e^{it}) = F(e^{i\varphi(t)}) \) where \( \varphi \) is a real-valued continuous nondecreasing function on \( \mathbb{R} \) such that \( \varphi(t + 2\pi) = \varphi(t) + 2\pi \). Define the functions \( \varphi_n \) as in the proof of (ii), and recall that \( \varphi_n(t + 2\pi) = \varphi_n(t) + 2\pi \) and that \( \varphi'_n \) is always positive. Observe that, since \( \varphi \) is continuous, \( \varphi_n \to \varphi \) uniformly in \( \mathbb{R} \). Hence, with \( f_n(e^{it}) = F(e^{i\varphi_n(t)}) \), each \( f_n \) is a sense-preserving homeomorphism of \( \mathbb{T} \) onto \( C \) and \( f_n \to f \) uniformly on \( \mathbb{T} \). This concludes (iii). \( \square \)
Let $f$ be a function of $\mathcal{A}(\rho, 1)$ into $\hat{C}$, and let $\xi \in \mathbb{T}$. We say that $f$ has the unrestricted limit $a \in \hat{C}$ at if
\[
f(z) \to a \quad z \to \xi, \quad z \in \mathcal{A}(\rho, 1);
\]
by defining $f(\xi) = a$ the function $f$ becomes continuous at $\xi$ as a function in $\mathcal{A}(\rho, 1) \cup \{\xi\}$. We shall use $f(\xi)$ to denote the unrestricted limit whenever it exists, and call the resulting function, on its domain of definition in $\mathbb{T}$, the unrestricted limit function $f$. We also define the cluster set $C(f, \xi)$ of $f$ at $\xi$ as the set of all $b \in \hat{C}$ for which there are sequences $\{z_n\}$ such that
\[
z_n \in \mathcal{A}(\rho, 1), \quad z_n \to \xi, \quad f(z_n) \to b \quad \text{as } n \to \infty.
\]
Moreover, If $F$ is a subset of $\mathbb{T}$, then we define the cluster set $C(f, F)$ of $f$ at $F$ as the set-union of the cluster sets $C(f, \xi)$ for $\xi \in E$.

Sense-preserving quasihomomorphisms are essential for describing the boundary behaviour of univalent harmonic mappings of ring domains onto bounded convex domains. Suppose $f$ is a univalent harmonic mapping of $\mathcal{A}(\rho, 1)$ onto a ring domain $G \setminus \{\zeta\}$, $\zeta \in \mathbb{G}$. Then either $\lim_{|z| \to 1} f(z) = \zeta$ and $C(f, \mathbb{T}_\rho) = \partial G$, or $\lim_{|z| \to \rho} f(z) = \zeta$ and $C(f, \mathbb{T}) = \partial G$. In the first case, $f(1/z)$ becomes a univalent harmonic mapping of $\mathcal{A}(\rho, 1)$ onto $G \setminus \{\zeta\}$ with $\lim_{|z| \to \rho} f(1/z) = \zeta$ and $C(f(1/z), \mathbb{T}) = \partial G$. For our study, this leads us to consider, without loss of generality, only univalent harmonic mappings of $\mathcal{A}(\rho, 1)$ onto ring domains $G \setminus \{\zeta\}$, $\zeta \in \mathbb{G}$, with $\lim_{|z| \to \rho} f(z) = \zeta$.

**Definition 2.2.** Denote by $\mathcal{H}_u(\rho, \mathbb{G})$ the class of all univalent harmonic mappings $f$ of $\mathcal{A}(\rho, 1)$ onto ring domains $G \setminus \{\zeta\}$, $\zeta \in \mathbb{G}$, with $f(\mathbb{T}_\rho) = \zeta$.

The boundary behavior of functions $f \in \mathcal{H}_u(\rho, \mathbb{G})$ is given as follows.

**Theorem 2.1.** Let $f \in \mathcal{H}_u(\rho, \mathbb{G})$. Then there is a countable set $E \subset \mathbb{T}$ such that the following hold:

(i) For each $e^{i\theta} \in \mathbb{T} \setminus E$, the unrestricted limit $f(e^{i\theta})$ exists and belongs to $\partial \mathbb{G}$. Furthermore, $f$ is continuous in $\mathcal{A}(\rho, 1) \setminus E$.

(ii) For each $e^{i\theta_0} \in E$, the side-limits $\lim_{\theta \to \theta_0} f(e^{i\theta})$ and $\lim_{\theta \to \theta_0} f(e^{i\theta})$ exist in $\partial \mathbb{G}$ and are distinct.

(iii) For each $e^{i\theta_0} \in E$, the cluster set $C(f, e^{i\theta_0})$ lies in $\partial \mathbb{G}$ and is the straight-line segment joining the side-limits $\lim_{\theta \to \theta_0} f(e^{i\theta})$ and $\lim_{\theta \to \theta_0} f(e^{i\theta})$.

(iv) $\overline{\mathbb{T}}(f(\mathbb{T} \setminus E)) = \mathbb{G}$.

(v) There is a sense-preserving quasihomomorphism of $\mathbb{T}$ into $\partial \mathbb{G}$ that coincides with the unrestricted limit function $f$ on $\mathbb{T} \setminus E$.

(vi) $f$ is the Dirichlet solution in $\mathcal{A}(\rho, 1)$ of the function $f^* \in \mathcal{A}(\rho, 1)$ defined by the unrestricted limit function of $f$ on $\mathbb{T}$ and the value of $f$ on $\mathbb{T}_\rho$. 
The fact that \( f^* \) is not defined on \( E \) in (vi) is insignificant. Indeed, Dirichlet solutions in multiply connected domains coincide whenever their boundary functions coincide almost everywhere.

**Proof.** Applying [8, Theorem 4.3] to \( f \) locally at each \( e^{i\theta} \) yields (i), (ii), and (iii) except for the inclusion \( C(f, e^{i\theta}) \subset \partial G \) which follows because \( f : \mathbb{A}(\rho, 1) \rightarrow \partial G \) is onto. Also, (vi) follows from the maximum principle.

(iv) Since \( \mathfrak{T} \) is convex and each unrestricted limit \( f(e^{i\theta}) \) belongs to \( \partial G \), \( \varTheta(f(\mathbb{T} \setminus E)) \subset \mathfrak{T} \). Let \( w \in \partial G \). Because \( f : \mathbb{A}(\rho, 1) \rightarrow \mathbb{G} \) is onto, \( w \) belongs to the cluster set of some point \( \xi \in \mathbb{T} \). If \( \xi \in E \), then \( w \) is the unrestricted limit \( f(\xi) \). If \( \xi \notin E \), then \( w \) belongs to the boundary straight-line segment joining the side-limits at \( \xi \) of the unrestricted function \( f \). Note that these limits belong to \( \varTheta(f(\mathbb{T} \setminus E)) \); consequently \( w \in \varTheta(f(\mathbb{T} \setminus E)) \). In either case \( \mathfrak{T} \subset \varTheta(f(\mathbb{T} \setminus E)) \), and (iv) follows.

(v) Since \( \mathbb{G} \) is a Jordan domain, there is a homeomorphism \( F \) of \( \mathbb{D} \) onto \( \mathfrak{T} \) which maps \( \mathbb{D} \) conformally onto \( \mathbb{G} \). Let \( G(z) = F^{-1} \circ f(z), z \in \mathbb{A}(\rho, 1) \). Observe that \( G \) is sense-preserving homeomorphism of \( \mathbb{A}(\rho, 1) \) into \( \mathbb{D} \) which extends continuously to a mapping, also denoted by \( G \), from \( \mathbb{A}(\rho, 1) \cup (\mathbb{T} \setminus E) \) to \( \mathbb{T} \). Let \( I = \{ t : -\infty < t < \infty, e^{it} \in \mathbb{T} \setminus E \} \). We conclude that there is a continuous nondecreasing function \( \varphi \) on \( I \) such that \( G(e^{it}) = e^{i\varphi(t)}, t \in I \), and

\[
\sup\{\varphi(t) : t \in I \cap [0, 2\pi)\} - \inf\{\varphi(t) : t \in I \cap [0, 2\pi)\} \leq 2\pi.
\]

Extend \( \varphi \) to \( (-\infty, \infty) \) by defining \( \varphi(\tau) = \inf\{\varphi(t) : t \in I \cap [0, 2\pi), t > \tau\} \) if \( \tau \in [0, 2\pi) \setminus I \), and by letting \( \varphi(t + 2\pi) = \varphi(t) + 2\pi \). It is immediate that the new \( \varphi \) is a nondecreasing function on \( (-\infty, \infty) \) with period \( 2\pi \) that is continuous only on \( I \). Using Proposition 2.1(ii), the function \( F(e^{i\varphi(t)}) \) coincides with a sense-preserving quasihomoeomorphism of \( \mathbb{T} \) into \( \partial G \) on \( \mathbb{T} \setminus E \). \( \square \)

Now let \( f^* \) be a sense-preserving quasihomoeomorphism of \( \mathbb{T} \) into \( \partial G \). Throughout the paper we denote by \( \mathcal{E}(f^*) \) the set of points \( e^{i\theta} \) at which \( f^* \) is continuous. Our second purpose in this section is to show that if \( \mathfrak{T} \) is the closed convex hull of \( f^*(\mathcal{E}(f^*)) \), then \( f^* \) yields a univalent harmonic mapping of \( \mathbb{A}(\rho, 1) \) onto the convex domain \( \mathbb{G} \) minus one point. This extends Theorem B to sense-preserving quasihomoeomorphisms \( f^* \) of \( \mathbb{T} \) into \( \partial G \).

**Theorem 2.2.** Suppose that the following are true:

(i) \( f^* \) is a sense-preserving quasihomoeomorphism of \( \mathbb{T} \) into \( \partial G \), and the constant \( \zeta_0 \) defined by (1.3) on \( \mathbb{T} \).

(ii) \( \varTheta f^*(\mathcal{E}(f^*)) = \mathfrak{T} \).

(iii) \( f \) is the Dirichlet solution of \( f^* \) in \( \mathbb{A}(\rho, 1) \).
Then \( \zeta_0 \in \mathcal{G} \), \( f \in \mathcal{H}_w(\rho, \mathcal{G}) \), and there is an analytic function \( h \) of \( \mathbb{A}(\rho^2, 1) \) such that
\[
(2.2) \quad f(z) = h(z) - h(\rho^2/\overline{z}) + \zeta_0, \quad (z \in \mathbb{A}(\rho, 1)).
\]

The proof of the theorem requires two lemmas. Let \( f^* \) be a sense-preserving quasihomoeomorphism of \( T \) into \( \partial \mathbb{G} \), and let \( F \) be a homeomorphism of \( \mathbb{D} \) onto \( \overline{\mathbb{C}} \) that maps \( \mathbb{D} \) conformally onto \( \mathbb{G} \). By Proposition 2.1(i), there is a real-valued nondecreasing function \( \varphi \) on \( \mathbb{R} \) such that \( \varphi(\theta + 2\pi) = \varphi(\theta) + 2\pi \) and \( f^*(e^{i\varphi}) = F(e^{i\varphi(\theta)}) \). If \( E \) is the set of points of discontinuity of \( e^{i\varphi(\theta)} \) in \( T \), then Proposition 2.1(ii) yields a sequence \( \{\varphi_n\} \) of real-valued infinite-differentiable functions on \( \mathbb{R} \) such that \( \varphi_n(\theta + 2\pi) = \varphi_n(\theta) + 2\pi \), \( \varphi_n'(\theta) > 0 \), and \( F(e^{i\varphi_n(\theta)}) \to F(e^{i\theta}) \) pointwise on \( T \setminus E \). Let \( \{r_n\}, \rho < r_n \leq 1 \), be a sequence converging to 1, and let \( f^*_n(e^{i\theta}) = F(r_n e^{i\varphi_n(\theta)}) \). Since \( F \) is uniformly continuous on \( \mathbb{D} \), we conclude that \( f^*_n(e^{i\theta}) \to f(e^{i\theta}) \) pointwise on \( T \setminus E \). Note that since \( F \) is a convex univalent function, if \( r_n < 1 \) then \( f^*_n \) is an infinite-differentiable sense-preserving homeomorphism of \( T \) onto a convex curve in \( \mathbb{G} \), and \( (f^*_n)'(e^{i\theta}) \) is nonvanishing. Define \( f^* \) and each \( f^*_n \) on \( T_\rho \) by the constants \( \zeta_0 \) and \( \zeta_n \) respectively, where \( \zeta_0 \) is given by (1.3) and
\[
(2.3) \quad \zeta_n = \frac{1}{2\pi} \int_0^{2\pi} f^*_n(e^{i\theta}) \, d\theta.
\]

By the bounded convergence theorem, \( \zeta_n \to \zeta_0 \). Now let \( f \) and \( f_n \) be the solutions of the Dirichlet problems of \( f^* \) and \( f^*_n \) in \( \mathbb{A}(\rho, 1) \) respectively. By Theorem A, we can represent \( f \) by (1.2) and write each \( f_n \) as
\[
(2.4) \quad f_n(z) = h_n(z) - h_n(\rho^2/\overline{z}) + \zeta_n
\]
where \( h_n \) is analytic in \( \mathbb{A}(\rho^2, 1) \). Moreover, Theorem B implies that each \( f_n : \mathbb{A}(\rho, 1) \to \mathbb{G} \setminus \{\zeta_0\} \) is a homeomorphism.

Under the above assumptions, we prove the requisite lemmas.

**Lemma 2.1.** \( f_n \to f \) locally uniformly in \( \mathbb{A}(\rho, 1) \).

**Proof.** Let \( \Phi \) be a local homeomorphism of \( \mathbb{D} \setminus \{\pm 1\} \) onto \( \overline{\mathbb{A}(\rho, 1)} \) that maps \( \mathbb{D} \) conformally onto \( \mathbb{A}(\rho, 1) \), the upper semi-circle: \( |\xi| = 1, \Re \xi > 0 \) onto \( T \), and the lower semi-circle: \( |\xi| = 1, \Re \xi < 0 \) onto \( T_\rho \). Put \( T_n^* = f^*_n \circ \Phi \), \( T^* = f^* \circ \Phi \), \( T_n = f_n \circ \Phi \), and \( T = f \circ \Phi \). Note that \( T_n \) and \( T \) are the Dirichlet solutions of \( T_n^* \) and \( T^* \) in \( \mathbb{D} \) respectively, and that \( T_n \to T^* \) pointwise a.e. in \( T \) since \( \zeta_n \to \zeta_0 \). Hence, for \( \eta = Re^{i\Theta} \), we can write
\[
T_n(\eta) = \frac{1}{2\pi} \int_0^{2\pi} P(R, \tau - \Theta)T_n^*(e^{i\tau}) \, d\tau
\]
and
\[
T(\eta) = \frac{1}{2\pi} \int_0^{2\pi} P(R, \tau - \Theta)T^*(e^{i\tau}) \, d\tau.
\]
Let $K \subset A(\rho,1)$ be a compact disc, and let $\tilde{K}$ be a connected component of $\Phi^{-1}(K)$. Then $\tilde{K}$ is also compact with a distance $\sigma > 0$ from $\mathbb{T}$. Then, for $\eta \in \tilde{K}$, 
\begin{equation*}
|T_n(\eta) - T(\eta)| \leq \frac{1}{(1 - \sigma)\pi} \int_0^{2\pi} |T_n^*(e^{i\tau}) - T(e^{i\tau})| d\tau,
\end{equation*}
and $T_n \to T$ uniformly on $\tilde{K}$ by the bounded convergence theorem. It follows at once that $f_n \to f$ uniformly on $K$.

**Remark 2.1.** The above proof uses only the pointwise convergence a.e. of $T_n^*$ to $T^*$ in $\mathbb{T}$ which follows at once from the the pointwise convergence of $f_n^*$ to $f^*$ in $\tilde{E}(f^*)$. We conclude that if $f_n^*$, $n = 1, 2, \ldots$, are sense-preserving quasihomoeomorphisms of $\mathbb{T}$ into $\partial G$ such that $f_n^* \to f^*$ pointwise a.e. in $\mathbb{T}$, then $f_n \to f$ locally uniformly in $\mathbb{A}(\rho,1)$ where $f$ and each $f_n$ are as defined above.

**Lemma 2.2.** (a) $h_n \to h$ locally uniformly in $\mathbb{A}(\rho^2,1)$.
(b) For $z \in A(\rho,1)$, 
\begin{equation}
 f(z) = h(z) - h(\rho^2/z) + \zeta_0 = \sum_{k \neq 0} c_k(f^*) \frac{r^{2k} - \rho^{2k}}{1 - \rho^{2k}} r^{-k} e^{ik\theta}
 \end{equation}
where $c_k(f^*)$, $k = \pm 1, \pm 2, \ldots$, is the $k$-th Fourier coefficient of $f^*$.

**Proof.** For $z = re^{i\theta}$, $\rho^2 < r < 1$, and $n = 1, 2, \ldots$, we have 
\begin{equation}
 h_n(z) = \sum_{k=-\infty}^{\infty} a_k(f_n^*) r^k e^{ik\theta}
 \end{equation}
which, with (2.4), yields 
\begin{equation}
 f_n(z) = \zeta_0 + \sum_{k \neq 0} a_k(f_n^*) \left( r^k - \frac{\rho^{2k}}{r^{2k}} \right) e^{ik\theta}.
 \end{equation}
The uniqueness of the Fourier series of $f_n(re^{i\theta})$ gives 
\begin{equation*}
 a_k(f_n^*) \left( r^k - \frac{\rho^{2k}}{r^{2k}} \right) = \frac{1}{2\pi} \int_0^{2\pi} f_n^*(re^{it})e^{-ikt} dt, \quad (k \neq 0).
 \end{equation*}
Letting $r \to 1$, the bounded convergence theorem yields 
\begin{equation*}
 a_k(f_n^*) (1 - \rho^{2k}) = \frac{1}{2\pi} \int_0^{2\pi} f_n^*(e^{it})e^{-ikt} dt = c_k(f_n^*), \quad (k \neq 0).
 \end{equation*}
Hence 
\begin{equation}
 a_k(f_n^*) = \frac{c_k(f_n^*)}{1 - \rho^{2k}}, \quad (k \neq 0).
 \end{equation}
Substituting this in (2.6) yields

\[ f_n(z) = \zeta_n + \sum_{k \neq 0} c_k (f^*_n) r^{2k} \frac{\rho^{2k} - r^{-k} e^{ik\theta}}{1 - \rho^{2k}}. \]

Proceeding likewise for \( h \), we conclude that if

\[ h(z) = \sum_{k = -\infty}^{\infty} a_k (f^*) r^k e^{ik\theta}, \]

then

\[ a_k (f^*) = c_k (f^*) \frac{1 - \rho^{2k}}{1 - \rho^{2k}}, \quad (k \neq 0). \] (2.8)

Now since \( f^*_n(e^{it}) \rightarrow f^*(e^{it}) \) pointwise in \( T \setminus E \), \( c_k (f^*_n) \rightarrow c_k (f^*) \) uniformly relative to \( k \) as \( n \rightarrow \infty \). It follows, by (2.8) and (2.7), that \( a_k (f^*_n) \rightarrow a_k (f^*) \) uniformly relative to \( k \) as \( n \rightarrow \infty \). This proves (a). Now since \( h_n(z) \rightarrow h(z) \) and \( f_n(z) \rightarrow f(z) \) uniformly in \( A(\rho, 1) \), and \( \zeta_n \rightarrow \zeta \), we conclude (2.5) by taking the limits of both sides in (2.4).

**Proof of Theorem 2.2.** First, we show that \( \zeta_0 \in G \). Obviously, \( \zeta_0 \in \mathbb{G} \).

Suppose that \( \zeta_0 \in \partial G \). Since \( G \) is convex, there is a real \( \theta_0 \) such that

\[ \Re \left\{ e^{i\theta_0} [f^*(e^{i\theta}) - \zeta_0] \right\} \geq 0, \quad (0 \leq \theta \leq 2\pi). \]

By virtue of (ii), we conclude that this inequality must be strict in some open interval \((\alpha, \beta)\), where \( 0 \leq \alpha < \beta \leq 2\pi \). This implies that

\[ \frac{1}{2\pi} \int_{0}^{2\pi} \Re \left\{ e^{i\theta_0} [f^*(e^{i\theta}) - \zeta_0] \right\} d\theta > 0, \]

and consequently

\[ \frac{1}{2\pi} \int_{0}^{2\pi} e^{i\theta_0} [f^*(e^{i\theta}) - \zeta_0] d\theta \neq 0. \]

This yields at once

\[ \zeta_0 \neq \frac{1}{2\pi} \int_{0}^{2\pi} f^*(e^{i\theta}) d\theta \]

which gives a contradiction. Hence, \( \zeta_0 \in \mathbb{G} \).

In view of Lemma 2.2(b), it remains to show that \( f : A(\rho, 1) \rightarrow \mathbb{G} \setminus \{\zeta_0\} \) is a homeomorphism.

We show that \( f \) is univalent. Let \( f^*_n \) and \( f_n \), \( n = 1, 2, \ldots \), be the functions defined in the first paragraph succeeding the statement of the theorem but with each \( r_n = 1 \). Using (2.4), the Jacobian of \( f_n \) can be written as

\[ J_n(z) = [zh_n'(z)]^2 - |\rho^2 h_n'(\rho^2/z)|^2/|z|^2 > 0, \quad z \in A(\rho, 1). \]

Since \( f_n \) is univalent and sense-preserving, Lewy's theorem [13] implies \( J_n(z) > 0 \); so \( h_n'(z) \neq 0 \) for \( z \in A(\rho, 1) \). But, by Lemma 2.2(a), \( h_n' \rightarrow h' \)
locally uniformly in $A(\rho, 1)$. This implies, by Hurwitz’s theorem, that either $h'$ is nonvanishing or is identically zero in $A(\rho, 1)$. If the latter case holds, then $f$ is constant. This yields, by [21, Theorem IV.3] and (ii), that $\partial G$ is a singleton which contradicts (ii). Hence $h'(z) \neq 0$ for $z \in A(\rho, 1)$. Now the Jacobian of $f$ is given by

$$J(z) = \frac{|zh'(z)|^2 - |\rho^2 h'(\rho^2/z)|^2}{|z|^2}, \quad z \in A(\rho, 1).$$

For $z \in A(\rho, 1)$, $J(z) \geq 0$ since $J_n(z) \to J(z)$. If $J(z) = 0$ for some $z$, then $J$ is identically zero; this follows by applying the maximum principle to the dilatation of $f$ given by

$$(2.9) \quad \omega(z) = \frac{(\rho^2/z)h'(/ho^2/z)}{zh'(z)}, \quad (z \in A(\rho, 1)).$$

This implies, by [14, Lemma 2], that $f$ maps $A(\rho, 1)$ into a straight-line $L$. It follows that the unrestricted limits of $f$ lie in $L$. By [21, Theorem IV.3] and (ii), $f^*(\tilde{E}(f^*))) \subset L$ which yields $\partial G \subset L$. This gives a contradiction. Hence $J(z) > 0$ for $z \in A(\rho, 1)$, and Lewy’s theorem [13] yields $f$ locally univalent. Now the univalence of $f_n$, together with Lemma 2.1, yields $f$ univalent in $A(\rho, 1)$.

Next, we show that $f : A(\rho, 1) \to G \setminus \{\zeta_0\}$ is onto. Let $\xi \in T$. The cluster set $C(f, \xi)$ of $f$ at $\xi$ is the singleton $f(\xi) \in \partial G$ if $f$ has an unrestricted limit at $\xi$, or the straight-line segment $\ell$ joining the points $\lim_{\theta \to \theta_0} f(e^{i\theta})$ and $\lim_{\theta \to \theta_0} f(e^{i\theta})$, where $\xi = e^{i\theta_0}$, which belong to $\partial G$ by (ii). Suppose that the latter case holds. If $\ell \not\subset \partial G$, then $\ell$ is a crosscut of $G$ which separates $G$ into two Jordan domains of which one contains $f(A(\rho, 1))$. If $L$ is the straight-line containing $\ell$, then the cluster set $C(f, \ell)$ of $f$ on $\ell$ lies completely in the closed half-plane bounded by $L$ and containing $f(A(\rho, 1))$. Consequently, $\overline{\sigma f^*(\tilde{E}(f^*)))}$ is a proper subset of $\ell$ which contradicts (ii). Hence, $\ell \subset \partial G$, and $C(f, \xi) \subset T$ for every $\xi \in T$. Now if $f : A(\rho, 1) \to G \setminus \{\zeta_0\}$ is not onto, then there is $\xi \in T$ such that $C(f, \xi) \cap G \neq \emptyset$ which yields a contradiction. Therefore, $f : A(\rho, 1) \to G \setminus \{\zeta_0\}$ is a homeomorphism. This completes the proof. □

**Remark 2.2.** The last paragraph of the above proof is indeed a proof for the following result: Let $f$ be the Dirichlet solution in $A(\rho, 1)$ of a boundary function $f^*$ defined on $T$ by a sense-preserving quasiconformal mapping into $\partial G$ with $\sigma f^*(\tilde{E}(f^*))) = \overline{\sigma}$, and on $T_\rho$ by a constant $\zeta \in G$. If $f$ is univalent, then $f : A(\rho, 1) \to G \setminus \{\zeta\}$ is a homeomorphism.

Theorems 2.1 and 2.2 provide an interesting relationship between sense-preserving quasiconformal mappings of $T$ into $\partial G$ and univalent harmonic mappings of $A(\rho, 1)$ onto once punctured $G$. View two sense-preserving quasiconformal mappings $f^*$ of $T$ into $\partial G$ equivalent if they coincide almost everywhere. Let $f^*$ and $k^*$ be sense-preserving quasiconformal mappings of $T$ into
\( \partial G \). Using Proposition 2.1(i), it is easily seen that \( f^* \) and \( k^* \) are equivalent if and only if \( \tilde{E}(f^*) = \tilde{E}(k^*) \) and \( f^* \) and \( k^* \) are identical on \( \tilde{E}(f^*) \). Denote by \( \mathcal{Q}(G) \) the class of all (equivalence classes of) sense-preserving quasihomomorphisms \( f^* \) of \( \mathbb{T} \) into \( \partial G \) satisfying (ii) of Theorem 2.2. It is immediate that if \( f^* \in \mathcal{Q}(G) \) and \( k^* \) is equivalent to \( f^* \), then \( k^* \in \mathcal{Q}(G) \).

**Definition 2.3.** Denote by \( \mathcal{H}_0(\rho, G) \) be the class of all Dirichlet solutions \( f \) satisfying (i), (ii) and (iii) of Theorem 2.2.

The classes \( \mathcal{Q}(G) \) and \( \mathcal{H}_0(\rho, G) \) are related as follows.

**Theorem 2.3.** Define \( T : \mathcal{Q}(G) \to \mathcal{H}_0(\rho, G) \) by \( T(f^*) = f \), where \( f \) is the Dirichlet solution in \( \mathbb{H}(\rho, 1) \) of the boundary function which is \( f^* \) on \( \mathbb{T} \) and the average of \( f^* \) on \( \mathbb{T}_\rho \). Then \( T \) is bijective. Furthermore, for a sequence \( \{f^*_n\} \) in \( \mathcal{Q}(G) \), the following statements are equivalent:

(a) \( f^*_n \to f^* \) a.e.
(b) \( f^*_n \to f^* \) in \( L^1 \).
(c) \( f_n \to f \) locally uniformly in \( \mathbb{H}(\rho, 1) \).

**Proof.** Suppose that for \( f^*_1, f^*_2 \in \mathcal{Q}(G) \), \( T(f^*_1) = f_1 \), \( T(f^*_2) = f_2 \), and \( f_1 = f_2 \). Then, by Theorem 2.1(i), \( f^*_1 = f^*_2 \) everywhere except possibly on a countable set. This makes \( T \) injective. Also, by Theorem 2.1, \( T \) is surjective. Hence \( T \) is bijective.

The implication (a) \( \Rightarrow \) (b) follows by the bounded convergence theorem. Conversely, by Proposition 2.1(i) and Helly’s selection theorem, there is a subsequence \( \{n_j\} \) of positive integers such that the sequence \( \{f^*_{n_j}\} \) converges pointwise to a bounded function \( k^* \). So, \( |f^*_{n_j} - f^*| \to |k^* - f^*| \) pointwise. Then, by (b) and the bounded convergence theorem, \( \int_0^{2\pi} |k^* - f^*| = 0 \). This gives (a), and we conclude (a) \( \iff \) (b). On the other hand, by Remark 2.1, (a) \( \Rightarrow \) (c). It remains to show (c) \( \Rightarrow \) (a). Using Theorem 2.1(i), there exist sense-preserving quasihomomorphisms \( f^* \) and \( f^*_n \) of \( \mathbb{T} \) into \( G \) that coincide with the boundary functions of \( f \) and \( f_n \) everywhere except possibly on countable sets, respectively. By Helly’s selection theorem, every subsequence of \( \{f^*_n\} \) contains a subsequence \( \{f^*_{n_j}\} \) that converges pointwise to some bounded function \( k^* \). Denote by \( k \) the Dirichlet solution in \( \mathbb{H}(\rho, 1) \) of the boundary function defined on \( \mathbb{T} \) by \( k^* \) and on \( \mathbb{T}_\rho \) by the average of \( k^* \) on \( \mathbb{T} \). Then, by Remark 2.1, \( f_{n_j} \to k \) locally uniformly on \( \mathbb{H}(\rho, 1) \). Hence \( k = f \). This implies \( \tilde{E}(f^*) = \tilde{E}(k^*) \) and \( f^*(e^{i\theta}) = k^*(e^{i\theta}) \) for \( e^{i\theta} \in \tilde{E}(f^*) \). Hence \( f^*_{n_j} \to f^* \) pointwise a.e.. It follows that \( f^*_n \to f^* \) pointwise a.e. and (c) \( \Rightarrow \) (a). This ends the proof. \[ \Box \]**
3. Geometry of Analytic Parts of Univalent Harmonic Mappings onto Punctured Convex Domains

Let \( h \) be the analytic part of \( f \in \mathcal{H}_u(\rho, \mathbb{G}) \). The purpose of this section is two-fold: First, to show that \( h \) has a nonvanishing derivative on \( T_\rho \), and that it maps \( T_\rho \) homeomorphically onto a sense-preserving convex Jordan curve whose diameter admits a universal upper bound, and second, to prove that \( h \) is a composition of a univalent close-to-convex function and a homeomorphism of \( A(\rho, 1) \cup T \) onto a ring subdomain of \( \mathbb{D} \) that maps \( T \) homeomorphically onto itself.

Our first result in this section relates univalent harmonic maps in \( \mathcal{H}_u(\rho, \mathbb{G}) \) to their average associates in \( \mathcal{H}_0(\rho, \mathbb{G}) \).

**Proposition 3.1.** Suppose that the following are true:

(i) \( f^* \) is a sense-preserving quasihomeomorphism of \( T \) into \( \partial \mathbb{G} \) such that 
\[
\overline{co}(f^*(\mathbb{E}(f^*))) = \mathbb{T}.
\]
(ii) \( f \) is the Dirichlet solution in \( A(\rho, 1) \) of the function defined by \( f^* \) on \( T \) and a constant \( \zeta \in \mathbb{G} \) on \( T_\rho \).
(iii) \( f_0 \) is the Dirichlet solution of the function defined by \( f^* \) on \( T \) and the average \( \zeta_0 \) of \( f^* \) on \( T_\rho \).

Then there is an analytic function \( h \) in \( A(\rho^2, 1) \) such that
\[
\begin{align*}
f(z) &= h(z) - h(\rho^2/z) + \zeta + 2c_\zeta \log(|z|/\rho) \quad (z \in A(\rho, 1)) \\
f_0(z) &= h_0(z) - h_0(\rho^2/z) + \zeta_0, \quad (z \in A(\rho, 1)).
\end{align*}
\]

where \( c_\zeta \) is given by (5.4).

**Proof.** By Theorem A, there is a constant \( c \) and an analytic function \( h \) of \( A(\rho^2, 1) \) such that
\[
f(z) = h(z) - h(\rho^2/z) + \zeta + 2c \log(|z|/\rho), \quad (z \in A(\rho, 1)).
\]

By Theorem 2.2, there is an analytic function \( h_0 \) of \( A(\rho^2, 1) \) such that
\[
f_0(z) = h_0(z) - h_0(\rho^2/z) + \zeta_0, \quad (z \in A(\rho, 1)).
\]

Then
\[
(f - f_0)(z) = (h - h_0)(z) - (h - h_0)(\rho^2/z) + \zeta - \zeta_0 + 2c \log(|z|/\rho)
\]
is a bounded harmonic mapping in \( A(\rho, 1) \). We conclude, by Schwarz’s theorem, that the unrestricted limit function of \( (f - f_0) \) exists everywhere on \( T \) except possibly on a countable subset \( E \). Furthermore, it is identically zero on \( T \setminus E \) by the definition of \( f \) and \( f_0 \). Since \( (f - f_0)(T_\rho) = \zeta - \zeta_0 \), the maximum principle yields
\[
(f - f_0)(z) = \zeta - \zeta_0 + 2c_\zeta \log(|z|/\rho)
\]
where \( c \) is as given in (5.4). By comparing (3.3) and (3.4), it follows that \( c = c_\zeta \) and \( h - h_0 \) is constant. This yields (3.1) and (3.2), and the proof is complete. 

Note that Proposition 3.1 does not require \( f \) to be in \( \mathcal{H}_a (\rho, \mathbb{G}) \). If this however is the case, then we obtain Corollary 3.1.

**Corollary 3.1.** Let \( f_0 \) be the average associate of \( f \in \mathcal{H}_a (\rho, \mathbb{G}) \) with \( f (T_\rho) = \zeta \) and \( f_0 (T_\rho) = \zeta_0 \). Then there is an analytic function \( h \) in \( \mathbb{A} (\rho^2, 1) \) such that (3.1) and (3.2) hold simultaneously.

Suppose now that \( f \in \mathcal{H}_a (\rho, \mathbb{G}) \) has form (1.2). According to Proposition 3.1, \( f \) and its average associate \( f_0 \) have the same analytic and coanalytic part \( h \). Since our interest in this section is exclusively in \( h \), we restrict ourselves to functions \( f \in \mathcal{H}_0 (\rho, \mathbb{G}) \) of form (2.2).

We shall need the notion of the *module* \( M (R) \) of a ring domain \( R \) [18]. It is known that \( R \) is conformally equivalent to a unique annulus \( A (r, 1) \), \( 0 < r < 1 \). In this case \( M (R) \) is defined by \( \log (1/r) \) if \( r \neq 0 \) and by \( \infty \) if \( r = 0 \). It is immediate that \( M \) is a conformal invariant, and that if \( R \subset R' \) where \( R' \) is also a ring domain, then \( M (R) \leq M (R') \) with equality if and only if \( R = R' \). The Grötzsch's ring domain, \( B (t), 0 < t < 1, \) of \( R \) is the doubly-connected open subset of \( \mathbb{D} \) whose boundary components are \( T \) and the segment \( \{ x : 0 \leq x \leq t \} \). Observe that \( B (t) \) is unique. The module of \( B (t) \) is usually denoted by \( \mu (t) \). It follows that if \( B (s) \) is the Grötzsch's ring domain of \( \mathbb{A} (\rho, 1) \), then \( \mu (s) = \log (1/\rho) \). It is known that \( \mu \) is a strictly decreasing function of \( [0, 1) \).

Let \( S \) be a subset of \( \mathbb{C} \). The *diameter* of \( S \) is the least upper bound of the distances between any two points of \( S \). If \( \ell_\alpha, \alpha \in \mathbb{R}, \) is a straight-line in the direction of \( e^{i\alpha} \) perpendicular to two support lines \( \pi \) and \( \pi' \) of \( S \), then we call the distance between \( \ell_\alpha \cap \pi \) and \( \ell_\alpha \cap \pi' \) the *width of \( S \) in the direction of \( e^{i\alpha} \).* It is known that if \( S \) is compact, then the diameter of \( S \) is equal to its maximum width [6, p. 77]. In what follows, we denote by \( d \) the diameter of \( G \) and by \( d_\alpha \) its diameter in the direction of \( e^{i\alpha} \). We call a Jordan curve *convex* if it is the boundary of a bounded convex domain.

Using these notions, our result states as follows.

**Theorem 3.1.** Suppose \( f \in \mathcal{H}_0 (\rho, \mathbb{G}) \) has form (2.2). Then

(a) \( h' \) is nonvanishing on \( T_\rho \) and \( h \) maps \( T_\rho \) homeomorphically onto a convex curve whose diameter is bounded above by

\[
D = \left( 4d/\pi \right) \tanh^{-1} \left( \mu^{-1} (\log (1/\rho)) \right).
\]

(b) If \( h(z) = \sum_{n=0}^{\infty} a_n z^n, \ z \in \mathbb{A} (\rho^2, 1) \), then

\[
\sum_{n=1}^{\infty} n |a_n| \rho^{-2n} \leq \sum_{n=1}^{\infty} n |a_n| \rho^{2n} \leq D^2/4 + \sum_{n=1}^{\infty} n |a_n| \rho^{-2n}.
\]
The proof of the theorem needs two lemmas. The first is due to Bshouty and Hengartner [1, Theorem 2.5]. To state this result, we call a ring domain $\Omega$ a **slit domain convex in the direction of the real axis** if it is obtained by removing a horizontal slit from a domain convex in the direction of the real axis.

**Lemma 3.1.** Suppose $f \in H_u(\rho, G)$ has form (2.2), and let

$$(3.5) \quad \Phi_\alpha(z) = e^{i\alpha}h(z) + e^{-i\alpha}\overline{h(\rho^2/z)}, \quad (z \in A(\rho^2, 1)).$$

Then $\Phi_\alpha$ is univalent in $A(\rho, 1)$ and it maps $A(\rho, 1)$ onto a slit domain convex in the direction of the real axis.

Our second lemma is intuitive and geometric in nature, and it needs some basic notions. A **closed curve** is a continuous image of $\mathbb{T}$; we use the same notation for the curve and its defining function. Let $\gamma$ be a closed curve, and let $\ell$ be a straight line. A point $w \in \gamma \cap \ell$ is called a **meeting point** of $\gamma$ and $\ell$ of multiplicity $n$ if $|\gamma^{-1}(w)| = n$. For a meeting point $w$ of $\gamma$ and $\ell$, we call $w$ a **crossing point** of $\gamma$ and $\ell$ if there is an open subarc $I$ of $\mathbb{T}$ such that $\gamma^{-1}(w) \cap \ell$ is a singleton and $\ell$ separates $\gamma(I \setminus \{w\})$.

**Lemma 3.2.** If every straight-line through the origin meets a closed curve $\gamma$ exactly twice, counting multiplicity, and at crossing points only, then $\gamma$ is a Jordan curve whose inner domain is starlike with respect to the origin.

**Proof.** We show first that $\gamma$ is a Jordan curve. Suppose that there are points $z_1, z_2 \in \mathbb{T}$ such that $\gamma(z_1) = \gamma(z_2) = w$. If $w = 0$, then any straight-line passing through the origin and some other point of $\gamma$ meets $\gamma$ in at least three points, counting multiplicity. If $w \neq 0$, for convenience $w > 0$, then $\gamma$ does not meet the negative real axis. This implies, by the compactness of $\gamma$, that $\gamma$ lies within a minimal sector vertexed at the origin whose sides meet $\gamma$ without crossing. In either case, we have a contradiction and the claim holds.

Next, we show that the winding number $n(\gamma, 0)$ is $\pm 1$. We consider two cases.

(i) $0 \in \gamma$: In this case $\gamma$ meets only one of the positive and negative real axes.

(ii) $0 \notin \gamma$: In this case $\gamma$ meets $\mathbb{R}$ in two distinct points $a$ and $b$, say $a < b$. Here also we consider two cases.

(a) $0 < a < b$ or $a < b < 0$: In the first case $\gamma$ does not meet the negative real axis, and in the second it does not meet the positive real axis.

(b) $a < 0 < b$.

In (i) and (ii.a), the above compactness argument yield a contradiction. Thus only (ii.b) holds. It is immediate then that $n(\gamma, x) = 0$ for all $x \in (-\infty, a) \cup (b, \infty)$. Because $\mathbb{R} \cap \gamma = \{a, b\}$ for all $a < x < b$, either $n(\gamma, x) \neq 0$ or $n(\gamma, x) = 0$. In the latter case $\gamma \setminus \{a, b\}$ lies completely in one of the upper- or lower-half planes and $\mathbb{R}$ fails to cross $\gamma$ at $a$ or $b$. Hence $|n(\gamma, 0)| = 1$. 
We further conclude that any straight-line passing through the origin meets the inner domain of $\gamma$ in an open segment containing the origin. Therefore, the inner domain of $\gamma$ is starlike with respect to the origin. \hfill \Box

Proof of Theorem 3.1. (a) Fix $\alpha \in \mathbb{R}$ and let $\Phi_\alpha$ be given as in (3.5). Then we can write

$$\Phi_\alpha(\rho e^{i\theta}) = 2\Re\{e^{i\alpha}h(\rho e^{i\theta})\}, \quad (0 \leq \theta \leq 2\pi).$$

Let $m_\alpha = \min_{\theta} \Phi_\alpha(\rho e^{i\theta})$, $M_\alpha = \max_{\theta} \Phi_\alpha(\rho e^{i\theta})$, and $\Gamma$ be the curve defined by $\Gamma(\theta) = h(\rho e^{i\theta})$, $0 \leq \theta \leq 2\pi$. Observe that $M_\alpha - m_\alpha$ is the width of $\Gamma$ in the direction of $e^{-i\alpha}$, and that $\Phi_\alpha$ maps $\mathbb{T}_\rho$ onto the real interval $I_\alpha = [m_\alpha, M_\alpha]$ which is the inner boundary of the ring domain $\Phi_\alpha(\mathbb{A}(\rho, 1))$. Since $\Phi_\alpha$ is univalent by Lemma 3.1, $\Phi_\alpha$ admits two simple zeros $\rho e^{i\alpha_1}$ and $\rho e^{i\alpha_2}$, where $\alpha_1 < \alpha_2 < \alpha_1 + 2\pi$, such that $\Phi_\alpha(\rho e^{i\alpha_1}) = m_\alpha$ and $\Phi_\alpha(\rho e^{i\alpha_2}) = M_\alpha$.

Letting $\Psi(\theta) = \Phi_\alpha(\rho e^{i\theta})$, $0 \leq \theta \leq 2\pi$, we obtain

$$\Psi'(\theta) = i\rho e^{i\theta}\Phi'_\alpha(\rho e^{i\theta}) = -2\Re\left\{e^{i\alpha}[pe^{i\theta}h'(pe^{i\theta})]\right\}.$$

The first equality yields $\Psi'(\alpha_1) = \Psi'(\alpha_2) = 0$, $\Psi'(\theta) > 0$ for $\alpha_1 < \theta < \alpha_2$, and $\Psi'(\theta) < 0$ for $\alpha_2 < \theta < \alpha_1 + 2\pi$. Denote by $\gamma$ the curve defined by $\gamma(\theta) = \rho e^{i\theta}h'(\rho e^{i\theta})$, $0 \leq \theta \leq 2\pi$. The second equality implies that the real axis meets the curve $e^{i\theta}\gamma$ exactly twice, counting multiplicity, and only at crossing points; namely $\rho e^{i\alpha_1}h'(\rho e^{i\alpha_1})$ and $\rho e^{i\alpha_2}h'(\rho e^{i\alpha_2})$. This means that the line in the direction of $e^{-i\alpha}$ meets $\gamma$ exactly twice and only at crossing points. Since $\alpha$ is arbitrary, this property also holds for all straight-lines passing through origin. Using Lemma 3.2, we conclude that $\gamma$ is a Jordan curve that bounds a starlike domain with respect to the origin. Thus $h'$ is nonvanishing on $\mathbb{T}_\rho$ and

$$\frac{d}{d\theta} \arg\rho e^{i\theta}h'(\rho e^{i\theta}) = \Re\left\{1 + \rho e^{i\theta}h''(\rho e^{i\theta})\right\} \frac{h'(\rho e^{i\theta})}{h''(\rho e^{i\theta})}$$

is always either nonpositive or nonnegative. Hence $\Gamma$ is a convex curve as claimed.

Now we show that the diameter of $\Gamma$ is bounded by $D$. With a fixed $\alpha$ again, we can write

$$\Phi_\alpha(z) = e^{i\alpha}(f(z) - \zeta_0) + 2\Re\{e^{i\alpha\rho^2/2}\}.$$

Geometrically, this means that for every $z \in \mathbb{A}(\rho, 1)$ the value $\Phi_\alpha(z)$ can be obtained from the point $e^{i\alpha}(f(z) - \zeta_0)$ by a horizontal shift by $2\Re\{e^{i\alpha\rho^2/2}\}$. Recall $d_\alpha$, $d$, and $I_\alpha$. We conclude that the ring domain $\Phi_\alpha(\mathbb{A}(\rho, 1))$ is properly contained in a horizontal strip of width $d_\beta$, $\beta = \pi/2 - \alpha$, and with a slit along $I_\alpha$. Let $S_\alpha$ and $S$ be the horizontal strips symmetric about $\mathbb{R}$ and of widths $2d_\beta$ and $2d$ respectively. Obviously, $\Phi_\alpha(\mathbb{A}(\rho, 1))$ is a proper subset of $S_\alpha \setminus I_\alpha$, $S_\alpha \setminus I_\alpha \subset S \setminus I_\alpha$, and
$S \setminus I_\alpha$ is conformally equivalent to $S \setminus [(m_\alpha - M_\alpha)/2,(M_\alpha - m_\alpha)/2]$. Observe that the length of the boundary slit of the Grötzsch’s domain of $S \setminus [(m_\alpha - M_\alpha)/2,(M_\alpha - m_\alpha)/2]$ is $\tanh[\pi(M_\alpha - m_\alpha)/(4d)]$. Then
\[
\log(1/\rho) = M(A(\rho,1)) < M(S_\alpha \setminus I_\alpha) \leq M(S \setminus [(m_\alpha - M_\alpha)/2,(M_\alpha - m_\alpha)/2]) = \mu(\tanh[\pi(M_\alpha - m_\alpha)/(4d)]).
\]
Since $\mu$ is a decreasing function, we obtain
\[
\tanh[\pi(M_\alpha - m_\alpha)/(4d)] < \mu^{-1}(\log(1/\rho)),
\]
or
\[
M_\alpha - m_\alpha < \frac{4d}{\pi} \tanh^{-1} (\mu^{-1}(\log(1/\rho))) = D.
\]
Note that $\alpha$ may be chosen so that $d = M_\alpha - m_\alpha$. This concludes (a).

(b) Let $\Omega$ be the closed region bounded by the curve $\Gamma$ defined in the proof of (a). We show first that the area $A(\Omega)$ of $\Omega$ is at most $\pi D^2/4$. By [6, Theorem 54], $\Omega$ is contained in a convex region $\Omega'$ of constant width $D$ in every direction. Then Cauchy’s theorem [6, p. 127] implies that the perimeter of $\Omega'$ is $\pi D$. But the area of $\Omega'$ is at most $\pi D^2/4$ by the isoperimetric inequality [6, p. 108]. This proves our claim.

On the other hand,
\[
A(\Omega) = \frac{1}{2i} \int_{|z|=\rho} \overline{h(z)} h'(z) \, dz
\]
\[
= \frac{1}{2} \int_0^{2\pi} \left\{ \sum_{n=-\infty}^{\infty} \overline{a_n} \rho^n e^{-in\theta} \right\} \left\{ \sum_{n=-\infty}^{\infty} n a_n \rho^n e^{in\theta} \right\} \, d\theta
\]
\[
= \pi \left\{ \sum_{n=1}^{\infty} n|a_n|^2 \rho^{2n} - \sum_{n=1}^{\infty} n|a_{-n}|^2 \rho^{-2n} \right\}.
\]
Therefore,
\[
\sum_{n=1}^{\infty} n|a_n|^2 \rho^{2n} - \sum_{n=1}^{\infty} n|a_{-n}|^2 \rho^{-2n} < D^2/4
\]
and (b) follows. \hfill \Box

Next, we embark on proving that the analytic part of every harmonic mapping in $H_0(\rho, G)$ is a univalent close-to-convex function of $D$ precomposed with a homeomorphism of $A(\rho,1) \cup T$ onto a ring subdomain of $D$ that maps $T$ homeomorphically onto itself. As above, it suffices to consider harmonic mappings $f \in H_0(\rho, G)$.

**Theorem 3.2.** Suppose $f \in H_0(\rho, G)$ has form (2.2). Then there is a univalent close-to-convex function $H$ of $D$ and a homeomorphism $\phi$ of $A(\rho,1) \cup T$ into $D$ with $\phi(T) = T$ such that $h = H \circ \phi$. 

Observe that if $f \in \mathcal{H}_0(\rho, \mathbb{G})$ is given by (2.2), then the dilatation of $f$ is given by (2.9).

The proof of the theorem needs two lemmas. The first states as follows.

**Lemma 3.3.** Fix $p$, $p = 2, 3, \ldots$. Suppose $f \in \mathcal{H}_0(\rho, \mathbb{G})$ has form (2.2) and an unrestricted limit function that satisfies the following properties:

(i) $f$ is a sense-preserving local homeomorphism of $\mathbb{T}$ onto $\partial \mathbb{G}$.
(ii) $f^{(p)}$ exists and is absolutely continuous on $\mathbb{T}$.
(iii) $f'$ is nonvanishing on $\mathbb{T}$.

Then

(a) $h$ extends to $\mathbb{H}(\rho^2, 1)$ such that $h(e^{i\theta})$ and $h(\rho^2 e^{i\theta})$ are continuously $(p - 1)$-differentiable with

$$\lim_{z \to re^{i\theta}} h^{(k)}(z) = h^{(k)}(re^{i\theta}), \quad (z \in \mathbb{H}(\rho^2, 1)), \quad (3.6)$$

where $r$ is either 1 or $\rho^2$.

(b) $h'(e^{i\theta}) \neq 0$ and $h'(\rho^2 e^{i\theta}) \neq 0$ for all $\theta$.

(c) $\omega$ extends continuously to $\mathbb{H}(\rho, 1)$ such that $\omega(e^{i\theta}) \neq -1$ and $|\omega(z)| \leq 1$ for $z \in \overline{\mathbb{H}(\rho, 1)}$.

**Proof.** (a) If

$$h(z) = \sum_{n=-\infty}^{\infty} a_n z^n, \quad (z \in \mathbb{H}(\rho^2, 1)), \quad (3.7)$$

then for $z = re^{i\theta}$, $\rho < r < 1$,

$$f(z) = \zeta_0 + \sum_{n \neq 0} a_n [r^n - (\rho^2/r)^n] e^{in\theta}. \quad (3.8)$$

Since $f'(e^{i\theta})$ exists for all $\theta$, [7, Theorem 55] gives

$$f(e^{i\theta}) = \zeta_0 + \sum_{n \neq 0} c_n e^{in\theta}$$

where, by the bounded convergence theorem and (3.7),

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\theta}) e^{-in\theta} d\theta = \lim_{r \to 1} \frac{1}{2\pi} \int_0^{2\pi} f(re^{i\theta}) e^{-in\theta} d\theta$$

$$= \lim_{r \to 1} a_n [r^n - (\rho^2/r)^n] = a_n(1 - \rho^{2n}).$$

Using this in (3.8), we obtain

$$f(e^{i\theta}) = \zeta_0 + \sum_{n \neq 0} a_n(1 - \rho^{2n}) e^{in\theta}. \quad (3.8)$$

Since $f^{(p)}$ is absolutely continuous, [7, Theorem 40] yields

$$a_n(1 - \rho^{2n}) = o(|n|^{-(p+1)}).$$
This gives for \( k = 1, 2, \ldots, p - 1 \),
\begin{equation} \label{eq:3.9}
 n(n - 1) \cdots (n - k + 1)a_n = o(|n|^{-p+k-1}) = o(|n|^{-2}).
\end{equation}

Define, for \( r = 1 \) or \( \rho^2 \),
\[ h(re^{i\theta}) = \sum_{n=-\infty}^{\infty} a_n r^n e^{in\theta}. \]

Observe that term by term differentiation of the latter series yields, by (3.9), a uniformly convergent series. Now term by term integration of the resulting series yields \( h(re^{i\theta}) \) continuously differentiable. Repeating the same procedure with \( h'(re^{i\theta}) \) yields \( h(re^{i\theta}) \) continuously 2-differentiable. Observe, again because of (3.9), that the same procedure can be repeated \( p - 1 \) times proving \( h(re^{i\theta}) \) continuously \((p - 1)\)-differentiable. Using (3.9) once again, together with the uniform convergence of \( k \)-th, \( k = 1, 2, \ldots, p - 1 \), derivatives of \( h(re^{i\theta}) \) and the above Laurent’s series of \( h(z) \), yields (3.6).

(b) The Jacobian of \( f \) is given by
\[ J(z) = |zh'(z)|^2 - |\rho^2 h'(\rho^2/z)|^2/|z|^2, \quad (z \in \mathbb{H}(\rho, 1)). \]

Since \( f \) is univalent in \( \mathbb{H}(\rho, 1) \), Lewy’s theorem \([13]\) yields \( J(z) > 0 \) for \( z \in \mathbb{H}(\rho, 1) \); that is,
\[ \rho^2|h'(\rho^2/z)| < |zh'(z)|, \quad (z \in \mathbb{H}(\rho, 1)), \]
which implies \( h'(z) \neq 0 \) in \( \mathbb{H}(\rho, 1) \). Using (a), we conclude
\[ \rho^2|h'(\rho^2 e^{i\theta})| \leq |h'(e^{i\theta})|, \quad (0 \leq \theta \leq 2\pi). \]

We infer that if \( h'(e^{i\theta}) = 0 \) for some \( \theta \), then \( \rho^2 e^{2i\theta} h'(\rho^2 e^{i\theta}) = 0 \). Note that
\begin{equation} \label{eq:3.10}
 f'(e^{i\theta}) = h'(e^{i\theta}) + \rho^2 e^{2i\theta} h'(\rho^2 e^{i\theta}).
\end{equation}

Thus \( f'(e^{i\theta}) = 0 \) which gives a contradiction. Hence \( h'(e^{i\theta}) \neq 0 \) for all \( \theta \). On the other hand, Theorem 3.1 yields \( h'(pe^{i\theta}) \neq 0 \) for all \( \theta \). This concludes (b).

(c) It is immediate from (a), (b), and (2.9) that \( \omega \) extends continuously to \( \mathbb{H}(\rho, 1) \). If \( \omega(e^{i\theta}) = -1 \) for some \( \theta \), then (2.9) and (3.10) give \( f'(e^{i\theta}) = 0 \) which leads to a contradiction. Now since \( f \) is univalent, Lewy’s theorem \([13]\) implies \( |\omega(z)| < 1 \) for \( z \in \mathbb{H}(\rho, 1) \). Using (2.9) once more, with (b), gives \( |\omega(z)| \leq 1 \) for \( z \in \mathbb{H}(\rho, 1) \). This completes the proof. \( \square \)

Our second lemma is a weaker form of Theorem 3.2.

**Lemma 3.4.** Suppose \( f \in \mathcal{H}_0(\rho, \mathbb{C}) \) has form (2.2), \( \partial \mathbb{G} \) an analytic curve, and \( f(e^{i\theta}) \) an infinite-differentiable function with a nonvanishing derivative. Let \( \Gamma \) be the convex curve defined by \( \Gamma(\theta) = h(pe^{i\theta}), \ 0 \leq \theta \leq 2\pi \) (see Theorem 3.1(a)). Then \( h \) is a sense-preserving homeomorphism of \( \mathbb{T}_\rho \) onto...
\( \Gamma \), and \( h = H \circ \phi \) where \( H \) is a univalent close-to-convex function of \( D \) and \( \phi \) is a homeomorphism of \( \mathbb{A}(\rho, 1) \cup T \) into \( \overline{\mathbb{D}} \) with \( \phi(T) = T \).

**Proof.** From Lemma 3.3, we infer that \( h(e^{i\theta}) \) is infinite-differentiable with a nonvanishing derivative. Using (3.10) and Lemma 3.3(c), we can write

\[
e^{i\theta} h'(e^{i\theta}) = \frac{e^{i\theta} f'(e^{i\theta})}{1 + \omega(e^{i\theta})}.
\]

Differentiation of both sides yields

\[
\Re \left\{ 1 + e^{i\theta} \frac{h''(e^{i\theta})}{h'(e^{i\theta})} \right\} = \Re \left\{ 1 + e^{i\theta} \frac{f''(e^{i\theta})}{f'(e^{i\theta})} \right\} - \Re \left\{ \frac{e^{i\theta} \omega'(e^{i\theta})}{1 + \omega(e^{i\theta})} \right\}.
\]

Denote by \( F \) a conformal map of \( D \) onto \( G \). Since \( \partial G \) is an analytic curve, \( F \) extends to a conformal map of \( D \) onto \( G \). Using the bounded convergence theorem and [18, p. 65], we obtain

\[
\Re \left\{ 1 + e^{i\Theta} \frac{F''(e^{i\Theta})}{F'(e^{i\Theta})} \right\} \geq 0, \quad (\Theta \in (-\infty, \infty)).
\]

Observe that we can write \( f(e^{i\theta}) = F(e^{i\Theta(\theta)}) \) where \( \Theta(\theta) \) is an increasing differentiable function of \( (-\infty, \infty) \) such that \( \Theta(\theta + 2\pi) = \Theta(\theta) + 2\pi \). It is easy to verify that

\[
\Re \left\{ 1 + e^{i\Theta} \frac{F''(e^{i\Theta})}{F'(e^{i\Theta})} \right\} = \Theta'(\theta) \Re \left\{ 1 + e^{i\Theta} \frac{F''(e^{i\Theta})}{F'(e^{i\Theta})} \right\}.
\]

Thus

\[
\int_0^{2\pi} \Re \left\{ 1 + e^{i\theta} \frac{f''(e^{i\theta})}{f'(e^{i\theta})} \right\} d\theta = \int_0^{2\pi} \Re \left\{ 1 + e^{i\Theta} \frac{F''(e^{i\Theta})}{F'(e^{i\Theta})} \right\} d\Theta = 2\pi,
\]

and

\[
\Re \left\{ 1 + e^{i\theta} \frac{f''(e^{i\theta})}{f'(e^{i\theta})} \right\} \geq 0, \quad (\theta \in (-\infty, \infty))
\]

since \( \Theta'(\theta) > 0 \). On the other hand, by Lemma 3.3(c), \( \Re [1 + \omega(e^{i\theta})] > 0 \). Since

\[
\Re \left\{ \frac{e^{i\theta} \omega'(e^{i\theta})}{1 + \omega(e^{i\theta})} \right\} = \frac{d}{d\theta} \arg [1 + \omega(e^{i\theta})],
\]

we conclude

\[
\int_0^{2\pi} \Re \left\{ \frac{e^{i\theta} \omega'(e^{i\theta})}{1 + \omega(e^{i\theta})} \right\} d\theta = 0
\]

and, for \( \theta_1 \leq \theta_2 < \theta_1 + 2\pi \),

\[
\int_{\theta_1}^{\theta_2} \Re \left\{ \frac{\omega'(e^{i\theta})}{1 + \omega(e^{i\theta})} \right\} d\theta = \arg \left\{ \frac{1 + \omega(e^{i\theta_2})}{1 + \omega(e^{i\theta_1})} \right\} > -\pi.
\]
Using (3.12), with (3.15) and (3.17), we get

\[ \int_0^{2\pi} \Re \left\{ 1 + e^{i\theta} \frac{h''(e^{i\theta})}{h'(e^{i\theta})} \right\} \, d\theta = 2\pi, \tag{3.19} \]

and, with (3.16) and (3.18), we get

\[ \int_{\theta_1}^{\theta_2} \Re \left\{ 1 + e^{i\theta} \frac{h''(e^{i\theta})}{h'(e^{i\theta})} \right\} \, d\theta > -\pi, \quad (\theta_1 \leq \theta_2 < \theta_1 + 2\pi). \tag{3.20} \]

Using Lemma 3.3(b) and (3.19), the argument principle gives

\[ \int_0^{2\pi} \Re \left\{ 1 + e^{i\tau} \frac{h''(e^{i\tau})}{h'(e^{i\tau})} \right\} \, d\tau = 2\pi. \tag{3.21} \]

This, together with Theorem 3.1(a), implies

\[ \Re \left\{ 1 + e^{i\tau} \frac{h''(e^{i\tau})}{h'(e^{i\tau})} \right\} \geq 0 \tag{3.22} \]

for all \( \tau \), and consequently the convex curve \( \Gamma \) is positively-oriented.

Now let \( \Omega \) be the convex domain bounded by \( \Gamma \). Since \( h \) is a sense-preserving homeomorphism of \( \mathbb{T} \) onto \( \Gamma \), Schoenflies theorem [18, p. 25] extends \( h \) to a local homeomorphism of \( \mathbb{D} \) which maps the closed disc bounded by \( \mathbb{T}_\rho \) homeomorphically onto \( \Omega \). Let \( W \) be the image surface of \( h \) in \( \mathbb{D} \). Note that \( W \) is a simply connected hyperbolic covering of \( \mathbb{C} \). Hence, by the Uniformization theorem, there is a locally univalent function \( H \) of \( \mathbb{D} \) with image surface \( W \). Define \( \phi = H^{-1} \circ h \); \( \phi \) is obviously a conformal map of \( A(\rho, 1) \) onto a ring subdomain of \( \mathbb{D} \) that extends conformally between the unit circles. Write \( \phi(e^{i\theta}) = e^{i\tau}, \ 0 \leq \theta, \tau \leq 2\pi \). Observe that \( H(e^{i\tau}) \) is infinite-differentiable with \( H'(e^{i\tau}) \neq 0 \), since both \( h \) and \( \phi \) are, and that \( \Im[e^{i\theta} \phi'(e^{i\theta})/\phi(e^{i\theta})] = 0 \) for all \( \theta \). Then direct computation yields

\[ \Re \left\{ 1 + e^{i\tau} \frac{H''(e^{i\tau})}{H'(e^{i\tau})} \right\} \, d\tau = \Re \left\{ 1 + e^{i\theta} \frac{h''(e^{i\theta})}{h'(e^{i\theta})} \right\} \, d\theta, \]

which, with (3.19), gives

\[ \int_0^{2\pi} \Re \left\{ 1 + e^{i\tau} \frac{H''(e^{i\tau})}{H'(e^{i\tau})} \right\} \, d\tau = 2\pi, \]

and, with (3.20), gives

\[ \int_{\theta_1}^{\theta_2} \Re \left\{ 1 + e^{i\tau} \frac{H''(e^{i\tau})}{H'(e^{i\tau})} \right\} \, d\tau > -\pi, \quad (\theta_1 \leq \theta_2 < \theta_1 + 2\pi). \]

It follows from Kaplan’s proof [11, Theorem 2] that \( H \) is a univalent close-to-convex function. This completes the proof. \( \square \)
Proof of Theorem 3.2. Using the ideas in the paragraph succeeding the statement of Theorem 2.2, there exists a sequence \( \{f_n\} \) of functions in \( \mathcal{H}_0(\rho, G) \) with form (2.4) such that each \( f_n(e^{i\theta}) \) is an infinite-differentiable function in \( T \), and \( f_n \to f \) and \( h_n \to h \) locally uniformly in \( \mathbb{A}(\rho, 1) \) and \( \mathbb{A}(\rho^2, 1) \) respectively. Let \( \Gamma \) be the convex curve defined by \( \Gamma(\theta) = h(e^{i\theta}), \) \( 0 \leq \theta \leq 2\pi \), and let \( \Omega \) be the convex domain bounded by \( \Gamma \). Also, let \( \Gamma_n \) be the convex curve defined by the function \( h_n(e^{i\theta}), \) \( 0 \leq \theta \leq 2\pi \). By Lemma 3.4, each \( h_n \) is a sense-preserving homeomorphism of \( T_\rho \) onto \( \Gamma_n \) with

\[
\mathfrak{R} \left\{ 1 + e^{i\theta} \frac{h''(e^{i\theta})}{h'(e^{i\theta})} \right\} \geq 0
\]

for all \( \theta \); see (3.22). Using this, with Lemma 2.2(a) and Theorem 3.1(a), we conclude that \( h \) satisfies (3.22) and, consequently, \( h \) is also a sense-preserving homeomorphism of \( T_\rho \) onto \( \Gamma \). Also, by Lemma 3.4, we have each \( h_n \) univalent in \( \mathbb{A}(\rho, 1) \). Hence, by Hurwitz’s theorem, \( h \) is also a univalent function on \( \mathbb{A}(\rho, 1) \) or else \( f \) is a constant. Define \( W \) as above, \( \Omega_n \) as the convex domain bounded by \( \Gamma_n \), and \( W_n = h_n(\mathbb{A}(\rho, 1)) \cup \Omega_n \). It is immediate that \( W \) and each \( W_n \) are simply connected domains in \( \mathbb{C} \). Fix a point \( \rho \in \Omega \). We show that

\[(3.23) \quad W_n \to W \quad \text{as} \quad n \to \infty \quad \text{with respect to} \quad \rho \]

in the sense of Carathéodary’s kernel convergence [18, pp. 13-15]. Let \( w_0 \in W \). We show first that \( w_0 \in W_n \) for sufficiently large \( n \). Let \( \gamma \) be a separating Jordan curve in \( \mathbb{A}(\rho, 1) \) with \( w_0 \) in the interior domain of \( h(\gamma) \). Since \( h_n \to h \) uniformly on \( \gamma \), \( w_0 \) belongs to the interior domain of the Jordan curve \( h_n(\gamma) \) for sufficiently large \( n \). Since each \( W_n \) is simply connected, \( w_0 \in W_n \) for sufficiently large \( n \). Now let \( w_0 \in \partial W \). We show that \( w_0 \) is the limit point of a sequence \( \{w_n\} \) where \( w_n \in \partial W_n \). Suppose that this is false. Then there is an increasing sequence of positive integers \( \{n_\nu\} \) and an open neighborhood \( V \) of \( w_0 \) such that \( \partial W_{n_\nu} \cap V = \emptyset \). Also, choose \( V \) so that \( \overline{\Omega} \cap \overline{V} = \emptyset \); this is possible since \( \Gamma_n \to \Gamma \). It follows that, for each \( n_\nu \), either \( V \cap W_{n_\nu} = \emptyset \) or \( V \subset W_{n_\nu} \). Suppose that the first case happens infinitely often, say, without loss of generality, for all \( \nu \). Then \( h_{n_\nu}(z) \notin V \) for \( z \in \mathbb{A}(\rho, 1) \). Since \( h_{n_\nu}(z) \to h(z) \), \( h(\mathbb{A}(\rho, 1)) \cap V = \emptyset \) and we have a contradiction. Now suppose, without loss of generality, that \( V \subset W_{n_\nu} \) for all \( \nu \). Then the inverse function \( \psi_{n_\nu}(w) = h_{n_\nu}^{-1}(w) \) is analytic in \( V \) with \( |\psi_{n_\nu}(w)| < 1 \). By Montel’s theorem, we can find a subsequence of \( \{\psi_{n_\nu}\} \) that converges locally uniformly in \( V \). Suppose, without loss of generality, that \( \{\psi_{n_\nu}\} \) converges locally uniformly in \( V \). Then the limit function \( \psi \) satisfies \( \rho \leq |\psi(w)| \leq 1 \) for \( w \in V \). By Hurwitz’s theorem, either \( \psi \) is a constant or is a univalent function in \( V \). We show that the latter holds. To do so, we show first that \( \{\psi_{n_\nu}\} \) converges locally uniformly in \( h(\mathbb{A}(\rho, 1)) \) even though these functions may not be defined in \( h(\mathbb{A}(\rho, 1)) \) in the proper sense. Let
Δ be a closed Jordan region in \( h(\mathcal{H}(\rho, 1)) \), and let \( K \) be a compact subset of the interior \( \Delta \). Since \( h \) is univalent, \( h^{-1}(\Delta) \) is a closed Jordan region in \( \mathcal{H}(\rho, 1) \) whose interior contains \( h^{-1}(K) \). Since \( h_{n_{\nu}} \to h \) uniformly on \( h^{-1}(\Delta) \), an argument using Rouche’s theorem implies that \( K \subset h_{n_{\nu}} \circ h^{-1}(\Delta) \) or \( K \subset h_{n_{\nu}}(\mathcal{H}(\rho, 1)) \) for sufficiently large \( n_{\nu} \). A compactness argument also yields the same conclusion for any compact subset \( K \) of \( h(\mathcal{H}(\rho, 1)) \). So, for a given compact subset \( K \) of \( h(\mathcal{H}(\rho, 1)) \), the functions \( \psi_{n_{\nu}} \) are defined on \( K \) for sufficiently large \( n_{\nu} \). Since the range of each \( \psi_{n_{\nu}} \) is \( \mathcal{H}(\rho, 1) \), the sequence \( \{ \psi_{n_{\nu}} \} \) is a normal family in \( h(\mathcal{H}(\rho, 1)) \). Since \( V \cap h(\mathcal{H}(\rho, 1)) \neq \emptyset \), \( \psi_{n_{\nu}} \to \psi \) in \( h(\mathcal{H}(\rho, 1)) \). Recall the above curve \( \gamma \). If \( \psi \) is constant, then \( \psi_{n_{\nu}}(h(\gamma)) \) admits an arbitrarily small diameter for large \( n_{\nu} \) which is impossible since each curve \( \psi_{n_{\nu}}(\gamma) \) separates \( \mathcal{H}(\rho, 1) \). Hence \( \psi \) is univalent in \( V \) and \( \rho < |\psi(w)| < 1 \) for \( w \in V \), in particular \( \psi(w_0) \in \mathcal{H}(\rho, 1) \). It follows that \( \{ h_{n_{\nu}} \} \) converges locally uniformly near \( \psi(w_0) \). Since \( \psi_{n_{\nu}}(w_0) \to \psi(w_0) \) and \( w_0 = h_{n_{\nu}} \circ \psi_{n_{\nu}}(w_0) \), we conclude \( w_0 = h(\psi(w_0)) \). This contradicts \( w_0 \in \partial V \) and (3.23) holds.

Now define \( H \) as above but with the additional conditions \( H(0) = \rho \) and \( H'(0) > 0 \). Also, let \( H_n \) be the conformal map of \( \mathbb{D} \) onto \( W_n \) satisfying \( H_n(0) = \rho \) and \( H_n'(0) > 0 \). By Carathéodory’s kernel theorem [18, pp. 13-15], \( H_n \to H \) locally uniformly in \( \mathbb{D} \). Since, by Lemma 3.4, each \( H_n \) is a univalent close-to-convex function, \( H \) is also a univalent close-to-convex function. Letting \( \phi = H^{-1} \circ h \). It is easily seen that \( \phi \) satisfies the desired properties. This ends the proof. \( \square \)

4. Univalent Harmonic Mappings onto Punctured Convex Domains.

Let \( f \) be the Dirichlet solution in \( \mathcal{H}(\rho, 1) \) of a function \( f^* \) of \( \partial \mathcal{H}(\rho, 1) \) defined on \( \mathbb{T} \) be a sense-preserving quasihomoeomorphism into \( \partial \mathcal{G} \) satisfying \( \overline{c} f^*(\overline{E}(f^*)) = \overline{G} \), and on \( \mathbb{T}_\rho \) by a constant \( \zeta \in \mathcal{G} \). Theorem 2.2 asserts that \( f \) belongs to \( \mathcal{H}_u(\rho, \mathcal{G}) \) if \( \zeta = \zeta_0 \), where \( \zeta_0 \) is the average of \( f^* \) on \( \mathbb{T}_\rho \) given by (1.3). Recently however, Duren and Hengartner [5, Example 1] observed that this condition is not necessary, and showed that the harmonic mapping

\[
F(z) = (z - \rho^2/\mathbb{T})/(1 - \rho^2) + 2c \log |z|, \quad (z \in \mathcal{H}(\rho, 1)),
\]

belongs to \( \mathcal{H}_u(\rho, \mathbb{D}) \) with \( f(0) = 2c \log \rho \) if \( |c| < \rho/(1 - \rho^2) \). Note that the boundary function of \( F \) is the identity map on \( \mathbb{T} \) and the constant \( 2c \log \rho \) on \( \mathbb{T}_\rho \). In view of this, Hengartner [2, Problem 15] suggested the problem of finding the set of values \( \zeta \in \mathcal{G} \) that yields \( f : \mathcal{H}(\rho, 1) \to G \setminus \{ \zeta \} \) a homeomorphism.

Now, let \( f^* \) be a sense-preserving quasihomoeomorphism of \( \mathbb{T} \) into \( \partial \mathcal{G} \) with \( \overline{c} f^*(\overline{E}(f^*)) = \overline{G} \). Denote by \( \mathcal{H}(\rho, f^*) \) the class of Dirichlet solutions in \( \mathcal{H}(\rho, 1) \) of functions of \( \partial \mathcal{H}(\rho, 1) \) defined on \( \mathbb{T} \) by \( f^* \) and on \( \mathbb{T}_\rho \) by some
constant ζ ∈ ℂ. Also, denote by \( \mathcal{H}_u(\rho, f^*) \) the subclass of \( \mathcal{H}(\rho, f^*) \) of univalent mappings. Of interest shall be the set \( \mathcal{K}(\rho, f^*) \) for which a function \( f \in \mathcal{H}(\rho, f^*) \) belongs to \( \mathcal{H}_u(\rho, f^*) \).

Our first result in this section states that \( \mathcal{K}(\rho, f^*) \) is compact. In view of Proposition 3.1, the class \( \mathcal{H}(\rho, f^*) \) yields an analytic function \( h \) in \( \mathbb{H}(\rho^2, 1) \), unique up to an additive constant, such that every \( f \in \mathcal{H}_u(\rho, f^*) \) is of the forms (3.1) and (3.2). In our second result, we characterize in terms of \( h \) and \( f^* \) the boundary points of \( \mathcal{K}(\rho, f^*) \) in a manner leading to a univalence criterion for functions \( f \in \mathcal{H}(\rho, f^*) \). Finally, we provide sufficient conditions on \( \rho \), \( G \), and \( f^* \) that warrant a nonempty interior for \( \mathcal{K}(\rho, f^*) \).

**Theorem 4.1.** \( \mathcal{K}(\rho, f^*) \) is a nonempty compact subset of \( \mathbb{G} \).

**Proof.** Let \( \zeta_0 \) be the average of \( f^* \) on \( T \). It is immediate from Theorem 2.2 that \( \zeta_0 \in \mathcal{K}(\rho, f^*) \). Hence \( \mathcal{K}(\rho, f^*) \neq \emptyset \).

Suppose that a sequence \( \{\zeta_n\}_{n=1}^\infty \) in \( \mathcal{K}(\rho, f^*) \) converges to \( \zeta \in \overline{\mathbb{G}} \). We show that \( \zeta \in \mathcal{K}(\rho, f^*) \). Clearly, there is a unique function \( f_n \in \mathcal{H}_u(\rho, f^*) \) such that \( f_n(T_\rho) = \zeta_n \). By Proposition 3.1, we can find an analytic function \( h \) in \( \mathbb{H}(\rho^2, 1) \), unique up to an additive constant, such that

\[
(4.2) \quad f_n(z) = h(z) - h(\rho^2/z) + \zeta_n + 2c_n \log(|z|/\rho), \quad (z \in \mathbb{H}(\rho, 1)),
\]

where \( c_n = (\zeta_n - \zeta_0)/(2\log \rho) \). Obviously, \( c_n \to c = (\zeta - \zeta_0)/(2\log \rho) \) as \( n \to \infty \). Using \( h \) and \( c \), we define the harmonic mapping \( f \) as in (1.2). If \( c = 0 \), then \( \zeta = \zeta_0 \in \mathcal{K}(\rho, f^*) \) by Theorem 2.2. So, suppose that \( c \neq 0 \). Then \( f_n \to f \) (locally) uniformly in \( \mathbb{H}(\rho, 1) \). It is easy to see that the Jacobian of \( f_n \) is given by

\[
(4.3) \quad J_n(z) = \frac{|zh'(z) + c_n|^2 - |(\rho^2/z)h'(\rho^2/z) + c_n|^2}{|z|^2}, \quad (z \in \mathbb{H}(\rho, 1)).
\]

Since \( f_n \) is univalent and sense-preserving, \( J_n(z) > 0 \), and consequently \( |zh'(z) + c_n| \neq 0 \) for \( z \in \mathbb{H}(\rho, 1) \). But \( zh'(z) + c_n \to zh'(z) + c \) uniformly in \( \mathbb{H}(\rho, 1) \). Hence, by Hurwitz’s theorem, either \( zh'(z) + c \neq 0 \) or \( zh'(z) = 0 \) for \( z \in \mathbb{H}(\rho, 1) \). If the latter holds, then \( h'(z) = -c/z \) which contradicts the analyticity of \( h \) in \( \mathbb{H}(\rho, 1) \). Hence \( zh'(z) + c \neq 0 \). The Jacobian of \( f \) is now given by

\[
J(z) = \frac{|zh'(z) + c|^2 - |(\rho^2/z)h'(\rho^2/z) + c|^2}{|z|^2}, \quad (z \in \mathbb{H}(\rho, 1)).
\]

Clearly, \( J_n z \to J(z) \). Since \( J_n(z) > 0 \), \( J(z) \geq 0 \). Thus

\[
|(\rho^2/z)h'(\rho^2/z) + c| \leq |zh'(z) + c|, \quad (z \in \mathbb{H}(\rho, 1)).
\]

We prove that this inequality must be strict. Suppose that equality holds for some \( z \). Then the maximum principle yields that the dilatation of \( f \)
given by

\[
\omega(z) = \frac{(\rho^2/\overline{z})h'(\rho^2/\overline{z}) + \overline{c}}{zh'(z) + c}, \quad (z \in \mathbb{A}(\rho, 1)),
\]

is a unimodular constant $e^{2i\alpha}$ for some real $\alpha$. That is,

\[
\frac{(\rho^2/\overline{z})h'(\rho^2/\overline{z}) + \overline{c}}{zh'(z) + c} = e^{2i\alpha}(zh'(z) + c), \quad (z \in \mathbb{A}(\rho, 1)).
\]

Since $h$ is analytic in $\mathbb{A}(\rho^2, 1)$, (4.5) holds for $z \in \mathbb{A}(\rho^2, 1)$. In particular, for all $\theta$,

\[
\rho e^{i\theta}h'(\rho e^{i\theta}) + c = e^{2i\alpha}[\rho e^{i\theta}h'(\rho e^{i\theta}) + c].
\]

This means that the function $zh'(z)$ maps $\mathbb{T}_\rho$ into the straight-line passing through $-c$ in the direction of $e^{-i\alpha}$. We conclude that $h(z)$ maps $\mathbb{T}_\rho$ to a straight-line in the direction of $e^{i(\pi/2 - \alpha)}$. This contradicts Theorem 3.1. Therefore,

\[
|(\rho^2/\overline{z})h'(\rho^2/\overline{z}) + c| < |zh'(z) + c| \quad (z \in \mathbb{A}(\rho, 1)).
\]

This yields $J(z) > 0$ for $z \in \mathbb{A}(\rho, 1)$, and consequently $f$ is locally univalent function by Lewy’s theorem [13]. Since each $f_n$ is univalent and $f_n \to f$ uniformly in $\mathbb{A}(\rho, 1)$, $f$ is univalent in $\mathbb{A}(\rho, 1)$. Using this, with the fact $\text{arg}^* f^*(E(f^*)) = \mathbb{T}$, we infer, by Remark 2.2, that $f : \mathbb{A}(\rho, 1) \to \mathbb{G} \setminus \{\zeta\}$ is a homeomorphism. Therefore $\zeta \in K(\rho, f^*)$ and the proof is complete. \qed

Our second result is Theorem 4.2.

**Theorem 4.2.** Let $f \in \mathcal{H}_u(\rho, f^*)$ be of form (3.1), where $f^* : \mathbb{T} \to \partial \mathbb{G}$ is a twice-differentiable function with nonvanishing derivative and absolutely continuous second derivative. Then the dilatation of $f$ and $zh'(z) + c_\zeta$ extend continuously to $\mathbb{A}(\rho, 1) \cup \mathbb{T}$ such that $e^{i\theta}h'(e^{i\theta}) + c_\zeta \neq 0$ for all $\theta$. Moreover, we have:

(a) If $\zeta \in \partial K(\rho, f^*)$, then either $\rho e^{i\theta_1}h'(\rho e^{i\theta_1}) + c_\zeta = 0$ for some $\theta_1$, or $|\omega(e^{i\theta_2})| = 1$ for some $\theta_2$.

(b) If $|\omega(e^{i\theta})| = 1$ for some $\theta$, then $\zeta \in \partial K(\rho, f^*)$.

(c) If in (a) and (b) the function $|\omega(e^{i\theta})|$ is replaced by the function

\[
2\Re \left\{ \frac{e^{i\theta}h'(e^{i\theta}) + c_\zeta}{e^{i\theta}f'(e^{i\theta})} \right\},
\]

then (a) and (b) continue to hold.

Regarding (a), a result of Hengartner and Szynal [10, Theorem 3.1] asserts that if $\zeta \in \partial K(\rho, f^*)$ then $\rho e^{i\theta_1}h'(\rho e^{i\theta_1}) + c_\zeta$ has at most one zero which is of order one.

**Proof.** The Jacobian of $f$ is given by

\[
J(z) = ||zh'(z) + c_\zeta|^2 - |(\rho^2/\overline{z})h'(\rho^2/\overline{z}) + c_\zeta||^2/|z|^2, \quad (z \in \mathbb{A}(\rho, 1)).
\]
Since $f$ is univalent and sense-preserving, Lewy’s theorem \([13]\) yields $J(z) > 0$. This implies $zh'(z) + c_{\zeta} \neq 0$ for $z \in A(\rho, 1)$. By Lemma 3.3, $h$ has a continuously differentiable extension to $A(\rho^2, 1)$ such that $h'(e^{i\theta}) \neq 0$ and $h'(\rho^2 e^{i\theta}) \neq 0$ for all $\theta$. It follows that $J$ has a continuous extension to $A(\rho, 1)$ such that

$$J(e^{i\theta}) = |e^{i\theta} h'(e^{i\theta}) + c_{\zeta}|^2 - |\rho^2 e^{i\theta} h'(\rho^2 e^{i\theta}) + c_{\zeta}|^2,$$

and $J(e^{i\theta}) \geq 0$ for all $\theta$. If for some $\theta$, $e^{i\theta} h'(e^{i\theta}) + c_{\zeta} = 0$, then $\rho^2 e^{i\theta} h'(\rho^2 e^{i\theta}) + c_{\zeta} = 0$, and consequently

$$e^{i\theta} f'(e^{i\theta}) = e^{i\theta} h'(e^{i\theta}) - \rho^2 e^{i\theta} h'(\rho^2 e^{i\theta}) = 0$$

which gives a contradiction. Hence, $e^{i\theta} h'(e^{i\theta}) + c_{\zeta} \neq 0$ for all $\theta$.

It also follows that the dilatation of $f$ given by

$$\omega(z) = \frac{(\rho^2/\overline{z}) h'(\rho^2/\overline{z}) + c_{\zeta}}{zh'(z) + c_{\zeta}}, \quad (z \in A(\rho, 1)),$$

extends continuously to $A(\rho, 1) \cup T$ such that

$$|\omega(e^{i\theta})| = \left| \frac{\rho^2 e^{i\theta} h'(\rho^2 e^{i\theta}) + c_{\zeta}}{e^{i\theta} h'(e^{i\theta}) + c_{\zeta}} \right|.$$

(a) We proceed to prove (a) by contraposition. Suppose that $|\omega(e^{i\theta})| < 1$ for all $\theta$. Then

$$|\rho^2 e^{i\theta} h'(\rho^2 e^{i\theta}) + c_{\zeta}| < |e^{i\theta} h'(e^{i\theta}) + c_{\zeta}|.$$

By the compactness of $T$, we can find $\delta > 0$ such that

$$|\rho^2 e^{i\theta} h'(\rho^2 e^{i\theta}) + c_{\zeta}| < |e^{i\theta} h'(e^{i\theta}) + c_{\zeta}| - \delta$$

for all $\theta$. It follows that, for $|\eta - \zeta| < \delta \log(1/\rho)$ and any $\theta$,

$$(4.7) \quad |\rho^2 e^{i\theta} h'(\rho^2 e^{i\theta}) + c_{\eta}| < |e^{i\theta} h'(e^{i\theta}) + c_{\eta}|$$

where $c_{\eta} = (\eta - \zeta)/(2 \log \rho)$ (see (5.4)).

Suppose now that $\rho e^{i\theta} h'(\rho e^{i\theta}) + c_{\zeta} \neq 0$ for all $\theta$. Then, in view of the above, $zh'(z) + c_{\zeta} \neq 0$ for $z \in A(\rho, 1)$. Since $A(\rho, 1)$ is compact, there is $\sigma > 0$ such that $|zh'(z) + c_{\zeta}| > \sigma$ for $z \in A(\rho, 1)$. It follows that, for $|\eta - \zeta| < 2\sigma \log(1/\rho)$,

$$(4.8) \quad |zh'(z) + c_{\eta}| > 0, \quad (z \in A(\rho, 1)).$$

Then (4.7) and (4.8) hold for every $\eta$ satisfying

$$|\eta - \zeta| < \tau = \min\{2\delta \log(1/\rho), 2\sigma \log(1/\rho)\}.$$

For each such $\eta$, let

$$f_{\eta}(z) = h(z) - h(\rho^2/\overline{z}) + \eta + 2c_{\eta} \log(||z||/\rho), \quad (z \in A(\rho, 1)).$$
Then $f_\eta$ is a harmonic mapping whose dilatation is given by

$$\omega_\eta(z) = \frac{(\rho^2/|z|)h'(\rho^2/|z|) + c_\eta}{zh'(z) + c_\eta}, \quad (z \in \mathbb{A}(\rho, 1)).$$

Clearly, by (4.7) and (4.8), $\omega_\eta$ is an analytic function that extends continuously to $\mathbb{A}(\rho, 1)$ such that $|\omega_\eta(e^{i\theta})| < 1$ and $|\omega_\eta(\rho e^{i\theta})| = 1$ for all $\theta$. Hence, by the maximum principle, $|\omega_\eta(z)| < 1$. This yields, because of (4.8), that the Jacobian of $f_\eta$ is positive in $\mathbb{A}(\rho, 1)$, and consequently $f_\eta$ is a univalent sense-preserving harmonic mapping. Now, by invoking Theorem 2.1 and Remark 2.2, we conclude that each $f_\eta : \mathbb{A}(\rho, 1) \to \mathbb{G} \setminus \{\eta\}$ is a homeomorphism. Since this holds whenever $|\eta - \zeta| < \tau$, $\zeta$ is an interior point of $K(\rho, f^*)$ and we have a contradiction. This proves (a).

(b) Suppose that $|\omega(e^{i\theta_1})| = 1$ for some $\theta_1$. Then the Möbius transformation

$$T(z) = \frac{\rho^2 e^{i\theta_1} h'(\rho^2 e^{i\theta_1}) + z}{e^{i\theta_1} h'(e^{i\theta_1}) + z}$$

satisfies $|T(c_\zeta)| = 1$. Since, by (5.4), $|\eta - \zeta| = 2(c_\eta - c_\zeta) \log \rho$, any open neighborhood of $\zeta$ contains an $\eta$ such that $|\omega_\eta(c_\eta)| > 1$, or equivalently, $\omega_\eta(e^{i\theta_1}) > 1$ where $\omega_\eta$ is as defined above. Therefore, $\eta \not\in K(\rho, f^*)$ and $\phi \in \partial K(\rho, f^*)$.

(c) Since $e^{i\theta} h'(e^{i\theta}) + c_\zeta \neq 0$ for all $\theta$, using (3.1), we obtain

$$e^{i\theta} f'(e^{i\theta}) = [e^{i\theta} h'(e^{i\theta}) + c_\zeta] - [\rho^2 e^{i\theta} h'(\rho^2 e^{i\theta}) + c_\zeta]$$

$$= [e^{i\theta} h'(e^{i\theta}) + c_\zeta]$$

$$\times \left[1 - \left[\rho^2 e^{i\theta} h'(\rho^2 e^{i\theta}) + c_\zeta\right]/[e^{i\theta} h'(e^{i\theta}) + c_\zeta]\right].$$

Since $f'(e^{i\theta}) \neq 0$ for all $\theta$, we obtain

$$\frac{e^{i\theta} h'(e^{i\theta}) + c_\zeta}{e^{i\theta} f'(e^{i\theta})} = \frac{1}{1 - \left[\rho^2 e^{i\theta} h'(\rho^2 e^{i\theta}) + c_\zeta\right]/[e^{i\theta} h'(e^{i\theta}) + c_\zeta]}.$$

This implies that

$$2\Re \left\{\frac{e^{i\theta} h'(e^{i\theta}) + c_\zeta}{e^{i\theta} f'(e^{i\theta})}\right\} = 1$$

for some $\theta$ if and only if $|\omega(e^{i\theta})| = 1$; see (4.6). This proves (c). \qed

We apply Theorem 4.2 to a function $f \in \mathcal{H}_u(\rho, f^*)$ of form (2.2). In this case, $\zeta$ is the average $\zeta_0$ of $f^*$ on $\mathbb{T}$, $c_\zeta = 0$, $\rho e^{i\theta} h'(\rho e^{i\theta}) \neq 0$ for all $\theta$ by Theorem 3.1(a), and $|\omega(e^{i\theta})| = 1$ for some $\theta$ if and only if $\rho^2|h'(\rho^2 e^{i\theta})| = |h'(e^{i\theta})|$. We conclude the following Corollary 4.1.

**Corollary 4.1.** Let $f \in \mathcal{H}_u(\rho, f^*)$ be of form (2.2), where $f^*$ is as in Theorem 4.2. Then the following statements are equivalent:

(a) $\zeta_0 \in \partial K(\rho, f^*)$. 


(b) $\rho^2|h'(\rho^2 e^{i\theta})| = |h'(e^{i\theta})|$ for some $\theta$.
(c) $2\Re\{h'(e^{i\theta})/f'(e^{i\theta})\} = 1$ for some $\theta$.

The arguments used in the proof of Theorem 4.2 yield at once sufficient conditions for the univalence of functions in $\mathcal{H}(\rho, f^*)$ where $f^*$ is as in Theorem 4.2.

**Theorem 4.3.** Let $f \in \mathcal{H}(\rho, f^*)$ be of form (3.1), where $f^*$ be smooth as in Theorem 4.2. Then $f \in \mathcal{H}_u(\rho, f^*)$ if $zh'(z) + c \zeta \neq 0$ for $z \in \mathbb{A}(\rho, 1)$, and if one of the following two inequalities holds for all $\theta$:
(a) $|\omega(e^{i\theta})| \leq 1$.
(b) $2\Re\{e^{i\theta}h'(e^{i\theta}) + c \zeta e^{i\theta}f'(e^{i\theta})\} \geq 1$.

We remark that $f^*$ as defined in Theorem 4.2 yields, by Lemma 3.3, $zh'(z) \neq 0$ for $z \in \mathbb{A}(\rho, 1)$. This makes the above sufficiency condition, $zh'(z) + c \zeta \neq 0$ for $z \in \mathbb{A}(\rho, 1)$, easily achievable for functions $f \in \mathcal{H}(\rho, f^*)$ with appropriately small $c \zeta$.

Finally, we prove the existence of a large family of triplets, $0 < \rho < 1$, $\mathbb{G}_\rho$, $f^*$, where $\mathbb{G}_\rho$ is a bounded convex domain and $f^*_\rho : T \rightarrow \partial \mathbb{G}_\rho$ is a sense-preserving homeomorphism, such that $K(\rho, f^*)$ has a nonempty interior containing the average of $f^*$.

**Theorem 4.4.** Let $\Omega$ be a bounded convex domain, and let $h$ be a homeomorphism of $\overline{D}$ onto $\overline{\Omega}$ that maps $D$ conformally onto $\Omega$. Suppose that $h''$ is continuous on $D$, $h''(e^{i\theta})$ is absolutely continuous, and
\begin{equation}
\Re\left\{1 + e^{i\theta}h''(e^{i\theta})/h'(e^{i\theta})\right\} > 0
\end{equation}
for all $\theta$. Then there exists $\delta > 0$ such that for each $0 < \rho < \delta$ we can find a bounded convex domain $\mathbb{G}_\rho$ such that the harmonic mapping
\begin{equation}
f_\rho(z) = h(z) - h(\rho^2/\overline{z}), \quad (z \in \mathbb{A}(\rho, 1)),
\end{equation}
satisfies the following properties:
(i) $f_\rho : T \rightarrow \partial \mathbb{G}_\rho$ is a sense-preserving homeomorphism.
(ii) $f_\rho$ is continuously twice-differentiable on $\mathbb{A}(\rho, 1)$.
(iii) $f_\rho \in \mathcal{H}_0(\rho, \mathbb{G}_\rho)$.
(iv) There is $\sigma > 0$, depending on $\rho$, such that for any $|\zeta| < \sigma$ the function
\begin{equation}
f_\zeta(z) = h(z) - h(\rho^2/\overline{z}) + \zeta + 2c \zeta \log(|z|/\rho)
\end{equation}
belongs to $\mathcal{H}_u(\rho, \mathbb{G}_\rho)$.
Remark 4.1. (i) Without (4.9), the hypothesis of the theorem yields the following weaker form of (4.9):

\[
\Re \left\{ 1 + e^{i\theta} \frac{h''(e^{i\theta})}{h'(e^{i\theta})} \right\} \geq 0.
\]

(4.12)

To see this, observe that \(zh'(z)\) is a univalent starlike function in \(D\) which gives

\[
\Re \left\{ 1 + h''(z) \frac{h'(z)}{h'(z)} \right\} > 0, \quad (z = re^{i\theta} \in \mathbb{D}).
\]

(4.13)

Now, because \(h''\) extends continuously to \(D\), the integral

\[
\int_0^z h''(\zeta) \, d\zeta, \quad (z \in \overline{D}),
\]

where the differentiable path of integration from 0 to \(z\) lies in \(\overline{D}\), yields, by Cauchy’s theorem, the continuous extension of \(h'(z)\) to \(\overline{D}\). On the other hand, since \(zh'(z)\) is univalent in \(D\) and maps the origin to itself, \(zh'(z) \neq 0\) for \(z \in \overline{D}\). Then (4.12) follows at once by letting \(r \to 1\) in (4.13).

(ii) Using Kellogg and Warschawski [18, Theorem 3.6, p. 49], the hypothesis that \(h''(z)\) admits a continuous extension to \(D\) with absolutely continuous \(h''(e^{i\theta})\) follows if \(\partial G\) has a parametrization \(w(t), 0 \leq t \leq 2\pi\), whose first derivative is nonvanishing and second derivative is Lipschitz of order \(\alpha, 0 < \alpha < 1\).

Proof of Theorem 4.4. By the compactness of \(T\), there is \(q > 0\) such that

\[
\Re \left\{ 1 + e^{i\theta} \frac{h''(e^{i\theta})}{h'(e^{i\theta})} \right\} > q
\]

(4.14)

for all \(\theta\). For a fixed \(0 < \rho < 1\), let

\[
k_\rho(z) = h(z) - h(\rho^2 z), \quad (z \in \overline{D}).
\]

(4.15)

Then \(k_\rho\) is an analytic function in \(D\) with \(k_\rho(0) = 0\). We can write

\[
1 + e^{i\theta} \frac{k_\rho''(e^{i\theta})}{k_\rho'(e^{i\theta})} = 1 + e^{i\theta} \frac{h''(e^{i\theta})}{h'(e^{i\theta})} + e^{i\theta} q_\rho(e^{i\theta}),
\]

(4.16)

where

\[
q_\rho(e^{i\theta}) = \rho^2 e^{i\theta} \frac{h'(\rho^2 e^{i\theta})h''(e^{i\theta}) - \rho^2 h'(e^{i\theta})h''(\rho^2 e^{i\theta})}{h'(e^{i\theta})[h'(e^{i\theta}) - \rho^2 h'(\rho^2 e^{i\theta})]}.
\]

Let \(m_1 = \min_{e^{i\theta}} |h'(e^{i\theta})|\), \(M_1 = \max_{e^{i\theta}} |h'(e^{i\theta})|\), \(M_2 = \max_{e^{i\theta}} |h''(e^{i\theta})|\), and

\[
\delta = \min \left\{ \sqrt{\frac{m_1}{2M_1}}, \frac{m_1}{2} \sqrt{\frac{q}{M_1 M_2}} \right\}.
\]
Then for $0 < \rho < \delta$, it is easy to verify that $|q_\rho(e^{i\theta})| < q$ which gives $\Re q_\rho(e^{i\theta}) > -q$. Using (4.14) and (4.16), we obtain

\begin{equation}
\Re \left\{ 1 + e^{i\theta} \frac{k''(e^{i\theta})}{k'_\rho(e^{i\theta})} \right\} > 0.
\end{equation}

Using (4.15), we conclude that $k'$ and $k''$ extend continuously to $\overline{D}$. Moreover, since $zh'(z)$ is univalent and $0 < \rho < \delta$, $k'(z) \neq 0$ for $z \in \overline{D}$. It follows by the maximum principle and (4.17) that $\Re \left\{ 1 + zk''(z) \frac{e^{i\theta}}{k'_\rho(z)} \right\} > 0, \quad (z \in \overline{D}).$

Let $G_\rho = k_\rho(D)$. We conclude that $G_\rho$ is a bounded convex domain, and that $k_\rho$ is a sense-preserving homeomorphism of $\overline{D}$ onto $G_\rho$ that maps $D$ conformally onto $G_\rho$. Now define $f_\rho$ as in (4.10). Then, by (4.15), $f_\rho(e^{i\theta}) = k_\rho(e^{i\theta})$ which yields (i) and (ii). Furthermore,

\[ 0 = f_\rho(\rho e^{i\theta}) = \frac{1}{2\pi} \int_0^{2\pi} f_\rho(e^{i\theta}) \, d\theta. \]

Then (iii) follows at once from Theorem B. On the other hand, by the definition of $\delta$, we obtain

\[ \rho^2 |h'(\rho^2 e^{i\theta})| < \rho^2 M_1 < \delta^2 M_1 \leq \frac{m_1}{2} \leq |h'(e^{i\theta})|. \]

This implies $f'(e^{i\theta}) \neq 0$. Since $h''(e^{i\theta})$ is absolutely continuous, $f''(e^{i\theta})$ is also absolutely continuous. Now an application of Corollary 4.1 implies (iv). This completes the proof. \qed

5. Nitsche’s Question Revisited.

In this section, we determine explicitly all harmonic mappings $f \in \mathcal{H}_u(\rho, \mathbb{G})$ whose analytic parts extend analytically throughout $\overline{D}$. As a consequence, we conclude that the function $f$ defined by (1.1) is the only harmonic mapping, up to rotation, in $\mathcal{H}_0(\rho, \mathbb{D})$, (here $\mathbb{G}$ is taken as $\mathbb{D}$), of $\Lambda(\rho, 1)$ onto $\Lambda(0, 1)$ whose analytic part is analytic in $\overline{D}$. This somehow justifies Nitsche’s question above.

**Definition 5.1.** Let $f \in \mathcal{H}_u(\rho, \mathbb{G})$. Then, by Theorem 2.1, the unrestricted limit function of $f$ coincides with a sense-preserving quasihomoeomorphism $f^*$ except possibly on a countable subset of $\mathbb{T}$. We call the value $\zeta_0$ given by (1.3) the average of $f$ on $\mathbb{T}$. Denote by $f_0$ the Dirichlet solution in $\Lambda(\rho, 1)$ of the boundary function which coincides with $f^*$ on $\mathbb{T}$ and is the constant $\zeta_0$ on $\mathbb{T}_\rho$. (By virtue of Theorem 2.2, $f_0 \in \mathcal{H}_0(\rho, \mathbb{G})$.) We call $f_0$ the average associate of $f$.

The result of this section is Theorem 5.1.
Theorem 5.1. Suppose \( f \in H_u(\rho, \mathcal{G}) \) has form (1.2) with \( \zeta_0 \) the average of \( f \) on \( T \). If \( h \) is analytic in \( \mathbb{D} \), then

\[
(5.1) \quad f(z) = \sum_{n=1}^{\infty} \frac{\lambda_n b_n}{1 - \rho^{2n}} [z^n - (\rho^n/z^n)^n] + \zeta + 2c_\zeta \log(|z|/\rho)
\]

\[
(5.2) \quad = \sum_{n=1}^{\infty} \frac{\lambda_n b_n}{1 - \rho^{2n}} [z^n - (\rho^n/z^n)^n] + \zeta_0 + 2c_\zeta \log |z|,
\]

where \( b_n, n = 1, 2, \ldots \), is the \( n \)-th coefficient of the conformal map

\[
(5.3) \quad F(z) = \zeta_0 + \sum_{n=1}^{\infty} b_n z^n
\]

of \( \mathbb{D} \) onto \( \mathcal{G} \) satisfying \( F(0) = \zeta_0 \), and

\[
(5.4) \quad c_\zeta = \frac{\zeta - \zeta_0}{2 \log \rho}.
\]

Proof. By virtue of Proposition 3.1, it suffices to prove the theorem for the average associate \( f_0 \) of \( f \). Using the proposition, we write

\[
(5.5) \quad f_0(z) = h(z) - h(\rho z) + \zeta_0, \quad (z \in \mathbb{A}(\rho, 1)).
\]

Since \( h \) is analytic in \( \mathbb{D} \), the function

\[
(5.6) \quad q(z) = h(z) - h(\rho z) + \zeta_0
\]

is analytic in \( \mathbb{D} \), maps the origin to \( \zeta_0 \), and satisfies

\[
(5.7) \quad \lim_{|z| \to 1} [f_0(z) - q(z)] = \lim_{|z| \to 1} [h(\rho z) - h(\rho z)] = 0.
\]

This implies that \( f_0 \) and \( q \) have the same cluster set at each \( \xi \in T \). But \( C(f_0, \xi) \subset \partial \mathcal{G} \) for \( \xi \in T \). Hence, by \([18, Corollary 2.10] \), \( q \), and consequently \( f_0 \) by (5.7), has a continuous extension to \( \overline{\mathbb{D}} \) that assumes every value of \( \mathbb{G} \) exactly \( m \) times in \( \overline{\mathbb{D}} \). It follows that \( f_0(e^{\iota t}) = F(e^{\iota \varphi(t)}) \) where \( \varphi \) is a continuous increasing function of \((-\infty, \infty)\) with \( \varphi(t + 2\pi) = \varphi(t) + 2m\pi \). Using Theorem 2.1(v), we conclude \( m = 1 \). This implies that \( q \), like \( F \), is a conformal map of \( \mathbb{D} \) onto \( \mathcal{G} \) with \( q(0) = \zeta_0 \). By Schwarz’s lemma,

\[
(5.8) \quad q(z) = F(\lambda z)
\]

for some unimodular constant \( \lambda \).

Suppose

\[
 h(z) = a_0 + \sum_{n=1}^{\infty} a_n z^n, \quad (z \in \mathbb{D}).
\]

Then (5.3), (5.6) and (5.8) yield

\[
 q(z) = \zeta_0 + \sum_{n=1}^{\infty} a_n (1 - \rho^{2n}) z^n = \zeta_0 + \sum_{n=1}^{\infty} \lambda^n b_n z^n.
\]
This gives
\[ a_n = \frac{\lambda^n b_n}{1 - \rho^{2n}} \quad (n = 1, 2, \ldots). \]

Using (5.5), we obtain
\[ f_0(z) = \zeta_0 + \sum_{n=1}^{\infty} \frac{\lambda^n b_n}{1 - \rho^{2n}} [z^n - (\rho^2/\bar{z})^n]. \]

This completes the proof. \(\square\)

If \(G = \mathbb{D}\), then Theorem 5.1 yields Corollary 5.1 by taking
\[ F(z) = \frac{z^2 + \zeta_0}{1 + \bar{\zeta}_0 z} = \zeta_0 + (1 - |\zeta_0|^2) \sum_{n=2}^{\infty} (-\zeta_0)^{n-1} z^n. \]

**Corollary 5.1.** Suppose \( f \in \mathcal{H}_u(\rho, \mathbb{D}) \) has form (1.2) with \( \zeta_0 \) the average of \( f \) on \( \mathbb{T} \) and \( h \) analytic in \( \mathbb{D} \). Then there is a unimodular constant \( \lambda \) such that
\[ f(z) = \lambda(1 - |\zeta_0|^2) \left\{ \frac{z^2 - \rho^2/\bar{z}}{1 - \rho^2} + \sum_{n=2}^{\infty} \frac{(-\lambda \zeta_0)^{n-1}}{1 - \rho^{2n}} [z^n - (\rho^2/\bar{z})^n] \right\} \]
\[ + \zeta + 2c_\zeta \log(|z|/\rho), \quad (z \in \mathbb{A}^*(\rho, 1)). \]

In particular, if \( \zeta_0 = 0 \), then
\[ f(z) = \lambda \frac{z^2 - \rho^2/\bar{z}}{1 - \rho^2} + 2c_\zeta \log|z|, \quad (z \in \mathbb{A}^*(\rho, 1)). \]

**References**


Received October 15, 1999.

**Department of Mathematics**
**American University of Beirut**
**Beirut**
**Lebanon**
E-mail address: lyzzaik@aub.edu.lb
We show that the Vassiliev invariants of the knots contained in an embedding of a graph $G$ into $\mathbb{R}^3$ satisfy certain equations that are independent of the choice of the embedding of $G$. By a similar observation we define certain edge-homotopy invariants and vertex-homotopy invariants of spatial graphs based on the Vassiliev invariants of the knots contained in a spatial graph. A graph $G$ is called adaptable if, given a knot type for each cycle of $G$, there is an embedding of $G$ into $\mathbb{R}^3$ that realizes all of these knot types. As an application we show that a certain planar graph is not adaptable.

1. Introduction.

Throughout this paper we work in the piecewise linear category. Let $G$ be a finite loopless graph. Let $\text{SE}(G)$ be the set of all embeddings of $G$ into the three-dimensional Euclidean space $\mathbb{R}^3$. An element of $\text{SE}(G)$ is called a spatial embedding of $G$. The image of a spatial embedding of $G$ is also called a spatial embedding of $G$ so long as no confusion occurs. A spatial embedding of a graph is called a spatial graph. The study of spatial graphs up to ambient isotopy is a branch of knot theory. The knots and links contained in a spatial graph are primitive invariants of the spatial graph. The following interesting theorem on knots and links in a spatial graph is shown in [3]. Let $K_n$ be the complete graph on $n$ vertices.

**Theorem 1.1 ([3]).**

1. Every spatial embedding of $K_6$ contains a non-trivial link.
2. Every spatial embedding of $K_7$ contains a nontrivial knot.

In fact the following theorem is shown in [3].

**Theorem 1.2 ([3]).**

1. For any element $f$ of $\text{SE}(K_6)$, the sum of the linking numbers of all two-component links contained in $f(K_6)$ is odd.
2. For any element $f$ of $\text{SE}(K_7)$, the sum of the Arf invariants of all knots in $f(K_7)$ containing all vertices of $f(K_7)$ is odd.
Since the linking number of a trivial two-component link is zero and the Arf invariant of a trivial knot is zero, Theorem 1.1 is an immediate corollary of Theorem 1.2. The above theorems imply that the knots and links in a spatial graph are mutually dependent in general. We refer the reader to [16], [7], [14] and [15] for related results and [21] for a higher dimensional analogue.

The main purpose of this paper is to give a generalization of Theorem 1.2. We show that the Vassiliev type invariants of order at most \( n \) of the knots contained in a spatial embedding of \( G \) satisfy certain equations that are independent of the choice of the embedding of \( G \) if the graph \( G \) is sufficiently large in comparison with \( n \). We remark here that the linking number and the Arf invariant are the Vassiliev type invariants of order 1 and 2 respectively.

We refer the reader to [23], [2], [1], [4], [17], [8], [12], [13] etc. for Vassiliev type invariants.

A graph is called planar if it is embeddable into a plane. A plane graph is a spatial graph that is ambient isotopic to a spatial graph on a plane in \( \mathbb{R}^3 \). As an application we show in Corollary 3.2 that the knot types in a spatial embedding of the graph in Figure 1.1 satisfy a nontrivial condition. In fact we have, for example, that there is no embedding of the graph in Figure 1.1 that contains just one trefoil knot containing all of the vertices of the graph, and contains no other nontrivial knots. Note that the graph in Figure 1.1 is the first planar graph which is known to have such a property.

![Figure 1.1.](image)

Then we define certain edge-homotopy invariants and vertex-homotopy invariants of spatial graphs based on the Vassiliev type invariants of the knots contained in a spatial graph. Here edge-homotopy allows crossing changes of an edge with itself and vertex-homotopy allows crossing changes of two adjacent edges. Then we show in Example 3.6 that none of the graphs in Figure 1.2 is edge-homotopic to a plane graph.
2. General results.

Let $\mathcal{R}$ be a commutative ring with unit 1. Let $v : \text{SE}(G) \to \mathcal{R}$ be an ambient isotopy invariant. Namely, $v$ is a map from $\text{SE}(G)$ to $\mathcal{R}$ such that $v(f) = v(g)$ for any ambient isotopic embeddings $f$ and $g$. Let $\text{SE}_i(G)$ be the set of all $i$-singular embeddings of $G$ into $\mathbb{R}^3$ where an $i$-singular embedding is a continuous map whose multiple points are exactly $i$ double points of edges spanning small flat planes. Such a double point is called a crossing vertex. Then, under a given edge orientation of $G$, $v$ is uniquely extended to an ambient isotopy invariant $v_i : \text{SE}_i(G) \to \mathcal{R}$ by the recursive formula

$$v_i(f_0) = v_{i-1}(f_+) - v_{i-1}(f_-)$$

where $f_0$, $f_+$ and $f_-$ are related as illustrated in Figure 2.1.

Here we only consider ambient isotopies that preserve a small flat plane at each crossing vertex. We say that $f_+$ (resp. $f_-$) is a positive (resp. negative)
resolution of $f_0$ at the crossing vertex illustrated in Figure 2.1. Let $n$ be a natural number. We say that $v$ is a Vassiliev type invariant of order at most $n$ if $v_{n+1} : SE_{n+1}(G) \to \mathcal{R}$ is a zero map. (Then we have that $v_m$ is a zero map for $m \geq n + 1$.) It is easy to check that the definition of Vassiliev type invariant of order at most $n$ is independent of the choice of the edge orientations.

Now we fix a ring $\mathcal{R}$ and a natural number $n$. Let $\Omega(G)$ be the set of all subgraphs of $G$. Let $\Gamma$ be a subset of $\Omega(G)$. Suppose that for each $\gamma \in \Gamma$, a Vassiliev type invariant of order at most $n$ $v_\gamma : SE(\gamma) \to \mathcal{R}$ is given. (Possibly $v_\gamma$ is a zero map.) Let $\omega : \Gamma \to \mathcal{R}$ be a map. Then we define an ambient isotopy invariant $v = v(\{v_\gamma\}, \omega) : SE(G) \to \mathcal{R}$ by

\[
v(f) = \sum_{\gamma \in \Gamma} \omega(\gamma)v_\gamma(f|_\gamma).
\]

Note that when $\Gamma = \{G\}$ and $\omega(G) = 1$ $v$ is just a Vassiliev type invariant of $G$ of order at most $n$. We include this case in the following. In general we have the following proposition.

**Proposition 2.1.** $v$ is a Vassiliev type invariant of order at most $n$.

**Proof.** Let $f$ be an element of $SE_{n+1}(G)$ and $c_1, \ldots, c_{n+1}$ the crossing vertices of $f$. Let $P$ be a subset of $\{1, \ldots, n+1\}$. By $f_P$ we denote an element of $SE(G)$ that is obtained from $f$ by resolving $c_i$ positively if $i$ is contained in $P$ and negatively if $i$ is not contained in $P$. Then we have

\[
v_{n+1}(f) = \sum_{P \subset \{1, \ldots, n+1\}} (-1)^{n+1-|P|}v(f_P)
\]

\[
= \sum_{P \subset \{1, \ldots, n+1\}} (-1)^{n+1-|P|} \left( \sum_{\gamma \in \Gamma} \omega(\gamma)v_\gamma(f_P|_\gamma) \right)
\]

\[
= \sum_{\gamma \in \Gamma} \omega(\gamma) \left( \sum_{P \subset \{1, \ldots, n+1\}} (-1)^{n+1-|P|}v_\gamma(f_P|_\gamma) \right).
\]

Therefore it is sufficient to show that

\[
\sum_{P \subset \{1, \ldots, n+1\}} (-1)^{n+1-|P|}v_\gamma(f_P|_\gamma) = 0
\]

for each $\gamma \in \Gamma \subset \Omega(G)$.

First suppose that some $c_i$ is not a crossing vertex of $f|_\gamma$. Then for each $P \subset \{1, \ldots, n+1\}$ with $i \in P$, $f_P|_\gamma$ and $f_{P-\{i\}}|_\gamma$ are ambient isotopic.
Therefore
\[
\sum_{P \subseteq \{1, \ldots, n+1\}} (-1)^{n+1-|P|} v_\gamma(f_P \mid \gamma)
= \sum_{P \subseteq \{1, \ldots, n+1\}, P \ni i} \left( (-1)^{n+1-|P|} v(f_P \mid \gamma) + (-1)^{n+1-|P|-1} v(f_{P\setminus\{i\}} \mid \gamma) \right) = 0.
\]

Next suppose that every \(c_i\) is a crossing vertex of \(f \mid \gamma\). Then we have that \(f \mid \gamma \in \text{SE}_{n+1}(\gamma)\) and
\[
\sum_{P \subseteq \{1, \ldots, n+1\}} (-1)^{n+1-|P|} v_\gamma(f_P \mid \gamma) = (v_\gamma)_{n+1}(f \mid \gamma) = 0.
\]

\[\square\]

An \(i\)-configuration on \(G\) is a pairing of \(2i\) points on the edges of \(G\). A realization of an \(i\)-configuration is an element of \(\text{SE}_i(G)\) whose crossing vertices correspond to the pairing. Let \(e_1\) and \(e_2\) be edges of \(G\) (possibly \(e_1 = e_2\)). We say that \(f\) and \(g\) in \(\text{SE}(G)\) are \((e_1, e_2)\)-homotopic if \(g\) is obtained from \(f\) by a series of crossing changes between \(e_1\) and \(e_2\) and ambient isotopy. Here a crossing change between \(e_1\) and \(e_2\) is a change of a crossing whose over-path belongs to an image of \(e_1\) and under-path belongs to an image of \(e_2\).

**Theorem S.** The following conditions are mutually equivalent.

1. \(v(f) = v(g)\) for any \((e_1, e_2)\)-homotopic embeddings \(f\) and \(g\) in \(\text{SE}(G)\),
2. \(v_i(f) = 0\) for any \(f\) in \(\text{SE}_i(G)\) with \(1 \leq i \leq n\) that has at least one crossing vertex of \(e_1\) and \(e_2\),
3. for any \(i\)-configuration \(C\) on \(G\) with \(1 \leq i \leq n\) that has at least one pair of points one on \(e_1\) and the other on \(e_2\), there is a realization \(f_C \in \text{SE}_i(G)\) of \(C\) such that \(v_i(f_C) = 0\).

**Proof.** (1) \(\rightarrow\) (2) It is sufficient to show that \(v_1(f) = 0\) for any 1-singular embedding \(f \in \text{SE}_1(G)\) that has just one crossing vertex of \(e_1\) and \(e_2\). The two resolutions \(f_+\) and \(f_-\) of \(f\) are \((e_1, e_2)\)-homotopic embeddings. Therefore \(v(f_+) = v(f_-)\). Thus \(v_1(f) = v(f_+) - v(f_-) = 0\).

(2) \(\rightarrow\) (1) It is sufficient to show the case that \(g\) is obtained from \(f\) by a crossing change of \(e_1\) and \(e_2\). Let \(h \in \text{SE}_1(G)\) be the 1-singular embedding that corresponds to the crossing change between \(f\) and \(g\). Then we have \(v(f) - v(g) = v_1(h) = 0\).

(2) \(\rightarrow\) (3) It is clear.

(3) \(\rightarrow\) (2) First we show that \(v_n(f) = 0\) for any \(f\) in \(\text{SE}_n(G)\) that has at least one crossing vertex of \(e_1\) and \(e_2\). Let \(C\) be the \(n\)-configuration on \(G\) that corresponds to \(f\). Let \(f_C\) be the realization of \(C\) with \(v_n(f_C) = 0\). We note that \(f\) and \(f_C\) are transformed into each other by a sequence of crossing changes and ambient isotopy. Since \(v_{n+1} : \text{SE}_{n+1}(G) \rightarrow \mathcal{R}\) is the
zero map (Proposition 2.1) crossing changes do not change the value of \( v_n \). Therefore we have \( v_n(f) = v_n(f_C) = 0 \).

Next suppose inductively on \( i \) that \( v_i(f) = 0 \) for any \( f \) in \( \text{SE}_i(G) \) that has at least one crossing vertex of \( e_1 \) and \( e_2 \). Then similarly we have that \( v_{i-1}(f) = 0 \) for any \( f \) in \( \text{SE}_{i-1}(G) \) that has at least one crossing vertex of \( e_1 \) and \( e_2 \). This completes the proof. □

As follows we can actually check whether or not these conditions hold. For each \( i \)-configuration \( C \) on \( G \) with \( 1 \leq i \leq n \) that has at least one pair of points one on \( e_1 \) and the other on \( e_2 \), choose any realization \( f_C \in \text{SE}_i(G) \) of \( C \). If \( v_i(f_C) \neq 0 \) for some configuration \( C \) then we have that the condition (2) does not hold. If \( v_i(f_C) = 0 \) for each configuration \( C \) then we have that the condition (3) holds.

We say that \( f \) and \( g \) in \( \text{SE}(G) \) are

(1) **edge-homotopic** if \( g \) is obtained from \( f \) by a series of “self-crossing changes” and ambient isotopy. Here a self-crossing change is a change of a crossing whose over-path and under-path belong to an edge of \( G \). Namely, edge-homotopy is the equivalence relation generated by \( (e, e) \)-homotopies for all edges \( e \),

(2) **vertex-homotopic** if \( g \) is obtained from \( f \) by a series of “crossing changes between adjacent edges” and ambient isotopy. Here a crossing change between adjacent edges is a change of a crossing whose over-path and under-path belong to two edges that have a common vertex. Namely, vertex-homotopy is the equivalence relation generated by \( (e_1, e_2) \)-homotopies for all pair of adjacent edges \( e_1, e_2 \).

We note that edge-homotopy and vertex-homotopy are equivalence relations on spatial graphs introduced in [18] as generalizations of Milnor’s link homotopy [9]. We remark that edge-homotopy implies vertex-homotopy since \( G \) is loopless [18].

The following three parallel theorems are immediate consequences of Theorem S.

**Theorem A.** The following conditions are mutually equivalent.

1. \( v(f) = v(g) \) for any \( f \) and \( g \) in \( \text{SE}(G) \),
2. \( v_i(f) = 0 \) for any \( f \) in \( \text{SE}_i(G) \) with \( 1 \leq i \leq n \),
3. for any \( i \)-configuration \( C \) on \( G \) with \( 1 \leq i \leq n \), there is a realization \( f_C \in \text{SE}_i(G) \) of \( C \) such that \( v_i(f_C) = 0 \).

**Theorem B.** The following conditions are mutually equivalent.

1. \( v(f) = v(g) \) for any edge-homotopic embeddings \( f \) and \( g \) in \( \text{SE}(G) \),
2. \( v_i(f) = 0 \) for any \( f \) in \( \text{SE}_i(G) \) with \( 1 \leq i \leq n \) that has at least one “self-crossing vertex”,
3. for any \( i \)-configuration \( C \) on \( G \) with \( 1 \leq i \leq n \) that has at least one pair of points on an edge, there is a realization \( f_C \in \text{SE}_i(G) \) of \( C \) such that \( v_i(f_C) = 0 \).
**Theorem C.** The following conditions are mutually equivalent.

1. \( v(f) = v(g) \) for any vertex-homotopic embeddings \( f \) and \( g \) in \( SE(G) \),
2. \( v_i(f) = 0 \) for any \( f \) in \( SE_i(G) \) with \( 1 \leq i \leq n \) that has at least one “crossing vertex of adjacent edges”,
3. for any \( i \)-configuration \( C \) on \( G \) with \( 1 \leq i \leq n \) that has at least one pair of points on two edges that have a common vertex, there is a realization \( f_C \in SE_i(G) \) of \( C \) such that \( v_i(f_C) = 0 \).

Here a self-crossing vertex is a crossing vertex of a single edge and a crossing vertex of adjacent edges is a crossing vertex of two edges that have a common vertex. We note that we can actually check whether or not the conditions in each of Theorem A, Theorem B and Theorem C hold as before. The proofs of Theorem A, Theorem B and Theorem C easily come from Theorem S and we omit them.

Suppose that for each \( \gamma \in \Gamma \subset \Omega(G) \), \( \phi_\gamma \in SE(\gamma) \) is given. Then we say that a set of spatial embeddings \( \{ \phi_\gamma \in SE(\gamma) \mid \gamma \in \Gamma \} \) is realizable up to ambient isotopy if there is an element \( f \) of \( SE(G) \) such that the restriction map \( f \mid _\gamma \) is ambient isotopic to \( \phi_\gamma \) for all \( \gamma \in \Gamma \). As an immediate corollary of Theorem A we have the following theorem.

**Theorem 2.2.** Suppose that the conditions of Theorem A hold. Let \( f \) be an element of \( SE(G) \). Then a set of spatial embeddings \( \{ \phi_\gamma \in SE(\gamma) \mid \gamma \in \Gamma \} \) is realizable up to ambient isotopy only if \( \sum_{\gamma \in \Gamma} \omega(\gamma) v_\gamma(\phi_\gamma) = v(f) = \sum_{\gamma \in \Gamma} \omega(\gamma) v_\gamma(f \mid _\gamma) \).

**Remark.** (1) Since the mod 2 linking number and the Arf invariant are order 1 and 2 Vassiliev type invariants respectively, Theorem 1.2 serves examples of Theorem 2.2 where the ring \( R = \mathbb{Z}/2\mathbb{Z} \). In fact it is not hard to check that the condition (3) of Theorem A holds for \( v \) the sum of the linking numbers of the links in a spatial embedding of \( K_6 \), and for \( v \) the sum of the Arf invariants of the knots in a spatial embedding of \( K_7 \) each of which contains all of the vertices of \( K_7 \). The spatial embedding of \( K_6 \) illustrated in Figure 2.2 contains just one nontrivial link as in Figure 2.2 and the spatial embedding of \( K_7 \) illustrated in Figure 2.3 contains just one nontrivial knot as in Figure 2.3, cf. [14]. Thus we have Theorem 1.2.

(2) Since the second coefficient of the Conway polynomial of a knot is an order 2 Vassiliev type invariant, the edge-homotopy (resp. vertex-homotopy) invariants defined in [19] are examples of Theorem B (resp. Theorem C).
We sometimes restrict our attention to certain class of embeddings of a graph $G$ and find some relations among the spatial embeddings of the subgraphs of $G$ that may not hold in an arbitrary embedding of $G$.

**Theorem D.** Let $T_i(G)$ be a subset of $SE_i(G)$ for each $i$ with $1 \leq i \leq n$ and $T_{n+1}(G) = SE_{n+1}(G)$. Suppose that for any $f \in T_i(G)$ with $1 \leq i \leq n$ there is an element $g \in T_i(G)$ with $v_i(g) = 0$ and a sequence of crossing changes from $f$ to $g$ such that the corresponding $(i+1)$-singular embeddings of $G$ are contained in $T_{i+1}(G)$. Then $v_i(h) = 0$ for any $h \in T_i(G)$ with $1 \leq i \leq n$.

**Proof.** First suppose that $h \in T_n(G)$. Then by the assumption there is $u \in T_n(G)$ with $v_n(u) = 0$ and a sequence of crossing changes from $h$ to $u$ such that the corresponding $(n + 1)$-singular embeddings are contained in $T_{n+1}(G) = SE_{n+1}(G)$. By Proposition 2.1 we have that $v_{n+1}$ is the zero map. Therefore we have $v_n(h) = v_n(u) = 0$ and $v_n(h) = 0$. Next suppose inductively that $v_{i+1}(h) = 0$ for any $h \in T_{i+1}(G)$. Let $h \in T_i(G)$. Then
there is \( u \in T_i(G) \) with \( v(u) = 0 \) and a sequence of crossing changes from \( h \) to \( u \) whose corresponding \((i+1)\)-singular embeddings are contained in \( T_{i+1}(G) \). Then we have \( v(h) - v(u) = 0 \) and \( v(h) = 0 \). \( \square \)

3. Graphs with multiple edges.

Let \( e_1, \ldots, e_m \) be some edges of \( G \). Let \( G(e_1, \ldots, e_m) \) be a graph obtained by adding edges \( d_1, \ldots, d_m \) to \( G \) so that \( e_i \) and \( d_i \) form a pair of multiple edges for each \( i \). Let \( \Lambda \) be a subset of \( \{1, \ldots, m\} \). Let \( G_\Lambda \) be a subgraph of \( G(e_1, \ldots, e_m) \) obtained from \( G \) by replacing \( e_i \) to \( d_i \) for each \( i \in \Lambda \). Then \( G_\Lambda \) is canonically isomorphic to \( G \). Let \( \varphi_\Lambda : G \to G_\Lambda \) be a canonical homeomorphism. Let \( \Gamma = \Gamma(e_1, \ldots, e_m) = \{ G_\Lambda \mid \Lambda \subset \{1, \ldots, m\} \} \). We remark here that \( G = G_\emptyset \) and the number of elements \( | \Gamma | = 2^m \).

Let \( v_G : SE(G) \to \mathcal{R} \) be a Vassiliev type invariant of order at most \( n \). Let \( v_{G_\Lambda} : SE(G_\Lambda) \to \mathcal{R} \) be the Vassiliev type invariant of order at most \( n \) defined by \( v_{G_\Lambda}(f) = v_G(f \circ \varphi_\Lambda) \). Let \( \omega : \Gamma \to \mathcal{R} \) be a map defined by \( \omega(G_\Lambda) = (-1)^{|\Lambda|} \).

Then we have the following theorem from Theorem A, Theorem B and Theorem C.

**Theorem 3.1.** Let \( v = v(\{v_{G_\Lambda}\}, \omega) : SE(G(e_1, \ldots, e_m)) \to \mathcal{R} \) be a map defined as above.

1. If \( n \) is less than \( m/2 \) then \( v \) is the zero map.
2. If \( n \) is less than \( (m+2)/2 \) then \( v \) is an edge-homotopy invariant.
3. If \( n \) is less than \( (m+1)/2 \) then \( v \) is a vertex-homotopy invariant.

Let \( e \) and \( d \) be a pair of multiple edges of \( G \) and \( f \in SE_i(G) \). We say that \( e \) and \( d \) are parallel in \( f \) if \( f(e \cup d) \) contains no crossing vertices of \( f \) and bounds a disk in \( R^3 \) whose interior is disjoint from \( f(G) \).

**Proof of Theorem 3.1 (1).** Suppose that some edges \( e_1 \) and \( d_1 \) are parallel in \( f \in SE_i(G(e_1, \ldots, e_m)) \). Let \( g \in SE(G(e_1, \ldots, e_m)) \) be any one of \( 2^i \) total resolutions of \( f \). Then we have that \( e_1 \) and \( d_1 \) are parallel in \( g \). Then

\[
v_i(g)
= \sum_{\Lambda \subset \{1, \ldots, m\}} (-1)^{|\Lambda|} v_{G_\Lambda}(g \mid_{G_\Lambda})
= \sum_{\Lambda \subset \{1, \ldots, m\}} (-1)^{|\Lambda|} v_{G}(g \mid_{G_\Lambda} \circ \varphi_\Lambda)
= \sum_{\Lambda \subset \{1, \ldots, m\}} \left( (-1)^{|\Lambda|} v_{G}(g \mid_{G_\Lambda} \circ \varphi_\Lambda) + (-1)^{|\Lambda - \{l\}|} v_{G}(g \mid_{G_{\Lambda - \{l\}}} \circ \varphi_{G_{\Lambda - \{l\}}}) \right).
\]

Since \( (-1)^{|\Lambda|} + (-1)^{|\Lambda - \{l\}|} = 0 \) and \( g \mid_{G_\Lambda} \circ \varphi_\Lambda \) and \( g \mid_{G_{\Lambda - \{l\}}} \circ \varphi_{G_{\Lambda - \{l\}}} \) are ambient isotopic we have \( v_i(g) = 0 \). Therefore we have \( v_i(f) = 0 \).
Now suppose that $C$ is an $i$-configuration on $G(e_1, \ldots, e_m)$ with $i \leq n$. Since $2i \leq 2n < m$ there is a pair of multiple edges $e_l$ and $d_l$ that have no points of $C$. Let $f_C \in \text{SE}_i(G(e_1, \ldots, e_m))$ be a realization of $C$ such that $e_l$ and $d_l$ are parallel in $f_C$. Then we have $v_i(f_C) = 0$. Thus the condition (3) of Theorem A holds. Since there is an embedding $f \in \text{SE}(G(e_1, \ldots, e_m))$ with $v(f) = 0$ we have the conclusion. □

A $j$-cycle graph is a graph on $j$ vertices that is homeomorphic to a circle. A $j$-cycle of $G$ is a subgraph of $G$ that is a $j$-cycle graph. Let $C_j(G) \subset \Omega(G)$ be the set of all $j$-cycles of $G$ and $C(G) = \bigcup_{j=2}^{\infty}C_j(G)$.

A graph $G$ is called adaptable if any set of embeddings $\{\phi_\gamma \in \text{SE}(\gamma) \mid \gamma \in C(G)\}$ is realizable up to ambient isotopy. In [5] Kinoshita showed that a graph on two vertices and some edges joining them is adaptable. Then in [24] Yamamoto showed that $K_4$ is adaptable. Then in [25] Yasuhara found a nice construction method of knots in a spatial graph using the fact that any knot is transformed into a trivial knot by delta unknotting operation introduced in [11]. For example the graph obtained from $K_5$ by deleting an edge is shown to be adaptable. On the other hand it is shown in [10] that no nonplanar graph is adaptable.

Let $C_m$ be an $m$-cycle graph and $e_1, \ldots, e_m$ the edges of $C_m$. Let $D_m = C_m(e_1, \ldots, e_m)$.

The following is a corollary of Theorem 3.1 (1).

**Corollary 3.2.** The graph $D_5$ is not adaptable.

We note that the graph $D_5$ is the first planar graph which is shown to be non-adaptable. A graph $H$ is called a minor of a graph $G$ if $H$ is obtained from $G$ by a sequence of edge-contraction, edge-deletion and vertex-deletion. If $H$ is a minor of $G$ then there is a natural injection from $C(H)$ into $C(G)$. Therefore it follows that a minor of an adaptable graph is adaptable. In [22] the second author and Yasuhara showed that any proper minor of $D_5$ is adaptable. We also note that after our work some other planar graphs are shown to be non-adaptable in [22].

**Proof.** Let $v_{C_5} : \text{SE}(C_5) \to Z$ be an invariant defined by $v_{C_5}(f) = a_2(f(C_5))$ for each element $f \in \text{SE}(C_5)$ where $a_2(f(C_5))$ denotes the second coefficient of the Conway polynomial of the knot $f(C_5)$. Then $v_{C_5}$ is a Vassiliev type invariant of order at most 2. Since $2 < \frac{5}{2}$ we have that $v = v(\{v_\gamma\}, \omega) : \text{SE}(D_5) \to Z$ is the zero map. On the other hand $a_2(\text{trefoil knot}) = 1$. Therefore a collection of knots, one trefoil knot by a 5-cycle and the others trivial, cannot be realized. This completes the proof. □

Let $f \in \text{SE}_i(G)$ and $c$ a self-crossing vertex of $f$ on an edge $e$ of $G$. Let $e'$ be a subarc of $e$ whose end points are $f^{-1}(c)$. We say that $c$ is immediate if $f(e')$ contains no crossing vertices of $f$ except $c$. We say that $c$ is nugatory
if $c$ is immediate and $f(e')$ bounds a disk in $R^3$ whose interior is disjoint from $f(G)$.

Let $f \in SE_i(G)$ and $c$ a crossing vertex of $f$ of two adjacent edges $e$ and $d$ of $G$. Let $p$ be a common vertex of $e$ and $d$. Let $e'$ and $d'$ be subarcs of $e$ and $d$ respectively whose end points are $f^{-1}(c)$ and $p$. We say that $c$ is immediate if $f(e' \cup d')$ contains no crossing vertices of $f$ except $c$. We say that $c$ is nugatory if $c$ is immediate and $f(e' \cup d')$ bounds a disk in $R^3$ whose interior is disjoint from $f(G)$.

We note that if $c$ is a nugatory self-crossing vertex or a nugatory crossing vertex of adjacent edges then the positive resolution and the negative resolution of $f$ at $c$ are mutually ambient isotopic. Therefore if $f \in SE_i(G)$ has a nugatory self-crossing vertex or a nugatory crossing vertex of adjacent edges then $v_i(f) = 0$.

**Proof of Theorem 3.1 (2).** Let $C$ be an $i$-configuration on $G(e_1, \ldots, e_m)$ with $i \leq n$. Suppose that $C$ has a pair of points $p_1, p_2$ on an edge $e$ of $G(e_1, \ldots, e_m)$. If there are no other points of $C$ on $e$ between $p_1$ and $p_2$, then there is a realization $f_C \in SE_i(G(e_1, \ldots, e_m))$ of $C$ that has a nugatory self-crossing vertex. Therefore we have $v_i(f_C) = 0$. If there are some other points of $C$ on $e$ then we have by the condition $i \leq n < (m+2)/2$ that there is a pair of multiple edges $e_l$ and $d_l$ that have no points of $C$. Let $f_C \in SE_i(G(e_1, \ldots, e_m))$ be a realization of $C$ such that $e_l$ and $d_l$ are parallel in $f_C$. Then we have $v_i(f_C) = 0$. Thus we have established the condition (3) of Theorem B. □

Now the proof of Theorem 3.1 (3) is quite similar and we omit it.

Let $\Lambda$ be a subset of $\{1, \ldots, m\}$. Then we say that an $i$-singular embedding $f$ of $G(e_1, \ldots, e_m)$ is locally parallel with respect to $\Lambda$ if $f(\bigcup_{l \in \Lambda} (e_l \cup d_l))$ contains no crossing vertices of $f$ and for each $l \in \Lambda$ $e_l$ and $d_l$ are parallel in the restriction map $f|_{\bigcup_{l \in \Lambda} (e_l \cup d_l)}$.

Similarly we say that an $i$-singular embedding $f$ of $G(e_1, \ldots, e_m)$ is locally parallel up to edge-homotopy (resp. vertex-homotopy) with respect to $\Lambda$ if $f(\bigcup_{l \in \Lambda} (e_l \cup d_l))$ contains no crossing vertices of $f$ and the restriction map $f|_{\bigcup_{l \in \Lambda} (e_l \cup d_l)}$ is edge-homotopic (resp. vertex-homotopic) to an element $h$ of $SE(\bigcup_{l \in \Lambda} (e_l \cup d_l))$ such that for each $l \in \Lambda$ $e_l$ and $d_l$ are parallel in $h$.

The following three theorems are applications of Theorem D.

**Theorem 3.3.** Let $\Lambda$ be a subset of $\{1, \ldots, m\}$ with $|\Lambda| = k$. If $n \leq k - 1$ and $f \in SE(G(e_1, \ldots, e_m))$ is locally parallel with respect to $\Lambda$ then $v(f) = v(\{v_\gamma\}, \omega)(f) = 0$.

**Theorem 3.4.** Let $k$ be a natural number with $k \leq m - 1$. Let $f$ and $g$ be elements of $SE(G(e_1, \ldots, e_m))$ each of which is locally parallel up to edge-homotopy with respect to any $\Lambda \subset \{1, \ldots, m\}$ with $|\Lambda| = k$. Suppose that $n \leq k$ and $f$ and $g$ are edge-homotopic. Then $v(f) = v(\{v_\gamma\}, \omega)(f) = v(g) = v(\{v_\gamma\}, \omega)(g)$. 

Theorem 3.5. Let $k$ be a natural number with $k \leq m - 2$. Let $f$ and $g$ be elements of $\text{SE}(G(e_1, \cdots, e_m))$ each of which is locally parallel up to vertex-homotopy with respect to any $\Lambda \subset \{1, \cdots, m\}$ with $|\Lambda| = k$. Suppose that $n < \frac{2}{3}(k + 2)$ and $f$ and $g$ are vertex-homotopic. Then $v(f) = v(\{v_\gamma\}, \omega)(f) = v(g) = v(\{v_\gamma\}, \omega)(g)$.

Note that in many situations these theorems actually present sharper results on the order $n$ of $v$ than that in Theorem 3.1. We also note that an edge-homotopy invariant for $D_3$ defined in [18] is an example of Theorem 3.4 where $k = n = 2$, $R = \mathbb{Z}/2\mathbb{Z}$ and $v_\gamma$ is the Arf invariant of knots. In general it follows from Theorem 3.4 that the spatial graphs in Figure 1.2 are not edge-homotopic to plane graphs. See Example 3.6.

Proof of Theorem 3.3. For $1 \leq i \leq k$ let $T_i(G(e_1, \cdots, e_m))$ be the $i$-singular embeddings of $G(e_1, \cdots, e_m)$ that is locally parallel with respect to some $\Delta \subset \Lambda$ with $|\Delta| = k - i$. Let $g \in T_i(G(e_1, \cdots, e_m))$. Suppose that $g$ is locally parallel with respect to $\Delta$ and $l \in \Delta$. Let $D$ be a disk that bounds $g(e_l \cup d_l)$. If the interior of $D$ is disjoint from $g(G(e_1, \cdots, e_m))$ then we have $v_i(g) = 0$ as in the proof of Theorem 3.1 (1). If not then we remove the intersection by a sequence of crossing changes. Let $h \in T_i(G(e_1, \cdots, e_m))$ be the result of the crossing changes. Then we have that $e_l$ and $d_l$ are parallel in $h$ and $v_i(h) = 0$. Moreover it is clear that the $(i + 1)$-singular embeddings corresponding to the crossing changes are contained in $T_{i+1}(G(e_1, \cdots, e_m))$. Therefore we have by Theorem D that $v_1(T_1(G(e_1, \cdots, e_m))) = 0$. Suppose that $f \in \text{SE}(G(e_1, \cdots, e_m))$ is locally parallel with respect to $\Lambda$. Then similarly there is a sequence of crossing changes from $f$ to some $u \in T(G(e_1, \cdots, e_m))$ with $v(u) = 0$ whose corresponding 1-singular embeddings are contained in $T_1(G(e_1, \cdots, e_m))$. Therefore we have $v(f) - v(u) = 0$ and $v(f) = 0$. □

Proof of Theorem 3.4. It is sufficient to show that the values of $v_1$ of the 1-singular embeddings corresponding to the edge-homotopy from $f$ to $g$ are zero. For $1 \leq i \leq k$ let $T_i(G(e_1, \cdots, e_m))$ be the $i$-singular embeddings of $G(e_1, \cdots, e_m)$ each of which is locally parallel up to edge-homotopy with respect to some $\Delta$ with $|\Delta| = k - i + 1$. Let $h \in T_i(G(e_1, \cdots, e_m))$. Suppose that $h$ is locally parallel up to edge-homotopy with respect to $\Delta$ and $l \in \Delta$. Then we choose a sequence of self-crossing changes from $h$ to an element $u \in T_i(G(e_1, \cdots, e_m))$ that is locally parallel with respect to $\Delta$ and then choose a sequence of crossing changes from $u$ to an element $w \in T_i(G(e_1, \cdots, e_m))$ in which $e_l$ and $d_l$ are parallel. Then we have that $v_i(w) = 0$ and all the corresponding $(i + 1)$-singular embeddings are contained in $T_{i+1}(G(e_1, \cdots, e_m))$. Then we have the result by Theorem D. □

Proof of Theorem 3.5. It is sufficient to show that the values of $v_1$ of the 1-singular embeddings corresponding to the vertex-homotopy from $f$ to $g$ are
zero. For \(1 \leq i \leq n\) let \(T_i(G(e_1, \ldots, e_m))\) be the \(i\)-singular embeddings of \(G(e_1, \ldots, e_m)\) each of which is locally parallel up to vertex-homotopy with respect to some \(\Delta\) with \(|\Delta| = s \geq 0\) and has at least \(t \geq 0\) immediate crossing vertices of adjacent edges for some \(s\) and \(t\) with \(2s + \frac{1}{3}t > n - i\).

Let \(h \in T_i(G(e_1, \ldots, e_m))\).

Case 1. \(h\) is locally parallel up to vertex-homotopy with respect to some \(\Delta \neq \emptyset\).

Let \(|\Delta| = s\). Then \(h\) has at least \(t\) immediate crossing vertices of adjacent edges and \(\frac{2}{3}s + \frac{1}{3}t > n - i\). We choose a sequence of crossing changes of adjacent edges from \(h\) to an element \(u\) in \(T_i(G(e_1, \ldots, e_m))\) that is locally parallel with respect to \(\Delta\). We note that each of the corresponding \((i + 1)\)-singular embeddings is locally parallel with respect to some (possibly empty) subset \(\Delta'\) of \(\Delta\) with \(|\Delta'| \geq s - 2\) and has at least \(t + 1\) immediate crossing vertices of adjacent edges. Therefore it is contained in \(T_{i+1}(G(e_1, \ldots, e_m))\).

Next we choose a sequence of crossing changes from \(u\) to an element \(w\) in \(T_i(G(e_1, \ldots, e_m))\) such that \(e_l\) and \(d_l\) are parallel in \(w\) for some \(l \in \Delta\). Then we have \(v_i(w) = 0\). We note that each of the \((i + 1)\)-singular embeddings corresponding to the crossing changes from \(h\) to \(w\) is locally parallel with respect to \(\Delta - \{l\}\) and has at least \(t - 1\) immediate crossing vertices of adjacent edges. Therefore they are contained in \(T_{i+1}(G(e_1, \ldots, e_m))\).

Case 2. \(h\) is not locally parallel up to vertex-homotopy with respect to any \(\Delta \neq \emptyset\).

Then we have that \(h\) has at least \(3(n - i) + 1\) immediate crossing vertices of adjacent edges. Then we choose a sequence of crossing changes from \(h\) to \(u\) that has a nugatory crossing vertex of adjacent edges. Then \(v_i(u) = 0\) and each of the corresponding \((i + 1)\)-singular embeddings still has at least \(3(n - i) + 1 - 2 > 3(n - (i + 1)) + 1\) immediate crossing vertices of adjacent edges. Therefore they are contained in \(T_{i+1}(G(e_1, \ldots, e_m))\).

Thus by Theorem D we have the result.

\[\square\]

**Example 3.6.** Let \(J_m\) be the knot illustrated in Figure 3.1.

By a calculation we have that the Jones polynomial of \(J_m\)

\[V_{J_m}(t) = 1 + (1 + t^{-2})^{m-3}(t^2 + t + 1)(1 - t^{-1})^{m-1}.\]

Then by a calculation we have that the \((m - 1)\)-th derivative of \(V_{J_m}(t)\) evaluated at 1 \(V_{J_m}^{(m-1)}(1) = 3 \cdot (m - 1)! \cdot 2^{m-3} \neq 0\). It is well known that \(V_{J}^{(m-1)}(1)\) is an order \(m - 1\) Vassiliev type invariant of a knot \(J\). We note that each of the spatial graphs in Figure 1.2 contains only one nontrivial knot \(J_m\). Therefore by Theorem 3.4 we have that the spatial graphs in Figure 1.2 are not edge-homotopic to plane graphs. We remark that they are vertex-homotopic to plane graphs.
References


**Department of Intelligence and Computer Science**
**Nagoya Institute of Technology**
**Gokiso, Showa-ku**
**Nagoya, 466-8555**
**JAPAN**

E-mail address: ohyama@math.kyy.nitech.ac.jp

**Department of Mathematics**
**Tokyo Woman’s Christian University**
**Zempukuji 2-6-1, Suginamiku**
**Tokyo 167-8585**
**JAPAN**

E-mail address: taniyama@twcu.ac.jp
SPLITTING FIELDS OF G-VARIETIES

ZINOVY REICHSTEIN AND BORIS YOUSSIN

Let $G$ be an algebraic group, $X$ a generically free $G$-variety, and $K = k(X)^G$. A field extension $L$ of $K$ is called a splitting field of $X$ if the image of the class of $X$ under the natural map $H^1(K, G) \to H^1(L, G)$ is trivial. If $L/K$ is a (finite) Galois extension then $\text{Gal}(L/K)$ is called a splitting group of $X$.

We prove a lower bound on the size of a splitting field of $X$ in terms of fixed points of nontoral abelian subgroups of $G$. A similar result holds for splitting groups. We give a number of applications, including a new construction of noncrossed product division algebras.

Contents

1. Introduction 207
2. $G$-varieties 211
3. Two interpretations of $H^1$ 216
4. Splitting fields 218
5. Examples 221
6. Splitting groups 226
7. Abelian subgroups of $\text{PGL}_n$ 229
8. Symplectic modules and division algebras 232
9. Constructing a $G$-variety with prescribed stabilizers 237
References 247

1. Introduction.

Let $k$ be an algebraically closed field of characteristic zero, let $K$ be a finitely generated field extension of $k$ and let $G$ is an algebraic group defined over $k$. Recall that elements of the nonabelian cohomology set $H^1(K, G)$ can be identified with (birational classes of) generically free $G$-varieties $X$ such that $k(X)^G = K$ (see [Po, Section 1.3]). The set $H^1(K, G)$ has no group structure in general; however, $H^1(K, G)$ is equipped with a marked element, which we shall denote by 1. This element is represented by the “split” $G$-variety $X \simeq X_0 \times G$, where $k(X_0) = K$ and $G$ acts by left multiplication on
the second factor. A field extension $L/K$ is said to be a splitting field for $u \in H^1(K, G)$ if $u \mapsto 1$ under the natural map $H^1(K, G) \longrightarrow H^1(L, G)$.

The nonabelian cohomology set $H^1(K, G)$ often allows a different interpretation: Its elements can be identified with certain algebraic objects defined over $K$, e.g., quadratic forms if $G = O_n$, central simple algebras if $G = \text{PGL}_n$, Cayley algebras, if $G = G_2$, etc. These objects may be viewed as “twisted forms” of a single “split” object. In such cases the above notion of a splitting field coincides with the usual one. We will review this interpretation of $H^1(K, G)$ in Section 3; see also [Se4, Chapter III], [Se2, Chapter X], [KMRT, Section 29] or [Re, Sections 6-8].

Recall that a subgroup of $G$ is called toral if it lies in a torus in $G$. Our main results on splitting fields are Theorems 1.1 and 1.2.

**Theorem 1.1.** Let $X$ be a generically free primitive $G$-variety, $K = k(X)^G$, and let $L/K$ be a splitting field for $X$. Suppose $X$ has a smooth point fixed by a finite abelian $p$-subgroup $H$ of $G$. Then $[L : K]$ is divisible by $[H : H_T]$ for some toral subgroup $H_T$ of $H$.

If $X$ is a generically free primitive $G$-variety, $K = k(X)^G$, and $L$ is a splitting field which is a (finite) Galois extension of $K$, then we shall refer to $\text{Gal}(L/K)$ as a splitting group for $X$.

**Theorem 1.2.** Let $X$ be a generically free primitive $G$-variety and let $A$ be a splitting group for $X$. Suppose $X$ has a smooth point fixed by a finite abelian subgroup $H$ of $G$. Then $A$ contains an isomorphic copy of $H/H_T$ for some toral subgroup $H_T$ of $H$.

Our proofs of Theorems 1.1 and 1.2 are based on the following results of [RY]: For any finite abelian subgroup $H$ of $G$, the existence of a smooth $H$-fixed point on a (complete smooth) $G$-variety $X$ is a birational invariant of $X$. Moreover, such points survive under dominant rational $G$-equivariant maps and under certain $G$-equivariant covers; see [RY, Section 5 and Appendix]. We review and further extend these results in Section 2; see Proposition 2.2 and Theorems 2.5, 2.6 and 2.7.

Informally speaking, Theorem 1.1 (respectively, Theorem 1.2) may be viewed as a “lower bound” on a splitting field (respectively, a splitting group) of $X$. In particular, if $X$ is a vector space and $G$ acts linearly on $X$ then $X$ has a smooth $G$-fixed point (namely, the origin) and, hence, in this case Theorem 1.1 (respectively, Theorem 1.2) can be applied to every finite abelian $p$-subgroup (respectively, subgroup) $H$ of $G$. Of course, Theorems 1.1 and 1.2 are only of interest if $H$ is nontoral, since otherwise $H/H_T$ may be trivial.

Elementary finite abelian subgroups of algebraic groups have been extensively studied (see [BS], [Bor1], [St], [Se5]); a complete classification was obtained by Griess [Gri]. To the best of our knowledge, nonelementary
finite abelian subgroups have not been classified. In Section 5 we apply Theorem 1.1 to a number of specific groups $G$, where we have sufficient information about the depth of certain nontoral subgroups (see Definition 4.5). In particular, for $G = E_8$ we give a new proof of a theorem of Serre; see Corollary 5.5. Note, however, that the examples we give in Section 5 are somewhat fragmentary, because we do not know any general results about the depth of finite abelian subgroups in exceptional groups. (Propositions 5.3 and 5.7 represent our best efforts in this direction; see also Corollary 4.10.) We hope that this question will attract the attention of group theorists in the future, and that a more complete picture will emerge.

“Upper bounds” on the degrees of splitting fields, i.e., results of the form “every $G$-variety can be split by a field extension of degree dividing $n(G)$”, can be found in the paper [T2] of Tits. For a discussion of these results, including a table of values for $n(G)$, see Remark 4.9.

In Sections 8 and 9 we apply Theorem 1.2, with $G = \text{PGL}_n$, to the theory of central simple algebras. Recall that an element $\alpha \in H^1(K, \text{PGL}_n)$ may be (functorially) identified with an $n^2$-dimensional central simple $K$-algebra $D_\alpha$; see Example 3.1. In particular, $L/K$ is a splitting field for $\alpha$ if and only if $L$ is a splitting field for $D_\alpha$, i.e., $D_\alpha \otimes_K L = M_n(L)$. Recall that $D$ is an $H$-crossed product iff $H$ is a splitting group for $D$ and $|H| = \deg(D)$.

Let $\text{UD}(n,k)$ be the universal division algebra of degree $n$, i.e., the division algebra generated by two generic matrices, $X = (x_{ij})$ and $Y = (y_{ij})$, in $M_n(k(x_{ij}, y_{ij}))$. Here $x_{ij}, y_{ij}$ are algebraically independent commuting variables over $k$. If the reference to $k$ is clear from the context, we shall write $\text{UD}(n)$ in place of $\text{UD}(n,k)$. A famous theorem of Amitsur asserts that $\text{UD}(n)$ is not a crossed product if $n$ is divisible by $p^3$ for some prime $p$.

As an application of Theorem 1.2 we will prove the following result.

**Theorem 1.3.** Let $Z(p^r)$ be the center of the universal division algebra $\text{UD}(p^r)$, let $K$ be a field extension of $Z(p^r)$ and let $D = \text{UD}(p^r) \otimes_{Z(p^r)} K$. Suppose $p^e$ is the highest power of $p$ dividing $[K : Z(p^r)]$, where $e$ is a nonnegative integer and $e \leq r - 1$. If $A$ is a splitting group for $D$ then

$$p^{2e-2e-2} | |A|.$$ 

In particular, if $r \geq 2e + 3$ then $D$ is a noncrossed product.

If $K = Z(p^r)$, i.e., $D = \text{UD}(p^r)$, we recover a theorem of Amitsur and Tignol; see [TA1, Theorem 7.3]. If $e = 0$, i.e., $D$ is a prime-to-$p$ extension of $\text{UD}(p^r)$, we recover a theorem of Rowen and Saltman [RS, Theorem 2.1] to the effect that $D$ is not a crossed product for any $r \geq 3$.

Abelian subgroups of $\text{PGL}_n$ carry a natural skew-symmetric form and their nontoral subgroups are isotropic with respect to this form; see Section 7. Thus symplectic modules and their Lagrangian submodules, used by Tignol and Amitsur to prove [TA1, Theorem 7.3], naturally arise in
our setting; in particular, they will be used in the proof of Theorem 1.3 in Section 8.

It is likely that Theorem 1.3 can also be proved by an application of Amit-
sur’s specialization technique, along the lines of [RS, Section 2] and that
such a proof will go through in prime characteristic (assuming $p / \text{char}(k)$).
We believe that our approach, based on the fixed points of nontoral sub-
groups, is of independent interest; in particular, it shows that Theorem 1.3
remains true if UD($p^r$) is replaced by any central simple algebra whose cor-
responding PGL$_n$-variety has points fixed by certain nontoral subgroups of
PGL$_n$; see Remark 8.4.

As another application of Theorem 1.2 with $G = \text{PGL}_n$, we construct
a noncrossed product division algebra over a “small” function field. Since
the time of Amitsur’s original examples, two other noncrossed product con-
structions have appeared in the literature, due, respectively, to Jacob—
Wadsworth [JW] and Brussel [Br]. Both of these examples have the prop-
erty that their centers are “smaller” and easier to describe than the center
of Amitsur’s “generic” example, UD($p^r, k$).

The problem we address here is one of constructing noncrossed product
examples over “small” fields in the geometric setting, i.e., noncrossed prod-
ucts $D$ with center $K$ such that $K$ is a function field over an algebraically
closed base field $k$ of characteristic 0. Moreover, we would like “the size of
$K$”, as measured by $\text{trdeg}_k(K)$, to be as small as possible.

Note that $\text{trdeg}_k(K)$ cannot be $\leq 1$ by Tsen’s theorem. Moreover, division
algebras $D$ with $\text{trdeg}_k(K) = 2$ are conjectured to be cyclic. At the other
extreme, if $D = \text{UD}(n)$ is Amitsur’s original noncrossed product example
(with $n$ divisible by $p^3$ for some prime $p$) then $\text{trdeg}_k(K) = n^2 + 1$.

In this paper we prove the following theorem.

**Theorem 1.4.** Let $p$ be a prime, $r \geq 2$ be an integer, and $k$ be an alge-
braically closed field of characteristic 0. Then there exists a division algebra
$D$ of degree $p^r$ with center $K$, such that;

(a) $K$ is a finitely generated extension of $k$ of transcendence degree 6 and;
(b) no prime-to-$p$ extension of $D$ is a crossed product.

The idea of the proof is as follows. We show (see Section 8) that it is
easy to construct smooth PGL$_{p^r}$-variety $X$ with two points whose stabilizers are “incompatible” symplectic modules $(\mathbb{Z}/p^r\mathbb{Z})^2$ and $(\mathbb{Z}/p^r\mathbb{Z})^6$;
such varieties are fairly easy to construct. The difficult part is to reduce
the dimension of $X/\text{PGL}_{p^r}$ to 6; this is done in Section 9. Our argument
there is based on a resolution result for the fixed point loci of finite abelian
subgroups (we show that the fixed-point set of a finite abelian subgroup $H$
can be resolved in such way that it has a component of the minimal possible
codimension, equal to rank $H$; see Theorem 9.3) and on a form of Bertini’s
theorem in the equivariant setting (Theorem 9.7). We believe Theorems 9.3 and 9.7 are of independent interest.

A.R. Wadsworth has pointed out to us that Theorem 1.4 can be proved by modifying the arguments of \([JW]\). This approach, based on valuation theory, cohomology, and the Merkurjev-Suslin theorem, yields the desired result under the assumption that \(p \not| \text{char}(k)\).

Throughout this paper we shall work over a fixed base field \(k\) which will be assumed to be algebraically closed and of characteristic zero. The assumption that \(k\) should be algebraically closed is usually not essential: Generally speaking, the problems we wish to consider (such as constructing noncrossed products or proving lower estimates on the size of splitting fields) can only become harder after passing to the algebraic closure. The characteristic zero assumption is more serious, since most of our proofs ultimately rely on canonical resolution of singularities (via Proposition 2.2 and Theorem 2.5).

Acknowledgements. We would like to thank D.J. Saltman, J.-P. Serre, and A.R. Wadsworth for helpful discussions.

2. \(G\)-varieties.

Preliminaries. A \(G\)-variety \(X\) is an algebraic variety with a \(G\)-action. Here \(G\), \(X\) and all other algebraic objects in this paper are assumed to be defined over a fixed base field \(k\). Unless otherwise specified, we shall assume that \(k\) is algebraically closed and of characteristic 0. The \(G\)-action on \(X\) is given by a morphism \(G \times X \rightarrow X\). If the reference to the action is clear from the context, we shall write \(gx\) for the image of \((g, x)\) under this map. Given \(x \in X\), the stabilizer of \(x\) is defined as \(\{g \in G \mid gx = x\}\); we will denote this subgroup of \(G\) by \(\text{Stab}_G(x)\) or simply \(\text{Stab}(x)\) if the reference to \(G\) is clear from the context.

By a morphism \(X \rightarrow Y\) of \(G\)-varieties, we shall mean a \(G\)-equivariant morphism from \(X\) to \(Y\). The same goes for rational morphism, isomorphism, birational morphism, etc., of \(G\)-varieties.

A \(G\)-variety \(X\) is called primitive if \(G\) transitively permutes the irreducible components of \(X\). Equivalently, \(X\) is primitive iff \(k(X)^G\) is a field. Note that an irreducible \(G\)-variety is necessarily primitive and that the converse holds if \(G\) is a connected group.

If \(X\) is a \(G\)-variety then any variety \(Y\) with \(k(Y) = k(X)^G\) is called a rational quotient variety for \(X\); we will often write \(Y = X/G\). Note that \(X/G\) is only defined up to birational isomorphism and that \(X\) is a primitive \(G\)-variety iff \(X/G\) is irreducible. The inclusion \(k(Y) = k(X)^G \hookrightarrow k(X)\) induces the rational quotient map \(\pi: X \dashrightarrow X/G\). By a theorem of Rosenlicht \(\pi^{-1}(y)\) is a single \(G\)-orbit for \(y\) in general position in \(X/G\) (see \([Ro1, \text{Theorem 2}], [Ro2]\)).
A $G$-variety $X$ is called generically free if $\operatorname{Stab}(x) = \{1\}$ for $x$ in general position in $X$. We will usually consider $G$-varieties that are both primitive and generically free. Up to birational isomorphism, a primitive generically free $G$-variety may be viewed as a principal $G$-bundle over $X/G$ and thus represents a class in $H^1(K,G)$, where $K = k(X)^G$; see [Po, Section 1.3]. We shall return to this connection in Section 3.

It is often convenient to have a concrete (biregular) model for $X/G$. If $G$ is a finite group then, under rather mild assumptions on $X$, we have such a model in the form of a geometric quotient, which we shall denote by $X//G$. Here the quotient map $\pi : X \to X//G$ is regular and each fiber of this map is a single $G$-orbit. For a precise definition and a detailed discussion of the geometric quotient we refer the reader to [PV, Section 4.2].

Lemma 2.1. Let $G$ be a finite group and $X$ be a normal quasiprojective $G$-variety. Then:

(a) $X$ is covered by affine open $G$-invariant subsets.
(b) There exists a geometric quotient map $\pi : X \to X//G$.
(c) Moreover, if $X$ is projective then so is $X//G$.

Proof. (a) By a theorem of Kambayashi, we may assume without loss of generality that $X \subset \mathbb{P}(V)$, where $V$ is a finite-dimensional vector space, and $G$-acts linearly on $X$, via a representation $G \to \text{GL}(V)$; see [Ka, Theorem 2.5] or [PV, Theorem 1.7].

We want to show that every $x \in X$ has an affine $G$-invariant neighborhood in $X$. To construct this neighborhood, choose a homogeneous polynomial $h \in k[V]$ such that $h(gx) \neq 0$ for every $g \in G$ but $h(y) = 0$ for every $y \in X - X$. After replacing $h$ by the product of $g^*h$ over all $g \in G$, we may assume $h$ is $G$-invariant. Now $\{z \in X | h(z) \neq 0\}$ is a desired affine $G$-invariant neighborhood of $x$.

(b) Follows from part (a) and [PV, Theorem 4.14].
(c) See [PV, Theorem 4.16].

The variety $X_L$. Let $X$ be a generically free primitive $G$-variety, let $K = k(X)^G$ and let $cl(X)$ be the class of $X$ in $H^1(K,G)$. Suppose $L$ is a finitely generated field extension of $K$. Then $X_L$ is defined as the $G$-variety representing the image of $cl(X)$ under the natural map $H^1(K,G) \to H^1(L,G)$. In other words, $cl(X) \mapsto cl(X_L)$ under this map.

To construct $X_L$ explicitly, let $Y \to X/G$ be a rational map such that $k(Y)/k(X/G)$ is precisely the extension $L/K$. Note that such a rational map exists because $L$ is finitely generated over $K$ and, hence, over $k$. Now we set $X_L = Y \times_{X/G} X$, where the $G$-action on $X_L$ is induced from the $G$-action on $X$; cf. [Re, Section 2.6].

We emphasize that $X_L$ is only defined up to birational isomorphism (of $G$-varieties). We will often want to work with a specific model for $X_L$ which...
is smooth or projective or has “small” stabilizers (or all of the above). The existence of such models is guaranteed by Proposition 2.2 and Theorem 2.5.

**Smooth projective models for G-varieties.**

**Proposition 2.2.** Every G-variety is birationally isomorphic to a smooth projective G-variety.

*Proof.* Let \( X \) be a G-variety. By [RY, Proposition 7.1], \( X \) is birationally isomorphic to a complete G-variety. (Note that the proof of [RY, Proposition 7.1] is based on Sumihiro’s equivariant completion theorem.) Thus we may assume without loss of generality that \( X \) is complete.

Now by [Ka, Theorem 2.5] there exists a projective representation \( G \hookrightarrow \operatorname{PGL}(V) \) and a closed G-invariant subvariety \( X' \) of \( \mathbb{P}(V) \) such that \( X \) and \( X' \) are birationally isomorphic as G-varieties. After replacing \( X \) by \( X' \), we may assume \( X \) is projective. Now apply the canonical resolution of singularities theorem (see either [V, Theorem 7.6.1] or [BM, Theorem 13.2]) to \( X \) to construct a smooth projective model. \( \square \)

**Definition 2.3.** We shall call an algebraic group \( H \) Levi-commutative if \( H \) is a semidirect product of a diagonalizable group \( D \) and a unipotent group \( U \), where \( U \triangleleft H \) is the unipotent radical of the identity component \( H_0 \) of \( H \).

We shall denote \( U = R_u(H_0) \).

**Lemma 2.4.** Let \( H \) be an algebraic group and let \( H_0 \) be the identity component of \( H \). The following conditions are equivalent.

(i) \( H \) is Levi-commutative,
(ii) \( H/R_u(H_0) \) is commutative,
(iii) every reductive subgroup of \( H \) is commutative, and
(iv) every linear representation of \( H \) has 1-dimensional \( H \)-invariant subspace.

*Proof.* The equivalence of (i), (ii) and (iii) follows from the Levi decomposition theorem (see [OV, Section 6.4]). The equivalence of (i) and (iv) is proved in [RY, Lemma A.1]. \( \square \)

**Theorem 2.5.** Let \( X \) be a G-variety. Then there exists a sequence of blowups

\[
X_n \rightarrow X_{n-1} \rightarrow \ldots \rightarrow X_0 = X
\]

such that \( X_n \) is smooth and \( \operatorname{Stab}(x) \) is Levi-commutative for every \( x \in X_n \).

*Proof.* See [RY, Theorem 1.1]. \( \square \)
Going up and going down.

**Theorem 2.6** (Going down). *Let $H$ be a Levi-commutative group (see Definition 2.3) and let $X \to Y$ be a rational map of $H$-varieties. Suppose $Y$ is complete and $X$ has a smooth point fixed by $H$. Then $Y$ has a smooth point fixed by $H$."

Proof. See [RY, Propositions 5.3 and A.2]. □

**Theorem 2.7** (Going up). *Let $H$ be a finite abelian group of prime power order $p^n$ and let $f : X \to Y$ be a dominant rational map of $H$-varieties. Suppose

(i) $Y$ is irreducible,

(ii) $X$ is complete,

(iii) $\dim(X) = \dim(Y)$,

(iv) $H$ has a smooth fixed point in $Y$, and

(v) $\deg(X/Y)$ is not divisible by $p^{e+1}$."

Then there exists a subgroup $H'$ of $H$ such that $|H'| \geq p^{n-e}$ and

(a) $H'$ has a fixed point in $X$. 

(b) Moreover, if $X \to Z$ is a rational map of complete $H$-varieties then $H'$ has a fixed point in $Z$."

This theorem is a generalization of [RY, Propositions 5.5 and A.4], where $e$ is assumed to be 0. Our proof below is based on an argument of Kollár and Szabó; cf. [RY, Proposition A.4]. Our applications will only use (a); however, part (b) is needed for the inductive argument.

A convenient way to visualize the setting of Theorem 2.7 is by means of the diagram

$$
\begin{array}{ccc}
X & \xrightarrow{h} & Z \\
\downarrow f & & \downarrow \\
Y & & \\
\end{array}
$$

If $e = 0$ the theorem allows us to lift an $H$-fixed point from $Y$ to $X$, then transport it to $Z$. (A similar but slightly weaker statement is true if $e \geq 1$.) We will make use of such diagrams in the proof.

Proof. The proof is by induction on $d = \dim Y$. If $\dim Y = 0$ then $Y$ is a point, $X$ is a set of $\deg(X/Y)$ points, and the desired result follows from a simple counting argument.

To perform the induction step, we assume that the theorem holds whenever $\dim(Y) \leq d - 1$. Let $y \in Y$ be a smooth fixed point, $B_y(Y)$ be the blowup of $Y$ at $y$, and $Y' \subset B_y Y$ be the exceptional divisor. Note that $H$ acts linearly on $Y' = \mathbb{P}^{\dim(Y)-1}$ and, hence, has a fixed point in $Y'$; see
Lemma 2.4(iv). This fixed point will be smooth because every point of $Y'$ is smooth.

Let $X = \bigcup_i X_i$ be the decomposition of $X$ into irreducible components. It is enough to find the required fixed point in one of the components $X_i$ which is mapped dominantly onto $Y$; thus, we may assume that all $X_i$ are mapped dominantly onto $Y$.

It follows that each map $X_i \rightarrow B_\mathcal{Y}$ is dominant; let $X_i$ be the normalization of $B_\mathcal{Y}$ in the field of rational functions on $X_i$, $\overline{X}$ be the disjoint union of all $X_i$ (in other words, $\overline{X}$ is the normalization of $B_\mathcal{Y}$ in the ring $k(\overline{X}) = \bigoplus_i k(\overline{X}_i)$), and $\overline{f} : \overline{X} \rightarrow B_\mathcal{Y}$ be the natural morphism. Clearly, $H$ acts on $\overline{X}$ and $\overline{f}$ is $H$-equivariant. Each $\overline{X}_i$, it is birationally isomorphic to $X_i$. Together these birational isomorphisms yield an $H$-equivariant birational isomorphism $\overline{X} \rightarrow X$.

Let $F_1, F_2, \ldots \subset \overline{X}$ be the divisors lying over $Y'$. Note that even though $X$ is not necessarily complete, each $F_i$ is complete since it is mapped finitely to a complete variety $Y'$.

The group $H$ acts on the set $\{F_i\}$. Let $F_j$ denote the $H$-orbits in $\{F_i\}$. Choose a divisor $F^*_j \in F_j$ in each orbit. By the ramification formula (see, e.g., [L, Corollary XII.6.2]),

$$\deg(X/Y) = \sum_j |F_j| \cdot \deg(F^*_j/E) \cdot e(\overline{f}, F^*_j),$$

where $e(\overline{f}, F^*_j)$ denotes the ramification index of $\overline{f}$ at the generic point of $F^*_j$. Since $\deg(X/Y)$ is not divisible by $p^{e+1}$, $|F_j| \cdot \deg(F^*_j/E)$ is not divisible by $p^{e+1}$ for some $j$. For this $j$, set $X' = HF^*_j = \bigcup_{F_i \in F_j} F_i$; this variety is complete since each $F_i$ is complete. Let $f' : X' \rightarrow Y'$ be the restriction of $\overline{f} : \overline{X} \rightarrow B_\mathcal{Y}$ to $X'$; the degree of $f'$ is equal to $|F_j| \cdot \deg(F^*_j/E)$, and hence, is not divisible by $p^{e+1}$. Let $h' : X' \rightarrow X$ be the restriction of $\overline{X} \rightarrow X$ to $X'$. Note that $h'$ is well-defined, since $\overline{X}$ is normal, $X$ is complete, and $X'$ is a divisor in $\overline{X}$.

By our construction, $\dim(Y') = d - 1$ and conditions (i)–(v) hold for the map $f' : X' \rightarrow Y'$. Applying the induction assumption to the diagram

$$
\begin{array}{ccc}
X' & \xrightarrow{h'} & X \\
| & & | \\
f' | & & | \\
\downarrow & & \downarrow \\
Y'
\end{array}
$$
we prove part (a). Applying the induction assumption to the diagram

\[
\begin{array}{c}
X' \xrightarrow{h'} X \xrightarrow{h} Z \\
\downarrow \quad \downarrow \\
f' \quad Y'
\end{array}
\]

we prove part (b).

\[\square\]

**Remark 2.8.** Theorems 2.6 and 2.7 are valid over an algebraically closed field of arbitrary characteristic; the proofs given above are characteristic-free.

3. Two interpretations of $H^1$.

Let $A$ be a finite-dimensional algebra over $k$. We do not assume that $A$ is commutative, associative or has an identity element. Let $K$ is a field extension of $k$. We shall say that a $K$-algebra $B$ is of type $A$ if $B_K \simeq A_K$, where $K$ is the algebraic closure of $K$. (Here, as usual, $B_K \overset{\text{def}}{=} B \otimes K_K$ and $A_K \overset{\text{def}}{=} A \otimes K_K$.)

Let $G = \text{Aut}_k(A)$. It is easy to see that $G$ is a closed subgroup of $\text{GL}_{\dim(A)}$; thus it is an algebraic group. We now have the following bijections.

\[
H^1(K, G) \leftrightarrow \left\{ \begin{array}{l}
\text{Birational isomorphism classes} \\
\text{of primitive generically free} \\
G\text{-varieties } X \text{ with } k(X)^G = K
\end{array} \right\}
\]

\[
\begin{array}{c}
\uparrow \\
\downarrow \\
\{K\text{-algebras of type } A\}
\end{array}
\]

The horizontal correspondence is described in [Po, Section 1.3]; the vertical one in [Se2, Section X.2] or [KMRT, Proposition 29.1]. Note that the above diagram is functorial in $K$ and that the correspondences in it are bijections of pointed sets: The identity elements of $H^1(K, G)$ corresponds to the “split” algebra $A_K$ and to the “split” variety $X = X_0 \times G$, where $k(X_0)^G = K$; see Definition 4.1.

In this paper we shall be primarily interested in passing back and forth between $G$-varieties and algebras of type $A$. In other words, we would like to construct an explicit correspondence $X \mapsto B$ which completes the triangle in the above diagram. Given a generically free primitive $G$-variety $X$ with $k(X)^G = K$, we define $B = RMaps_G(X, A)$. Here $RMaps_G(X, A)$ is the set of $G$-equivariant rational maps $X \dashrightarrow A$, where we view $A$ as a $k$-vector
space with a $G$-action. The $k$-algebra structure on $A$ gives rise to a $K$-algebra structure on $B$ via

$$
(f_1 + f_2)(x) \overset{\text{def}}{=} f_1(x) + f_2(x),
$$
$$
(f_1 f_2)(x) \overset{\text{def}}{=} f_1(x)f_2(x),
$$
$$
(\alpha f_2)(x) \overset{\text{def}}{=} \alpha(x)f_1(x).
$$

Here $f_1, f_2 \in RMaps_G(X, A)$, $\alpha \in K$, $x$ is an element of $X$ in general position, and the operations in the right hand sides of the above formulas are performed in $A$. The correspondence $X \mapsto RMaps_G(X, A)$ completes the triangle in the above diagram; see [Re, Proposition 8.6 and Lemma 12.3].

**Example 3.1.** $A = M_n(k)$ is the algebra of $n \times n$-matrices over $k$. By a theorem of Wedderburn, $B$ is a $K$-algebra of type $M_n$ if and only if $B$ is a central simple $K$-algebra; see e.g., [KMRT, Theorem 1.1] or [Se2, X.5, Proposition 7].

Note that $\text{Aut}_k(A) = \text{PGL}_n$. Thus, if $K$ is a finitely generated field extension of $k$, every central simple $K$-algebra $B$ of degree $n$ is of the form $RMaps_{\text{PGL}_n}(X, M_n(k))$ for some generically free irreducible $\text{PGL}_n$-variety $X$ with $k(X)^G = K$. Moreover, the $G$-variety $X$ is uniquely determined up to birational isomorphism. For example, if $B = \text{UD}(n)$ is the universal division algebra of degree $n$ generated by two generic matrices then $B = RMaps_{\text{PGL}_n}(X, M_n(k))$, where $X = M_n(k) \times M_n(k)$ and $\text{PGL}_n$ acts on $X$ by simultaneous conjugation. This description of $\text{UD}(n)$ is due to Procesi; see [Sa, Theorem 14.16] or [Pr, Theorem 2.1].

**Example 3.2.** $A = \mathbb{O}$ is the 8-dimensional split Cayley algebra (otherwise known as the split octonion algebra). Then $\text{Aut}(\mathbb{O})$ is the exceptional group $G_2$. By a theorem of Zorn, $B$ is a $K$-algebra of type $\mathbb{O}$ if and only if $B$ is a Cayley algebra over $K$. Cayley algebras are thus in natural 1—1 correspondence with generically free irreducible $G_2$-varieties; see [Re, Remark 11.4], [Se3, Section 8.1] and [KMRT, Proposition 33.24].

**Example 3.3.** $A$ is the 27-dimensional (split) Albert algebra (otherwise known as an exceptional simple Jordan algebra) defined over $k$. Then $\text{Aut}(A)$ is the exceptional group $F_4$. Algebras of type $A$ are precisely the Albert algebras, i.e., a 27-dimensional exceptional simple Jordan algebras; see e.g., [KMRT, p. 517], [Se3, Section 9].

**Remark 3.4.** The results of this section remain valid if the algebra $A$ is replaced by a more general algebraic object consisting of a vector space with a tensor on it. Such objects are called structured spaces in [Re]. We refer the reader there for details; see also [KMRT, Section 29].
4. Splitting fields.

Definition 4.1. Let $G$ be an algebraic group and $X$ be a primitive generically free $G$-variety, $K = k(X)^G$ and $cl(X) = \text{the class of } X \text{ in } H^1(K, G)$. We will say that $X$ is split if $X$ is birationally isomorphic to $X/G \times G$ (as a $G$-variety). Equivalently, $X$ is split if there exists a rational section $X/G \rightarrow X$ or, if $cl(X) = 1$; see, e.g., [Po, 1.4.1].

A field extension $L$ of $K$ is called a splitting field of $X$ if $X_L$ is split. Equivalently, $L$ is a splitting field if the image of $cl(X)$ under the natural map $H^1(K, G) \rightarrow H^1(L, G)$ is trivial.

Remark 4.2. Recall that an algebraic group $G$ is called special if every generically free $G$-variety is split. Special groups were studied by Serre [Se1] and classified by Grothendieck [Gro, Theorem 3]; see also [PV, Theorem 2.8]. In particular, $\text{GL}_n$, $\text{SL}_n$, $\text{Sp}_{2n}$, and the additive group $G_a$ are special. Moreover, it is easy to see that if $N$ is a normal subgroup of $G$ and both $N$, $G/N$ are special, then so is $G$. In particular, every connected solvable group is special; cf. [PV, Section 2.6].

Split $G$-varieties.

Lemma 4.3. Let $G$ be an algebraic group, let $X$ be a split $G$-variety and let $H$ be a Levi-commutative subgroup of $G$ (see Definition 2.3). If $H$ has a smooth fixed point in $X$ then

(a) $H$ is contained in a Borel subgroup of $G$. Moreover,

(b) if $H$ is diagonalizable then it is contained in a maximal torus of $G$.

Proof. Since $X$ is split, it is birationally isomorphic to $X_0 \times G$ for some variety $X_0$, where $G$ acts trivially on the first factor and by left translations on the second. Let $B$ be a Borel subgroup of $G$. Consider the rational $G$-equivariant map

$$X \simeq X_0 \times G \rightarrow G/B,$$

sending $(x_0, g)$ to $g \text{ mod } B$. Since $G/B$ is a complete variety, the Going Down Theorem 2.6 tells us that $H$ fixes a point of $G/B$. Consequently, $H$ is contained in a conjugate of $B$, and part (a) follows. Part (b) is immediate from part (a) and [Bor2, Theorem 10.6(5)].

Remark 4.4. Note that our proof of Lemma 4.3 relies only on the Going Down Theorem and not on Proposition 2.2. In particular, Lemma 4.3(a) is valid in arbitrary characteristic; see Remark 2.8.

Proof of Theorem 1.1. Consider the natural projection map $X_L \rightarrow X$ of $G$-varieties. Here $L$ is a splitting field for $X$, i.e., $X_L$ is split. By Proposition 2.2, we may assume without loss of generality that $X_L$ is smooth and projective. Let $p^e$ be the maximal power of $p$ dividing $[L : K] = \deg(X_L/X)$. 
By the Going Up Theorem 2.7(a), there exists a point \( y \in X_L \) and a subgroup \( H' \) of \( H \) such that (i) \([H : H']\) divides \( p^e\) and (ii) \( H' \subset \text{Stab}(y)\).

Now (i) says that \([H : H']\) divides \([L : K]\) and (ii), in combination with Lemma 4.3(b), says that \( H' \) is toral. \(\square\)

**Non-toral subgroups.**

**Definition 4.5.** Let \( G \) be an algebraic group and let \( H \) be an abelian \( p \)-subgroup of \( G \). The **depth** of \( H \) is defined as the smallest integer \( i \) such that \( H \) has a toral subgroup of index \( p^i \).

Recall that a prime number \( p \) is called a **torsion prime** for \( G \) if \( G \) has a non-toral abelian \( p \)-subgroup \( H \), i.e., an abelian \( p \)-subgroup \( H \) of depth \( \geq 1 \). Following [Se5, 1.3] we will denote the set of torsion primes for \( G \) by \( \text{Tors}(G) \).

**Remark 4.6.** Torsion primes have been extensively studied; see [Bor1], [SS], [St] and [Se5]. In particular, \( \text{Tors}(G) = \text{Tors}(G') \) if \( G' \) is a derived subgroup of \( G \); see [Se5, 1.3.2]. If \( f: \overline{G} \to G \) is the universal cover of \( G \) then \( \text{Tors}(G) \) is the union of \( \text{Tors}(\overline{G}) \) and the set of prime divisors of \( \text{Ker}(f) \); see [Se5, 1.3.3].

For simply connected simple groups the torsion primes are given by the following table:

<table>
<thead>
<tr>
<th>Type of ( G )</th>
<th>( \text{Tors}(G) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_n ) or ( C_n )</td>
<td>( \emptyset )</td>
</tr>
<tr>
<td>( B_n ) (( n \geq 3 )), ( D_n ) (( n \geq 4 )) or ( G_2 )</td>
<td>( {2} )</td>
</tr>
<tr>
<td>( F_4 ), ( E_6 ), or ( E_7 )</td>
<td>( {2, 3} )</td>
</tr>
<tr>
<td>( E_8 )</td>
<td>( {2, 3, 5} )</td>
</tr>
</tbody>
</table>

For details see [Bor1, Proposition 4.4], [St, Corollary 1.13] or [Se5, 1.3.3].

Using the terminology of Definition 4.5, we can rephrase Theorem 1.1 as follows.

**Theorem 4.7.** Let \( G \) be an algebraic group, \( H \) be a finite abelian \( p \)-subgroup of \( G \) of depth \( d \), \( X \) be a generically free \( G \)-variety and \( K = k(X)^G \). If \( X \) has a smooth \( H \)-fixed point and \( L/K \) is a splitting field of \( X \) then \( p^d \mid \left[ L : K \right] \).

In particular, if \( X = V \) is a generically free linear representation of \( G \) then \( [L : K] \) is divisible by every torsion prime of \( G \).

Let \( G \) be an algebraic group and let \( S \) be a special subgroup containing \( G \). (For example, \( S \) can be taken to be \( \text{GL}_n \), \( \text{SL}_n \) or \( \text{Sp}_{2n} \).) We shall view \( S \) as a \( G \)-variety with respect to the left multiplication action; it is easy to see that this variety is generically free.
Corollary 4.8. Let $G$ be an algebraic group, $H$ be a finite abelian $p$-subgroup of $G$ of depth $d$ and $S$ be a special group containing $G$, as above. Suppose $K = k(S)^G$ and $L/K$ is a splitting field for $S$ (as a $G$-variety). Then $p^d | [L : K]$.

Proof. Let $\overline{S}$ be a smooth projective model of $S$ (as an $S$-variety); see Proposition 2.2. Let $V$ be a generically free linear representation of $S$. Since $S$ is special, $V$ is split as an $S$-variety. Thus there exists an $S$-equivariant dominant rational map $f : V \to S \cong \overline{S}$. Since $G \subset S$, we can view $f : V \to \overline{S}$ as a dominant rational map of $G$-varieties. Since $V$ has a $H$-fixed point, the Going Down Theorem 2.6 tells us that so does $\overline{S}$. Applying Theorem 4.7 to the smooth variety $X = \overline{S}$, we conclude that $p^d | [L : K]$, as claimed. □

Remark 4.9. Let $G$ be a simple group. A theorem of Tits asserts that every $G$-variety $X$ can be split by an extension $L/K$, where $K = k(X)^G$ and every prime factor of $[L : K]$ lies in $\text{Tors}(G)$; see [Se3, 2.3]. This gives a partial converse to Theorem 4.7.

More precisely, the results of [T2] show that every $G$-variety can be split by an extension of degree dividing $n(G)$, where $n(G)$ is given by the following table.

<table>
<thead>
<tr>
<th>Type</th>
<th>Simply Connected</th>
<th>Not Simply Connected</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_n$</td>
<td>1</td>
<td>$n + 1$</td>
</tr>
<tr>
<td>$B_n$</td>
<td>$2^{\sup(1, n-4)}$</td>
<td>$2^n$</td>
</tr>
<tr>
<td>$C_n$</td>
<td>1</td>
<td>$2^{v_2(n)+1}$</td>
</tr>
<tr>
<td>$D_n$</td>
<td>$2^{\sup(1, n-5)}$</td>
<td>$2^{v_2(n)+n}$</td>
</tr>
<tr>
<td>$G_2$</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>$F_4$</td>
<td>6</td>
<td>–</td>
</tr>
<tr>
<td>$E_6$</td>
<td>6</td>
<td>$2 \cdot 3^4$</td>
</tr>
<tr>
<td>$E_7$</td>
<td>12</td>
<td>$2^5 \cdot 3$</td>
</tr>
<tr>
<td>$E_8$</td>
<td>$2^7 \cdot 3^3 \cdot 5$</td>
<td>–</td>
</tr>
</tbody>
</table>

Here $G$ is an almost simple group of the indicated type (recall that $G$ is almost simple if the center $Z(G)$ is finite and $G/Z(G)$ is simple) and $v_2(m)$ denotes the highest power of 2 dividing $m$. 
Note that the terminology of \[ T2 \] is somewhat different from ours. A primitive \( G \)-variety \( X \) corresponds to a group of inner type over \( k(X)^G \) (and if \( G \) is simply connected, then of strongly inner type). With these conventions, the entries for all group types other than \( E_8 \) come directly from \[ T2, \text{Proposition A1} \] or from \[ T2, \text{Propositions 1 and 2} \].

Our entry for \( E_8 \) follows from \[ T2, \text{Corollaire 2} \] and the fact that \( n(G) \) may be taken to be the degree of a splitting field for one particular “generic” \( E_8 \)-variety; see \[ T1, \text{Proposition 8} \]. (Recall that we are working in characteristic zero.) In fact \[ T2, \text{Corollaire 2} \] implies that \( n(E_8) \) may be taken to be one of the numbers \( 2^7 \cdot 3 \cdot 5 \), \( 2^6 \cdot 3^2 \cdot 5 \) or \( 2^4 \cdot 3^3 \cdot 5 \) (it is not currently known which one). The entry for \( n(E_8) \) in our table is the least common multiple of these three numbers.

Combining the above-mentioned results of \[ T2 \] with Theorem 4.7, we obtain the following upper bound on the depth of abelian \( p \)-subgroups of quasi-simple algebraic groups.

**Corollary 4.10.** Let \( G \) be an almost simple algebraic group and let \( H \) be a finite abelian \( p \)-subgroup of \( G \) of depth \( d \) (not necessarily elementary). Then \( p^d \) divides the number \( n(G) \) given in Remark 4.9.

## 5. Examples.

In this section we illustrate Theorem 1.1 for several classes of groups. The application of Theorem 1.1 to the case \( G = \text{PGL}_n \) will come up later, after we discuss the nontoral subgroups of \( \text{PGL}_n \) in Section 7; see Lemma 8.1 below.

**Orthogonal groups: splitting fields of quadratic forms.** Let \( K \) be a finitely generated field extension of \( k \). Recall that quadratic forms \( q \) over \( K \) are in 1—1 correspondence with primitive generically free \( O_n \)-varieties \( X \) such that \( k(X)^{O_n} = K \); see \[ \text{Se4, III. Appendix 2.2} \] or \[ \text{KMRT, Section 29E} \]. In particular, a field extension \( L/K \) splits the form if and only if it splits the corresponding variety.

**Proposition 5.1.** Let \( q = \langle a_1, \ldots, a_n \rangle \) be a quadratic form over \( K \). Then

(a) there exists a splitting field \( L \) for \( q \) such that \( L/K \) is a Galois extension with \( \text{Gal}(L/K) = (\mathbb{Z}/2\mathbb{Z})^l \) for some \( l \leq \left\lceil \frac{n+1}{2} \right\rceil \).

(b) Suppose \( a_1, \ldots, a_n \) are algebraically independent variables over \( k \), \( K = k(a_1, \ldots, a_n) \) and \( L/K \) is a splitting field for \( q \). Then \( [L : K] \) is divisible by \( 2^{\lceil \frac{n+1}{2} \rceil} \).

**Proof.** (a) Suppose \( n = 2m \) is even. Let

\[
L = K \left( \sqrt{-a_2/a_1}, \ldots, \sqrt{-a_{2m}/a_{2m-1}} \right).
\]
Note that $L$ is a Galois extension of $K$ and $\text{Gal}(L/K) = (\mathbb{Z}/2\mathbb{Z})^l$ for some $l \leq m$. Moreover, since $\langle a_{2i-1}, a_{2i} \rangle \simeq \langle a_{2i-1}, -a_{2i-1} \rangle$ for $i = 1, \ldots, m$,

$$q \simeq \langle a_1, -a_1 \rangle \oplus \cdots \oplus \langle a_{2m-1}, -a_{2m-1} \rangle \simeq 0$$

in the Witt ring of $L$. This shows that $q$ splits over $L$.

If $n = 2m + 1$ is odd then a similar argument shows that

$$L = k \left( \sqrt{-a_2/a_1}, \ldots, \sqrt{-a_{2m}/a_{2m-1}}, \sqrt{a_{2m+1}} \right)$$

is a splitting field for $q$ with $\text{Gal}(L/K) = (\mathbb{Z}/2\mathbb{Z})^l$ and $l \leq m + 1$.

(b) Consider the $O_n$-variety $X = M_n$, with $O_n$ acting (linearly) on $X$ by left multiplication. Recall that $k(X)^{O_n} = k(b_{ij} \mid 1 \leq i \leq j \leq n)$, where $b_{ij}(x)$ is the dot product of the $i$th and the $j$th columns of the matrix $x \in M_n$; see, e.g., [DC, Section 2.10] or [Re, Lemma 6.4]. Note that since $\dim(X/O_n) = \dim(X) - \dim(O_n) = \frac{n(n+1)}{2}$, the generators $b_{ij}$ are algebraically independent over $k$. The quadratic form corresponding to $X$ is the “generic form” $q_X = \sum_{i \leq j} b_{ij} x_i x_j$ defined over $k(b_{ij})$. Let $Y$ be the subvariety of $X$ consisting of $n \times n$-matrices with mutually orthogonal columns. Then $Y$ is irreducible (see [Re, Example 3.10]) and the quadratic form corresponding to $Y$ is the “generic diagonal” form $q = \sum_{i=1}^n a_i x_i^2$ which appears in the statement of part (b). Here $a_i = b_{ii}$ and $q$ is defined over $K = k(a_1, \ldots, a_n)$. We can now view the usual orthogonalization process in $k^n$ as an $O_n$-equivariant rational map $f : X \to Y$. That is, we view a matrix $x \in M_n$ as a collection of $n$ column vectors. To construct $f(x) \in Y$, we apply the orthogonalization process to this collection; the resulting $n$ mutually orthogonal vectors form the columns of $f(x)$ (see [Re, Example 3.10] for details).

Note that the point $0_{n \times n}$ is a smooth point of $X$ fixed by all of $O_n$; here $0_{n \times n}$ is the $n \times n$ zero matrix. Let $Y'$ be a smooth projective model of $Y$ (see Proposition 2.2); we can thus think of $f$ as an $O_n$-equivariant rational map $X \dasharrow Y'$. Let $H \simeq (\mathbb{Z}/2\mathbb{Z})^n$ be the diagonal subgroup of $O_n$. By the Going Down Theorem 2.6, $Y'$ has an $H$-fixed point. This point is smooth because every point of $Y'$ is smooth. Since $L$ splits $q$, it splits the $O_n$-variety $Y$ or, equivalently, the $O_n$-variety $Y'$. Thus Theorem 4.7 tells us that $2^d \mid [L : K]$, where $d$ is the depth of $H$. Since the dimension of any maximal torus in $O_n$ is $[\frac{n}{2}]$, $d \geq n - [\frac{n}{2}] = [\frac{n+1}{2}]$, as claimed. \hfill \Box

The same argument with the group $SO_n$ in place of $O_n$ yields the following variant of Proposition 5.1. Note that elements $H^1(K, SO_n)$ represent equivalence classes of quadratic forms of determinant 1; cf. [Re, Example 8.4(b)] or [KMRT, (29.29)].

**Proposition 5.2.** Let $K = k(a_1, \ldots, a_n)$, $q = \langle a_1, \ldots, a_n \rangle$ be a quadratic form of determinant 1 over $K$. Then:
(a) There exists a splitting field $L$ for $q$ such that $L/K$ is a Galois extension with $\text{Gal}(L/K) = (\mathbb{Z}/2\mathbb{Z})^l$ for some $l \leq \lceil n/2 \rceil$.

(b) Suppose $a_1, \ldots, a_{n-1}$ are algebraically independent variables over $k$, $a_n = (a_1 \ldots a_{n-1})^{-1}$, and $L/K$ is a splitting field for $q$. Then $[L : K] \geq 2^{n-1}$.

Exceptional groups $G_2$, $F_4$, $3E_6$ and $2E_7$. Let $V$ be a generically free linear representation of $G$, let $K = k(V)^G$ and let $L/K$ be a splitting field of $V$. Theorem 4.7 tells us that $[L : K]$ is divisible by 2 if $G = G_2$ and by 6 if $G = F_4$, $3E_6$ or $2E_7$. (Here $3E_6$ and $2E_7$ denote the simply connected groups of type $E_6$ and $E_7$ respectively.) If $G = G_2$, $F_4$ or $3E_6$ then this result is sharp. In other words, $V$ can be split by $L/K$ such that $[L : K]$ equals 2, if $G = G_2$ and 6, if $G = F_4$ or $3E_6$; see Remark 4.9.

Exceptional group $E_8$. Recall that by a theorem of Adams [Ad] $E_8$ has two maximal elementary abelian 2-subgroups (up to conjugacy): $D(T)$ of rank 9 and $EC^8$ of rank 8. Here we are following the notational conventions of [Ad, Section 2]; in particular, $D(T)$ means “double 2-torus” and $EC^8$ means “exotic candidate of rank 8”. By construction $D(T)$ has depth 1.

Proposition 5.3. The subgroup $EC^8 \simeq (\mathbb{Z}/2\mathbb{Z})^8$ of $E_8$ has depth 2.

Our proof of this proposition uses the theory of quadratic forms over $\mathbb{Z}/2\mathbb{Z}$. Recall that if $q$ is a quadratic form on $V = (\mathbb{Z}/2\mathbb{Z})^m$, the associated symmetric (or, equivalently, skew-symmetric) bilinear form $b_q : V \times V \rightarrow V$ is defined by $b_q(v, w) = q(v + w) - q(v) - q(w)$. Note that the “usual” relationship between $q$ and $b_q$ breaks down in characteristic 2: In particular, $b_q(v, v) = 0$ for any $v \in V$. The kernel of $b_q$ is called the radical of $q$.

We shall say that $v \in V$ is an anisotropic vector for $q$ if $q(v) = 1$ and an isotropic vector if $q(v) = 0$. In the sequel we shall be interested in counting the number of anisotropic vectors for a given form $q$. This is not a difficult task (at least in principle) because $q$ can always be written as a direct sum of quadratic forms of dimension 1 and 2 (see, e.g., [Pf, Theorem 1.4.3]), and if $q = r \oplus s$ then a simple counting argument shows that

$$|q^{-1}(0)| = |r^{-1}(0)| \cdot |s^{-1}(0)| + |r^{-1}(1)| \cdot |s^{-1}(1)|$$

(5.1)

$$|q^{-1}(1)| = |r^{-1}(1)| \cdot |s^{-1}(0)| + |r^{-1}(0)| \cdot |s^{-1}(1)|.$$

Lemma 5.4. Let $q$ be a quadratic form on $V = (\mathbb{Z}/2\mathbb{Z})^7$. Suppose the radical of $q$ has dimension 1. Then $q$ has 56, 64 or 72 anisotropic vectors in $V$.

Proof. Write $q \simeq q_1 \oplus q_2 \oplus q_3 \oplus \langle e \rangle$, where $q_1$, $q_2$ and $q_3$ are regular 2-dimensional quadratic forms and $e = 0$ or 1; see, e.g., [Pf, Theorem 1.4.3]. (Here, the 1-dimensional form $\langle e \rangle$ is the radical of $q$.) Note that over $\mathbb{Z}/2\mathbb{Z}$
there are only two classes of regular 2-dimensional quadratic forms: The hyperbolic form $h$ given by $h(x_1, x_2) = x_1 x_2$ and the anisotropic form $a(x_1, x_2) = x_1^2 + x_2^2 + x_1 x_2$. Since $a \oplus a \simeq h \oplus h$ (see [Pf, Example 2.4.5]), we may assume without loss of generality that $q_2 = q_3 = h$.

**Case 1.** $e = 1$. Using (5.1) it is easy to see that if $q_0$ is any quadratic form on $(\mathbb{Z}/2\mathbb{Z})^m$ then $q = q_0 \oplus \langle 1 \rangle$ has exactly $2^m$ anisotropic vectors in $(\mathbb{Z}/2\mathbb{Z})^{m+1}$. In our situation $q_0 = q_1 \oplus q_2 \oplus q_3$ and $m = 6$; thus we conclude that $|q^{-1}(1)| = 64$. From now on we shall assume that $e = 0$.

**Case 2.** $q = h \oplus h \oplus h \oplus \langle 0 \rangle$. We apply (5.1) to this form recursively. Since $|h^{-1}(0)| = 3$ and $|h^{-1}(1)| = 1$, we obtain $|q^{-1}(1)| = 56$.

**Case 3.** $q = a \oplus h \oplus h \oplus \langle 0 \rangle$. We note that $|a^{-1}(0)| = 1$ and $|a^{-1}(1)| = 3$, and apply (5.1) recursively, to conclude that $|q^{-1}(1)| = 72$.

This completes the proof of the lemma. 

**Proof of Proposition 5.3.** Recall that $EC^8 = A_1 \times A_2 \subset G_2 \times F_4 \subset E_8$, where $A_1 = (\mathbb{Z}/2\mathbb{Z})^3$ is the unique (up to conjugacy) nontoral abelian 2-subgroup of $G_2$ and $A_2$ is the unique (again, up to conjugacy) nontoral abelian 2-subgroup of $F_4$; see [Gri, Theorem 2.17]. Thus, $A_1$ has a subgroup of index 2 which is toral in $G_2$, and $A_2$ has a subgroup of index 2 which is toral in $F_4$. Taking a direct product of these toral subgroups, we construct a subgroup of $EC^8$ of index 4 which is toral in $G_2 \times F_4$ and, hence, in $E_8$. This proves that the depth of $EC^8$ is $\leq 2$.

It remains to show that the depth of $EC^8$ is $\geq 2$. Recall that elements of $E_8$ of order 2 fall into two conjugacy classes: class $A$ and class $B$; cf. [Ad, Section 5] or [Gri, (2.14)]. If $T$ is a maximal torus in $E_8$ and $T(2) = \{ t \in T : t^2 = 1 \}$ then we have a naturally defined $W_{E_8}$-invariant quadratic form $q$ on $T(2) \simeq (\mathbb{Z}/2\mathbb{Z})^8$; see [Gri, Definition 2.15]. By [Gri, Lemma 2.16] this form is nonsingular and has maximal Witt index; moreover, an element $x$ of $T(2)$ is of type $A$ in $E_8$ if $q(x) = 1$ and of type $B$ if $q(x) = 0$.

In particular, of the 255 nonidentity elements of $T(2)$, 120 are of type $A$ and 135 are of type $B$. On the other hand, of the 255 nonidentity elements of $EC^8$, 56 are of type $A$ and 199 are of type $B$; see [Ad, Section 5].

We now proceed to prove that the depth of $EC^8$ is $\geq 2$. Assume, to the contrary, that $EC^8$ has a toral subgroup $U$ of rank 7. Since $q$ is nonsingular (i.e., the associated symplectic form $b_q$ is nondegenerate) on $T(2)$, the radical of $q|_U$ is of dimension $\leq 1$. On the other hand, since dim$(U)$ is odd, the radical of $q|_U$ cannot be trivial; thus it has dimension exactly 1. By Lemma 5.4, $q$ has at least 56 anisotropic vectors in $U$, i.e., $U$ has at least 56 elements of type $A$. On the other hand, $EC^8$ has exactly 56 elements of type $A$. We therefore conclude that every element of type $A$ lies in $U$. We claim that this is impossible because the elements of type $A$ generate $EC^8$. This contradiction will complete the proof of the proposition.
To prove the claim, recall that $E_8 = A_1 \times A_2$, where $A_1 \simeq (\mathbb{Z}/2\mathbb{Z})^3$ lies in $G_2$ and $A_2 = (\mathbb{Z}/2\mathbb{Z})^5$ lies in $F_4$, as above. Moreover, $A_2$ has a subgroup $R$ of order 4 (called the radical of $E_8$) such that

$$S = (A_2 - R) \cup (A_1 R - R)$$

is precisely the set of elements of $E_8$ of type $A$; see [Gri, Theorem 2.17]. We want to show that $\langle S \rangle = E_8$. Indeed, $A_2 - R$ contains 28 of the 32 elements of $A_2$; these elements clearly generate all of $A_2$. Thus $A_2 \subset \langle S \rangle$. In particular, $R \subset \langle S \rangle$. Now $R$, together with $A_1 R$ generate $A_1$. We thus conclude that both $A_1$ and $A_2$ lie in $\langle S \rangle$. This proves that $E_8 = A_1 \times A_2 = \langle S \rangle$, as claimed. □

We are now ready to give an alternative proof of a theorem of Serre.

**Corollary 5.5** (Serre, see [T1, Proposition 9, p. 30] or [T2, p. 1132]). Suppose $E_8 \hookrightarrow S$, where $S = \text{GL}_n$, $\text{SL}_n$ or $\text{Sp}_{2n}$ for some $n$. We shall view $S$ as an $E_8$-variety via the left multiplication action. Suppose $K = k(S)^{E_8}$ and $L/K$ is a splitting field of $S$. Then $[L : K]$ is divisible by 60.

**Proof.** Recall 2 and 3 are torsion primes of $E_8$, i.e., $E_8$ has an abelian 3-subgroup and an abelian 5-subgroup, both of depth $\geq 1$. Moreover, by Proposition 5.3, $E_8$ contains an abelian 2-subgroup of depth 2. Thus Corollary 4.8 tells us that $[L : K]$ is divisible by $2^2 \cdot 3 \cdot 5 = 60$. □

**Exceptional group $E_7$ (adjoint).** We will now show that the (adjoint) group $E_7$ has an elementary abelian 2-subgroup of depth $\geq 2$. We begin with the following lemma.

**Lemma 5.6.** Let $f : G \rightarrow G'$ be a surjective homomorphism of algebraic groups, such that $\text{Ker}(f)$ is special. Suppose $H$ is a finite abelian subgroup of $G$ and $f(H)$ is toral in $G'$. Then $H$ is toral in $G$.

**Proof.** Suppose $f(H) \subset T' \subset G'$, where $T'$ be a torus in $G'$. Denote $f^{-1}(T')$ by $S$. Then $H \subset S \subset G$. Moreover, since both $\text{Ker}(f)$ and $S/\text{Ker}(f) \simeq T'$ are special, we conclude that $S$ is special as well; see Remark 4.2. This means that $H$ is toral in $S$ (see e.g., [Se5, 1.5.1] or Example 6.6); hence, $H$ is toral in $G$. □

**Proposition 5.7.** The (adjoint) group $E_7$ has an elementary abelian 2-subgroup of depth $\geq 2$.

Our proof uses the idea of Adams (see [Ad, Introduction]) to study non-toral 2-subgroups in groups of type $E_7$ by embedding $2E_7$ into $E_8$.

**Proof.** Let $E_8$ be a maximal elementary abelian subgroup of $E_8$ of rank 8, as in Proposition 5.3. As we mentioned in the proof of that proposition, $E_8$ has 56 elements of type $A$ (in $E_8$). Let $x$ be one of these 56 elements. Denote
the centralizer $C_{E_8}(x)$ by $C$. Note that $E_8^C \subset C$. Moreover, $C \simeq 2A_1E_7$; see [Gri, p. 280]. Thus there is an exact sequence

$$\{1\} \longrightarrow \text{SL}_2 \longrightarrow C \xrightarrow{f} E_7 \longrightarrow \{1\}.$$

We claim that $f(E_8^C)$ has depth $\geq 2$ in $E_7$. Indeed, assume the contrary. Then $f(E_8^C)$ contains a subgroup $H'$ of index 2 which is toral in $E_7$. By Lemma 5.6, $H = f^{-1}(H') \cap E_8^C$ is toral in $C$ and thus in $E_8$. Since $H$ is a toral subgroup of index 2 in $E_8^C$, this implies that $E_8^C$ has depth $\leq 1$, contradicting Proposition 5.3. \hfill $\Box$

We can now prove an analogue of Corollary 5.5 for $E_7$.

**Corollary 5.8.** Suppose $E_7 \hookrightarrow S$, where $S = \text{GL}_n$, $\text{SL}_n$ or $\text{Sp}_{2n}$ for some $n$. We shall view $S$ as an $E_7$-variety via the left multiplication action. Suppose $K = k(S)^{E_7}$ and $L/K$ is a splitting field of $S$. Then $[L : K]$ is divisible by 12.

**Proof.** Recall $E_7$ has a nontoral abelian 3-subgroup, i.e., a 3-subgroup of depth $\geq 1$; see, e.g., [Gri]. Moreover, by Proposition 5.7, $E_7$ contains an abelian 2-subgroup of depth $\geq 2$. Thus Corollary 4.8 tells us that $[L : K]$ is divisible by $2^3 : 3 = 12$. \hfill $\square$

### 6. Splitting groups.

**Definition and first examples.**

**Definition 6.1.** Let $X$ be a generically free primitive $G$-variety and let $K = k(X)^G$. We shall say that a finite group $A$ is a *splitting group* for $X$ if there exists a splitting field $L$ for $X$ such that $L/K$ is (finite) Galois and $\text{Gal}(L/K) = A$.

**Example 6.2.** Let $G$ be a finite group and let $X$ be a generically free irreducible $G$-variety. Then $G$ is a splitting group for $X$.

**Proof.** Suppose $L = k(X)$ and $K = k(X)^G$. Then $k(X_L) = L \otimes_K L = L \oplus \cdots \oplus L$ ($|G| = [L : K]$ times). In other words, up to birational equivalence, $X_L$ is the disjoint union of $|G|$ copies of $X$ and $G$ acts on $X_L$ by permuting these copies. Consequently, $X_L$ is split as a $G$-variety and $G = \text{Gal}(L/K)$ is a splitting group. \hfill $\Box$

**Example 6.3.** Let $G$ be a (connected) semisimple group, and let $W$ be the Weyl group of $G$. Then every irreducible generically free $G$-variety $X$ has a splitting group which is isomorphic to a subgroup of $W$.

**Proof.** Let $X$ be a generically free irreducible $G$-variety and let $\pi: X \dashrightarrow X/G$ be the rational quotient map. An irreducible subvariety $S$ of $X$ is called a *Galois section* if $GS$ is dense in $X$, i.e., $\pi|_S$ is dominant, and the
field extension \( k(S)/k(X)^G \) induced by \( \pi \), is a finite Galois extension. We shall denote the group \( \text{Gal}(k(S)/k(X)^G) \) by \( \text{Gal}(S) \).

A theorem of Galitskii asserts that every \( G \)-variety \( X \) has a Galois section \( S \); see [Ga]. Moreover, by [Po, Remark 1.6.3] \( S \) can be chosen so that \( \text{Gal}(S) \) is isomorphic to a subgroup \( H \) of \( W \). It is easy to see that in order to split \( X \) as a \( G \)-variety it is sufficient to split \( S \) as a \( \text{Gal}(S) \)-variety. Now Example 6.2 tells us that \( H \) is a splitting group for \( X \). □

If \( G \) is a connected but not necessarily semisimple then the assertion of Example 6.3 remains true if we define \( W \) as the Weyl group of \( G_{ss} = G/R(G) \), where \( R(G) \) is the radical of \( G \).

**Two elementary lemmas from group theory.** Our next goal is to prove Theorem 1.2. We begin with two elementary lemmas.

**Lemma 6.4.** Let \( P \) be a finite abelian group.

(a) Every quotient group of \( P \) is isomorphic to a subgroup of \( P \).

(b) Every subgroup of \( P \) is isomorphic to a quotient group of \( P \).

**Proof.** Suppose \( Q \) is a quotient group of \( P \). Then every character of \( Q \) lifts to a character of \( P \). This gives an inclusion \( Q^* \hookrightarrow P^* \) of dual groups. Since \( Q^* \cong Q \) and \( P^* \cong P \), part (a) follows. Part (b) is proved in a similar manner. □

**Lemma 6.5.** Suppose \( A \) and \( B \) are (abstract) groups and \( A \times B \) acts on a set \( Z \). Let \( W \) be the set of \( A \)-orbits in \( Z \) and let \( f: Z \longrightarrow W \) be the natural projection. Assume \( z \in Z \) and \( w = f(z) \). Then:

(a) \( \text{Stab}_B(z) \) is a normal subgroup of \( \text{Stab}_B(w) \).

(b) Let \( S = \text{Stab}_{A \times B}(z) \). Then \( \text{Stab}_B(z) \) is normal in \( S \) and \( S/\text{Stab}_B(z) \) is isomorphic to a subgroup of \( A \); we shall denote this subgroup by \( A_0 \).

(c) \( \text{Stab}_B(w)/\text{Stab}_B(z) \) is isomorphic to a quotient of \( A_0 \).

**Proof.** (a) Suppose \( b \in \text{Stab}_B(w) \). Since the actions of \( A \) and \( B \) on \( Z \) commute, \( f(bz) = bf(z) = bw = w \). Consequently, \( bz = az \) for some \( a \in A \) and thus

\[
b\text{Stab}_B(z)b^{-1} = \text{Stab}_B(bz) = \text{Stab}_B(az) = \text{Stab}_B(z).
\]

This proves part (a).

Let \( \pi_A \) and \( \pi_B \) be, respectively, the natural projections \( A \times B \longrightarrow A \) and \( A \times B \longrightarrow B \).

(b) The kernel of the map \( \pi_A: S \longrightarrow A \) is \( S \cap B = \text{Stab}_B(z) \), and part (b) follows.

(c) Note that \( b \in \text{Stab}_B(w) \) if and only if \( bz = az \) for some \( a \in A \) or, equivalently, if \( (a^{-1}, b) \in S \) for some \( a \in A \). In other words, \( \text{Stab}_B(w) = \pi_B(S) \). Consequently, we have a surjective homomorphism

\[
\pi_B: A_0 = S/\text{Stab}_B(z) \longrightarrow \text{Stab}_B(w)/\text{Stab}_B(z).
\]
This completes the proof of part (c).

Proof of Theorem 1.2. Let $K = k(X)^G$ and let $L/K$ be a Galois extension such that $\text{Gal}(L/K) = A$ and $X_L$ is split. Note that $A \times G$ acts rationally on $X_L$. By a theorem of Rosenlicht (see [Ro1, Theorem 1]), we can choose a birational model for $X_L$ so that this action becomes regular. Moreover, after applying Proposition 2.2 and Theorem 2.5 to $X_L$, we may assume that

(i) $X_L$ is smooth and projective, and

(ii) for every $z \in X_L$, $\text{Stab}_{A \times G}(z)$ is Levi-commutative (see Definition 2.3).

Note that by our construction the map $h: X_L \to X$ is a rational quotient map for the $A$-action on $X_L$. Since $A$ is a finite group and $X_L$ is projective, there exists a geometric quotient map $f: X_L \to W = X_L//A$ for the $A$-action on $X_L$ with $W$ projective; see Lemma 2.1. Note that by the universal property of categorical (and, hence, geometric) quotients, the $G$-action on $X_L$ descends to $W$; by our construction, $W$ and $X$ are birationally isomorphic as $G$-varieties. Applying the Going Down Theorem 2.6 to the birational isomorphism $X \cong W$, we conclude that $W$ has a $H$-fixed point. Denote this point by $w$.

Now choose $z \in f^{-1}(w) \subset X_L$ and apply Lemma 6.5 with $Z = X_L$ where we view $Z$ as an $A \times H$-variety via the obvious inclusion of $A \times H$ in $A \times G$. By Lemma 6.5(b), $A$ has a subgroup $A_0 \cong S/\text{Stab}_H(w)$, where $S = \text{Stab}_{A \times H}(z)$. Note that $S$ is a finite subgroup of $\text{Stab}_{A \times G}(z)$, and $\text{Stab}_{A \times G}(z)$ is Levi-commutative by our construction (see condition (ii) above). We conclude that $S$ is abelian; see Lemma 2.4. Thus $A_0$ is also abelian.

By Lemma 6.5(c), $A_0$ has a quotient of the form $S/\text{Stab}_H(z) = H/\text{Stab}_H(z)$. Denote $\text{Stab}_H(z)$ by $H'$. Since $Z = X_L$ is split, Lemma 4.3(b) says that $H'$ is toral in $G$. Thus $H/H'$ is a quotient of $A_0$, with $H'$ toral. By Lemma 6.4, $A_0$ (and, hence, $A$) has a subgroup isomorphic to $H/H'$, as claimed.

Examples.

Example 6.6 (cf. [Se5, 1.5.1]). Let $G$ be a special group (see Remark 4.2). Then every finite abelian subgroup of $G$ is toral.

Indeed, let $V$ be a generically free linear representation of $G$ and let $H$ be a finite abelian subgroup of $G$. Since $G$ is special, $V$ is split, i.e., $A = \{1\}$ is a splitting group for $V$. On the other hand, since the origin of $V$ is a smooth $H$-fixed point, there exists a toral subgroup $H_T$ of $H$ such that $H/H_T$ is isomorphic to a subgroup of $A = \{1\}$. In other words, $H = H_T$ is toral, as claimed.

Example 6.7. Let $a_1, \ldots, a_n$ be independent variables over $k$ and let $q = \langle a_1, \ldots, a_n \rangle$ be the generic quadratic form of dimension $n$. Then any splitting group of $q$ contains a copy of $(\mathbb{Z}/2\mathbb{Z})^{\frac{n+1}{2}}$. 
The proof is the same as in Proposition 5.1(b), with Theorem 1.2 used in place of Theorem 1.1.

**Example 6.8.** Let $G = E_7$ (adjoint) or $E_8$. Suppose $G \hookrightarrow S$, where $S = \text{GL}_n$, $\text{SL}_n$ or $\text{Sp}_{2n}$ for some $n$. Then any splitting group of $S$ (viewed as a $G$-variety with respect to the left multiplication action) contains a copy of $(\mathbb{Z}/2\mathbb{Z})^2$.

Indeed, $G$ has an elementary abelian 2-subgroup $H$ of depth $\geq 2$; see Propositions 5.3 and 5.7. If $X = S$ is a smooth projective model for $S$ (as an $S$-variety) then the argument of Corollary 4.8 shows that $H$ has a fixed point in $X$. Thus we can apply Theorem 1.2 to $X$.

7. **Abelian subgroups of $\text{PGL}_n$.**

The rest of this paper will be devoted to applications of Theorem 1.2 (with $G = \text{PGL}_n$) to the theory of central simple algebras. In this section we lay the foundation for these applications by studying finite abelian subgroups of $\text{PGL}_n$.

**Symplectic modules.** We begin by recalling the notion of a symplectic module from [TA2]. Let $H$ be an abelian group; in the sequel we shall refer to such groups as $\mathbb{Z}$-modules or just modules. We will always assume $H$ is finite. A skew-symmetric form on $H$ is a skew-symmetric $\mathbb{Z}$-bilinear map $\omega: H \times H \to \mathbb{Q}/\mathbb{Z}$. If $H$ is written multiplicatively, we will usually identify $\mathbb{Q}/\mathbb{Z}$ with the multiplicative group of roots of unity in $k^*$. A subgroup (or, equivalently, a $\mathbb{Z}$-submodule) $H'$ of $H$ is called *isotropic* if $\omega(h,h') = 0$ for every $h, h' \in H'$.

We will say that $\omega$ is *symplectic* if it is nondegenerate, i.e., the homomorphism $H \to H^*$ it defines, is an isomorphism. If $\omega$ is symplectic then a subgroup $H' \subset H$ is called Lagrangian if it is a maximal isotropic subgroup, i.e., if $H'$ is not contained in any other isotropic subgroup. It is easy to see that $H'$ is Lagrangian if and only if it is isotropic and $|H'|^2 = |H|$; cf. [TA2, Corollary 3.1].

**Lemma 7.1.** Let $(H, \omega)$ be a symplectic module of order $n^2$.

(a) If $\Lambda$ is a Lagrangian submodule then $H/\Lambda \simeq \Lambda$ (as abelian groups).

(b) Let $H_1$ be a subgroup of $H$, and let $I$ be an isotropic subgroup of $H_1$. Then $H_1/I$ contains an isomorphic copy of $I_1$, where $I_1$ is an isotropic subgroup of $H$ and $|H_1|$ divides $n|I_1|$.

**Proof.** (a) For $h \in H$, let $\chi_h: \Lambda \to k^*$ be the character given by $\chi_h(l) = \omega(h,l)$. Then $h \mapsto \chi_h$ is a group homomorphism $\phi: H \to \Lambda^*$. Since $\omega$ is nondegenerate, $\phi$ is onto. Since $\Lambda$ is Lagrangian, $\text{Ker}(\phi) = \Lambda$. Thus $H/\Lambda \simeq \Lambda^*$. Since $\Lambda^* \simeq \Lambda$, the lemma follows.

(b) We may assume without loss of generality that $I$ is a maximal isotropic subgroup of $H_1$. Indeed, let $I_{\text{max}}$ be a maximal isotropic subgroup of $H_1$...
containing \(I\). Suppose we can find an isotropic subgroup \(I_1\) such that \(|H_1|\) divides \(n|I_1|\) and \(H/I_{\text{max}}\) has a subgroup isomorphic to \(I_1\). Since \(H/I_{\text{max}}\) is isomorphic to a quotient, and hence a subgroup of \(H/I\) (see Lemma 6.4), the same \(I_1\) will work for \(I\).

Thus we may (and will) assume that \(I\) is a maximal isotropic subgroup of \(H_1\). Let \(\Lambda\) be a Lagrangian subgroup of \(H\) containing \(I\). Then \(\Lambda \cap H_1 = I\) and thus \(H_1/I \hookrightarrow H/\Lambda \simeq \Lambda\); the last isomorphism is given by part (a). Denote the image of \(H_1/I\) in \(\Lambda\) by \(I_1\). Then

\[
|H_1| = |I| \cdot |H_1/I| = |I| \cdot |I_1| \quad \text{divides} \quad |\Lambda| \cdot |I_1| = n|I_1|, \quad \text{as claimed.} \tag*{□}
\]

Definition 7.2 (cf. [TA2, Section 4]). Let \(A\) be an abelian group. We define a skew-symmetric form \(\omega_A\) on \(A \times A^*\) by

\[
\omega_1(a_1 \oplus \chi_1, a_2 \oplus \chi_2) = \chi_1(a_2) - \chi_2(a_1).
\]

Lemma 7.3. (a) \((A \times A^*, \omega_A)\) is a symplectic module, and \(A \times \{1\}\) is a Lagrangian submodule.

(b) Moreover, every symplectic module \(H\) is of the form \((A \times A^*, \omega_A)\) for a suitable Lagrangian submodule \(A\) of \(H\).

(c) Let \((H, \omega)\) be a symplectic module, \(H \simeq (\mathbb{Z}/p\mathbb{Z})^{2r}\). If \(s \leq r\) then \((H, \omega)\) has a symplectic submodule \((H_1, \omega_{|H_1})\) of rank \(2s\).

Proof. Parts (a) and (b) are proved in [TA2, Section 4]. Proof of (c): By part (b), we can write \((H, \omega)\) as \((A \times A^*, \omega_A)\), where \(A = (\mathbb{Z}/p\mathbb{Z})^r\). Let \(e_1, \ldots, e_r\) be an \(\mathbb{Z}/p\mathbb{Z}\)-basis of \(A\), viewed as an \(r\)-dimensional vector space over \(\mathbb{Z}/p\mathbb{Z}\). Then the module \(H_1\) spanned by \((e_i, 1)\) and \((1, e_j^*)\) as \(i, j\) range from 1 to \(s\), has the desired properties. \(\Box\)

The form \(\alpha_H\). Abelian subgroups of \(\text{PGL}_n\) are naturally endowed with a skew-symmetric bilinear form.

Definition 7.4. Let \(H\) be a finite abelian subgroup of \(\text{PGL}_n\). For \(a, b \in H\) define \(\alpha_H(a, b) = ABA^{-1}B^{-1}\), where \(A\) and \(B\) are elements of \(\text{GL}_n\) representing \(a\) and \(b\) respectively. It is easy to see that \(\alpha_H(a, b)\) does not depend of the choice of \(A\) and \(B\) and \(\alpha_H: H \times H \rightarrow k^*\) defined this way, is a skew-symmetric form on \(H\).

Lemma 7.5. Let \(H\) be a finite abelian subgroup of \(\text{PGL}_n\). Then the following conditions are equivalent.

(a) \(H\) lifts to an abelian subgroup of \(\text{SL}_n\).

(b) \(H\) is toral.

(c) The skew-symmetric form \(\alpha_H: H \times H \rightarrow k^*\) given in Definition 7.4 is trivial, i.e., \(\alpha_H(a, b) = 1\) for every \(a, b \in H\).
Proof. (a) \implies (b). Recall that every finite abelian subgroup of $SL_n$ can be simultaneously diagonalized and hence, is toral. (Alternatively, since $SL_n$ is a special group, this follows from Example 6.6.) The tori of $PGL_n$ are precisely the images of the tori in $SL_n$ under the natural projection $SL_n \twoheadrightarrow PGL_n$, and part (b) follows.

(b) \implies (c). Suppose $H$ is contained in a maximal torus $T \subset PGL_n$ and let $S$ be the preimage of $T$ in $SL_n$. Then $S$ is a maximal torus of $SL_n$. Thus any $a, b \in H$ can be lifted to, respectively, $A, B \in S$. Since $A$ and $B$ commute, we conclude that $\alpha_H(a, b) = ABA^{-1}B^{-1} = 1$, as claimed.

(c) \implies (a). Since $\alpha_H$ is trivial, the preimage of $H$ in $SL_n$ is a finite abelian group. □

The embedding $\phi$. We will now show that any symplectic module $H$ can be obtained from an abelian subgroup of $PGL_n$, as above, with $n = \sqrt{|H|}$. Note that by Lemma 7.3(b) we may assume $H = (A \times A^*, \omega_A)$ for some abelian group $A$.

Definition 7.6 (cf. [RY, Definitions 8.7 and 8.10]). Let $A$ be an abelian group of order $n$. We define the embedding

$$\phi: A \times A^* \hookrightarrow PGL_n$$

as follows. Identify $PGL_n$ with $PGL(V)$, where $V = k[A]$ = the group algebra of $A$. The group $A$ acts on $V$ by the regular representation $a \mapsto P_a \in GL(V)$, where

$$P_a \left( \sum_{b \in A} c_b b \right) = \sum_{b \in A} c_b ab$$

for any $a \in A^*$ and $c_b \in k$. The dual group $A^*$ acts on $V$ by the representation $\chi \mapsto D_\chi \in GL(V)$, where

$$D_\chi \left( \sum_{a \in A} c_a a \right) = \sum_{a \in A} c_a \chi(a) a$$

for any $\chi \in A^*$ and $c_a \in k$. We define $\phi$ by

$$\phi(a, \chi) = \text{the image of } P_a D_\chi \text{ in } PGL(V).$$

Lemma 7.7. Let $A$ be a finite abelian group, $a, b \in A$ and $\chi, \mu \in A^*$. Then:

(a) $D_\chi P_a = \chi(a) P_a D_\chi$.
(b) $(P_a D_\chi)(P_b D_\mu)(P_a D_\chi)^{-1} = \chi(b) \mu^{-1}(a)(P_b D_\mu)$.
(c) The embedding $\phi$ of Definition 7.6 is a monomorphism of groups, and $\phi(A \times A^*)$ is subgroup of $PGL_n$ isomorphic to $A \times A^*$.

Proof. See [RY, Lemmas 8.8 and 8.11(i)]. □
Lemma 7.8. Let $A$ be an abelian group of order $n$ and let $\phi: A \times A^* \hookrightarrow \PGL_n$ be the embedding of Definition 7.6. Then $\phi$ induces an isomorphism of $(A \times A^*, \omega_A)$ and $(\phi(A \times A^*), \alpha)$ as modules with skew-symmetric forms, where $\omega_A$ is as in Definition 7.2 and $\alpha = \alpha_{\phi(A \times A^*)}$ is as in Definition 7.4. In particular, $(\phi(A \times A^*), \alpha)$ is a symplectic module.

Proof. Let $h_1 = (a_1, \chi_1)$ and $h_2 = (a_2, \chi_2) \in A \times A^*$. Then we want to show that
\[
\alpha(\phi(h_1), \phi(h_2)) = \omega_A(h_1, h_2) = \chi_1(a_2)\chi_2(a_1)^{-1}.
\]
On the other hand, by definition of $\alpha$, we have
\[
\alpha(\phi(h_1), \phi(h_2)) = (P_{a_1}D_{\chi_1})(P_{a_2}D_{\chi_2})(P_{a_1}D_{\chi_1})^{-1}(P_{a_2}D_{\chi_2})^{-1}.
\]
The desired equality now follows from Lemma 7.7(b).

Corollary 7.9. Let $A$ be an abelian group of order $n = p^r$. Then the subgroup $H = \phi(A \times A^*) \subset \PGL_n$ is of depth $r$.

Proof. Let $H_T$ be any maximal (with respect to inclusion) toral subgroup of $H$. By Lemma 7.5, $H_T$ is isotropic; as it is maximal, it is Lagrangian. The index $[H : H_T] = n^2/n = n = p^r$, and hence, the depth of $H$ is $r$.

If $n = p^r$ then the depth of any $p$-subgroup of $\PGL_n$ is $\leq r$. This can be shown directly or, alternatively, derived from Theorem 1.1, since any central simple algebra of degree $n$ is split by a degree $n$ extension of its center.

8. Symplectic modules and division algebras.

We are now ready to proceed with our results on division algebras.

When is $RMaps_{PGln}(X, M_n)$ a division algebra? We begin with an application of Theorem 1.1.

Let $X$ be a generically free irreducible $PGL_n$-variety. Recall that $A = RMaps_{PGln}(X, M_n)$ is a central simple algebra with the center $Z(A) = k(X)^{PGln}$; $A$ is of the form $M_d(D)$, where $D$ is a division algebra. The degree $d$ of $D$ is called the index of $A$, and $sd = n$. The following lemma relates smooth points in $X$ fixed by finite abelian subgroups of $PGL_n$, to the index of $A$.

Let $H$ be a finite abelian subgroup of $PGL_n$. The skew-symmetric form $\alpha_H$ on $H$ may be singular; the quotient $H/\Ker(\alpha_H)$ is a symplectic module, and hence, $|H/\Ker(\alpha_H)| = m^2$ for some integer $m$.

Lemma 8.1. With the notations as above, suppose that $H$ has a smooth fixed point $x \in X$.

Then the index of $A$ is divisible by $m$. In particular, $m \mid n$, and if $H = \phi_P(P \times P^*)$ where $P$ is an abelian group of order $n$ (so that $m = n$), then $A$ is a division algebra.
Proof. Let $F$ be the center of $D$ (and of $A = M_s(D)$), and let $K$ be a maximal subfield of $D$. Recall that $[K : F] = d = \deg(D) = \text{index}(A)$ and that $K$ is a splitting field of $A$. By Theorem 1.1, $d$ is divisible by $|H/H_T|$ where $H_T$ is some toral subgroup of $H$. By Lemma 7.5, $H_T$ is isotropic in $H$; we may assume that $H_T$ is a maximal isotropic subgroup of $H$. Then $H_T \supset \text{Ker}(\alpha_H)$, and the image of $H_T$ in $H/\text{Ker}(\alpha_H)$ is Lagrangian; it follows that $|H/H_T| = m$, i.e., $d$ is divisible by $m$.

The equality $sd = n$ implies then that $m \mid n$, and if $m = n$ then $s = 1$, i.e., $A$ is a division algebra. □

Proof of Theorem 1.3. The following proposition is an application of Theorem 1.2.

Proposition 8.2. Let $H$ be a finite abelian subgroup of $\text{PGL}_n$ of order $n^2 = p^{2r}$, such that $(H, \alpha_H)$ is a symplectic module (i.e., $\alpha_H$ is nondegenerate on $H$; see Definition 7.4). Suppose $X' \to X$ is a rational cover of irreducible generically free $\text{PGL}_n$-varieties, $p^e$ is the largest power of $p$ dividing $\deg(X'/X)$, and $X$ has a smooth point fixed by $H$. Then any splitting group $A'$ for $X'$ contains an isomorphic copy of some isotropic subgroup $I_1 \subset H$, where $|I_1| \geq p^{r-e}$.

In particular, if $e = 0$, $A'$ contains an isomorphic copy of a Lagrangian subgroup of $H$.

Proof. By the Going Up Theorem 2.7(a), $X'$ has an $H_1$-fixed point for some subgroup $H_1$ of $H$ of order $p^{2r-e}$. By Theorem 1.2, $A'$ contains a copy of $H_1/I$, where $I$ is a toral subgroup of $H_1$. Lemma 7.5 says that $\alpha_H$ restricted to $I$ is trivial, i.e., $I$ is an isotropic subgroup. Thus by Lemma 7.1(b), $H_1/I$ contains a copy of $I_1$, where $I_1$ is an isotropic subgroup of $H_1$ and $|H_1|$ divides $p^r \cdot |I_1|$. Since $|H_1| \geq p^{2r-e}$, this translates into $|I_1| \geq p^{r-e}$, as claimed. □

We now continue with the proof of Theorem 1.3. Recall that $\text{UD}(n) = R\text{Maps}_{\text{PGL}_n}(X, M_n)$, where $X = M_n \times M_n$, with $\text{PGL}_n$ acting by simultaneous conjugation; see Example 3.1. Let $D = \text{UD}(n) \otimes \mathbb{Z}(n) K$, as in the statement of Theorem 1.3. Then $D = R\text{Maps}_{\text{PGL}_n}(X_K, M_n)$. Recall that we are assuming $n = p^r$, and $p^e$ is the highest power of $p$ dividing $[K : \mathbb{Z}(n)] = \deg(X_K/X)$. Also recall that $A$ is a splitting group for $D$ if and only if $A$ is a splitting group for $X_K$ (as a $\text{PGL}_n$-variety); see Definition 6.1.

Note that $X = M_n \times M_n$ has a smooth point (namely, the origin) fixed by all of $G$. Let $P$ be an abelian $p$-group of order $n = p^r$; then $H = \phi_P(P \times P^*)$ is an abelian $p$-subgroup of $\text{PGL}_n$. By Lemma 7.8, $(H, \alpha_H)$ is a symplectic module. Applying Proposition 8.2 to $H$ and remembering that every symplectic module of order $p^{2r}$ is isomorphic to one of the form $\phi_P(P \times P^*)$ for some $P$ (see Lemmas 7.3(b) and 7.8), we obtain the following generalization of [TA1, Corollary 7.2] (in characteristic 0):
Proposition 8.3. Let \( Z(p^r) \) be the center of the generic division algebra \( UD(p^r) \), let \( K \) be a field extension of \( Z(p^r) \) and let \( D = UD(p^r) \otimes_{Z(p^r)} K \). Suppose \( p^e \) is the highest power of \( p \) dividing \( [K : Z(p^r)] \), where \( e \) is a nonnegative integer and \( e \leq r - 1 \).

If \( A \) is a splitting group of \( D \) then for every symplectic module \( H \) of order \( p^{2r} \), there exists an isotropic submodule \( I_1 \) of order \( p^{r-e} \) such that \( A \) contains an isomorphic copy of \( I_1 \).

In order to finish the proof of Theorem 1.3 we use a comparison argument, as in the proof of [TA1, Theorem 7.3]. Let

\[
(8.1) \quad H_1 = \phi_{P_1}(P_1 \times P_1^r) \quad \text{and} \quad H_2 = \phi_{P_2}(P_2 \times P_2^r),
\]

where

\[
(8.2) \quad P_1 = (\mathbb{Z}/p\mathbb{Z})^r \quad \text{and} \quad P_2 = \mathbb{Z}/p^r\mathbb{Z}.
\]

By Proposition 8.3, \( A \) contains an isomorphic copy \( I_1 \) of an isotropic subgroup of \( H_1 \), and an isomorphic copy \( I_2 \) of an isotropic subgroup of \( H_2 \), such that \( |I_1| = |I_2| = p^{r-e} \). Since \( H_1 \cong (\mathbb{Z}/p\mathbb{Z})^{2r} \) and \( H_2 \cong (\mathbb{Z}/p^r\mathbb{Z})^2 \), \( I_1 \cong (\mathbb{Z}/p\mathbb{Z})^{r-e} \) and \( I_2 \) has rank \( \leq 2 \). We may assume without loss of generality that both \( I_1 \) and \( I_2 \) are contained in the same Sylow \( p \)-subgroup \( A_p \) of \( A \). Since the intersection of \( I_1 \) and \( I_2 \) has exponent \( p \) and rank \( \leq 2 \), we see that

\[
|A_p| \geq |I_1 I_2| = \frac{|I_1| |I_2|}{|I_1 \cap I_2|} \geq \frac{p^{2r-2e}}{p^2} = p^{2r-2e-2},
\]

where \( I_1 I_2 = \{ \gamma_1 \gamma_2 \mid \gamma_1 \in I_1, \gamma_2 \in I_2 \} \).

This shows that \( |A_p| \) is divisible by \( p^{2r-2e-2} \) and, hence, so is \( |A| \), as claimed. \( \square \)

Remark 8.4. The only property of \( X = M_n \times M_n \) used in the above proof is that each of the finite abelian subgroups \( H_1 \cong (\mathbb{Z}/p^r\mathbb{Z})^2 \) and \( H_2 \cong (\mathbb{Z}/p\mathbb{Z})^r \) of \( \text{PGL}_n \) has a smooth fixed point in \( X \). Thus our argument shows that Theorem 1.3 remains valid if the universal division algebra \( UD(n) \) is replaced by the algebra \( U' = RMaps_{\text{PGL}_n}(X, M_n) \), where \( X \) is an irreducible generically free \( \text{PGL}_n \)-variety \( X \) such that \( H_i \) has a smooth fixed point in \( X \) for \( i = 1, 2 \). There are many choices for such \( X \); in particular, by Proposition 8.6 \( X \) can be chosen so that \( \dim(X/\text{PGL}_n) = 2r \) or, equivalently, \( \text{trdeg}_k(Z(U')) = 2r \), where \( Z(U') \) is the center of \( U' \).

Remark 8.5. Tignol and Amitsur showed that if \( A \) is an abelian splitting group of \( UD(p^r) \) and

\[
(8.3) \quad A_p \cong \mathbb{Z}/p^{n_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p^{n_r}\mathbb{Z}
\]

is its Sylow \( p \)-subgroup then \( n_\nu + n_{\nu+1} \geq \lceil r/\nu \rceil \) for every \( \nu = 1, 2, \ldots; \)

see [TA1, Theorem 7.4]. Consequently, the order of \( A_p \) (and, hence, of \( A \))
is divisible by $p^{f(r)}$, where

$$f(r) = r + \sum_{\nu \geq 3} \left\{ \frac{[r/\nu]}{2} \right\};$$

see [TA1, Theorem 7.5]. Here $[x]$ is the greatest integer $\leq x$ and $\{x\}$ is the smallest nonnegative integer $\geq x$. Note that $f(r) = \frac{1}{2} r \ln(r) + O(r)$, as $r \to \infty$; see [TA1, Remark 7.5] (a more precise asymptotic estimate is given in [TA2, Corollary 6.2]).

The same assertions hold if $A$ is an abelian splitting group for any prime-to-$p$ extension of $UD(p^r)$: The proof given in [TA1, Theorem 7.5] goes through unchanged, except that we use Proposition 8.3 (with $e = 0$) in place of [TA1, Corollary 7.2].

Moreover, let $D = UD(p^r) \otimes_{Z(p^r)} K$, where $p^e$ is the highest power of $p$ which divides $[K : Z(p^r)]$, as in the statement of Theorem 1.3. Suppose $A$ is a splitting group of $D$ and $A_p$ is the Sylow $p$-subgroup of $A$. If $A_p$ is as in (8.3) then a slight modification of the proof of [TA2, Lemma 6.1] (again, based on Proposition 8.3) shows that

$$n \nu + n_{\nu+1} \geq \left\lfloor \frac{r}{\nu} \right\rfloor - e$$

for every $\nu = 1, 2, \ldots$ and consequently, the order of $A_p$ (and, hence, of $A$) is divisible by $p^{f_e(r)}$, where

$$f_e(r) = r - e + \sum_{\nu \geq 3} \left\{ \frac{[r/\nu] - e}{2} \right\}.$$

It is easy to see that for a fixed $e$ and large $r$, $f_e(r)$ also grows as $\frac{1}{2} r \ln(r) + O(r)$.

**Reduction of Theorem 1.4 to a geometric problem.**  Our proof of Theorem 1.4 will be based on Proposition 8.2. The idea is to construct a generically free $\text{PGL}_{p^r}$-variety $X$ with two smooth points $x_1$ and $x_2$ whose stabilizers contain “incompatible” symplectic modules $H_1$ and $H_2$. Let $P_1$ and $P_2$ be as in (8.2); this time we take

$$H_2 = \phi_{P_2}(P_2 \times P_2^*) \simeq (\mathbb{Z}/p^r\mathbb{Z})^2,$$

as in (8.1), but allow $H_1$ to be smaller:

$$H_1 = \text{rank} 6 \text{ symplectic subgroup of } \phi_{P_1}(P_1 \times P_1^*).$$

Note that $H_1 \simeq (\mathbb{Z}/p\mathbb{Z})^6$ with desired properties exists by Lemma 7.3(c).

Suppose $X$ is an irreducible generically free $\text{PGL}_n$-variety, and $x_1$, $x_2$ are smooth points of $X$ such that $x_i$ is fixed by $H_i$. Let $D$ be the algebra $RMaps_{\text{PGL}_n}(X, M_n)$. Since $X$ has a smooth point fixed by $H_2$, Lemma 8.1 tells us that $D$ is a division algebra. Moreover, in view of Proposition 8.2 (with $X = X'$) any splitting group $A$ of $X$ (or equivalently, of $D$) will contain subgroups $L_1$ and $L_2$ which are isomorphic to Lagrangian submodules of $H_1$ and $H_2$, respectively. Note that $L_1 \simeq (\mathbb{Z}/p\mathbb{Z})^3$ and $L_2 \simeq (\mathbb{Z}/p^r\mathbb{Z}) \times$
\((\mathbb{Z}/p^{r-i}\mathbb{Z})\), for some \(0 \leq i \leq r\). Then \(L_1 \cap L_2\) is an abelian group of exponent \(p\) and rank \(\leq 2\). Thus
\[
|A| \geq |L_1 L_2| = \frac{|L_1| |L_2|}{|L_1 \cap L_2|} \geq \frac{p^3 p^r}{p^2} = p^{r+1};
\]

this shows that \(D\) is not a crossed product. The same argument shows that any prime-to-\(p\) extension of \(D\) is not a crossed product.

Thus, in order to prove Theorem 1.4 it is sufficient to construct an irreducible generically free \(\text{PGL}_n\)-variety \(X\) such that \(\text{trdeg}_k \mathbb{Q}(X)_{\text{PGL}_n} = \dim(X/\text{PGL}_n) = 6\) and \(X\) has smooth points \(x_1\) and \(x_2\) such that \(x_i\) is fixed by \(H_i\).

Note that both \(H_1\) and \(H_2\) are contained in the finite subgroup \(G\) of \(\text{PGL}_n\) generated by the permutation matrices and by the diagonal matrices all of whose entries are \(p^{r}\)th roots of unity. We will construct \(X\) as \(\text{PGL}_n \rtimes G\), where \(G\) is a 6-dimensional primitive \(G\)-variety with two points, \(y_1\) and \(y_2\) such that \(H_i\) fixes \(y_i\). Indeed, if \(Y\) is as above then the points \(x_1 = (1_{\text{PGL}_n}, y_1)\) and \(x_2 = (1_{\text{PGL}_n}, y_2)\) of \(X\) have the desired properties. (Recall that \(\text{PGL}_n \rtimes G\) is defined as the geometric quotient of \(\text{PGL}_n \times Y\) by the \(G\)-action given by \(g \cdot (h, y) = (h g^{-1}, g y)\); see [PV, Section 4.8].)

Therefore, in order to prove Theorem 1.4 it is enough to establish the following result.

**Proposition 8.6.** Let \(G\) be a finite group and let \(H_1, \ldots, H_s\) be abelian subgroups of \(G\), \(r_i = \text{rank}(H_i)\) and \(r = \max\{r_i \mid i = 1, \ldots, s\}\). Then there exists a generically free primitive \(r\)-dimensional projective \(G\)-variety \(Y\) with smooth points \(y_1, \ldots, y_s\) such that \(H_i \subseteq \text{Stab}(y_i)\).

**Remark 8.7.** Note that \(\dim(Y)\) cannot be less than \(r\). More precisely, if \(H\) is a finite abelian group, \(Y\) is a quasiprojective \(H\)-variety, and \(y\) is a smooth point of \(Y\) fixed by \(H\) then
\[
(8.4) \quad \text{codim}_y(Y^H) \geq \text{rank}(H).
\]

Indeed, assume the contrary: \(\text{codim}_y(Y^H) < \text{rank}(H)\). By Lemma 2.1(a), \(y\) has an \(H\)-invariant affine neighborhood in \(Y\). Replacing \(Y\) by this neighborhood, we may assume \(Y\) is affine. By the Luna Slice Theorem [PV, Corollary to Theorem 6.4], \(Y^H\) is smooth at \(y\) and
\[
\text{dim}(T_y(Y)) - \text{dim}(T_y(Y^H)) = \text{codim}_y(Y^H) < \text{rank}(H);
\]
hence, the action of \(H\) on \(T_y(Y)\) cannot be faithful. In other words, there exists a subgroup \(H' \subset H\), \(H' \neq \{1\}\), which acts trivially on \(T_y(Y)\). Applying [PV, Corollary to Theorem 6.4] to the action of \(H'\) on \(Y\), we see that \(H'\) acts trivially on all of \(Y\). This contradicts our assumption that the \(G\)-action on \(Y\) is generically free.
9. Constructing a $G$-variety with prescribed stabilizers.

As we have just seen, Theorem 1.4 follows from Proposition 8.6. This section will thus be devoted to proving Proposition 8.6. Our general approach is to first construct a higher-dimensional variety with desired properties (this is easy), then replace it by a “generic” $G$-invariant hypersurface passing through $y_1, \ldots, y_s$, thus reducing the dimension by 1. To carry out this program, we first reduce to a situation where $Y^H_i$ has the highest possible dimension at $y_i$ (Theorem 9.3), then apply Theorem 9.7, which may be viewed as a weak form of Bertini’s theorem in the equivariant setting.

**A local system of parameters.** The following lemma summarizes some known facts about the local geometry of a smooth $G$-variety near a point fixed by a finite abelian group.

**Lemma 9.1.** Let $H$ be a finite abelian group, let $X$ be a smooth quasiprojective $H$-variety, and let $D_1, \ldots, D_l$ be $H$-invariant hypersurfaces passing through a point $x \in X$ and intersecting transversely at $x$.

1. There exists a local coordinate system (a regular system of parameters) $u_1, \ldots, u_n$ with the following properties:
   - (i) The group $H$ acts on each $u_i$ by a character $\xi_i$;
   - (ii) $u_i$ is the local equation of $D_i$ for $i = 1, \ldots, l$;
   - (iii) The germ of the fixed-point set $X^H$ at $x$ is given by the local equations $u_{i_1} = \cdots = u_{i_t} = 0$ where $\{i_1, \ldots, i_t\}$ is the set of all subscripts $i$ for which the character $\xi_i$ is nontrivial.

2. Let $\pi : X' \rightarrow X$ be the blowup with the center $Z$ given by the local equations $u_{j_1} = \cdots = u_{j_s} = 0$ at $x$. (In particular, we can take $Z = X^H$.) Let $\tilde{D}_i \subset X'$ be the strict transform of $D_i$. Then we have:
   - (i) $\tilde{D}_1, \ldots, \tilde{D}_l$ and the exceptional divisor $\pi^{-1}(Z)$ are in normal crossing in a neighborhood of $\pi^{-1}(x)$;
   - (ii) the natural isomorphism $\pi^{-1}(x) \cong \mathbb{P}(T_x X/T_x Z)$ identifies $\tilde{D}_i \cap \pi^{-1}(x)$ with $\mathbb{P}(L_i)$, where $L_i$ is an $H$-invariant subspace of $T_x X/T_x Z$ of codimension 0 or 1.

Note that if $u_i$ is a local equation of $D_i$ and $H$ acts on $u_i$ by a character $\xi_i$ then $H$ acts by the character $\xi_i$ on the conormal space $(T_x X/T_x D_i)^*$.

**Proof.** By Lemma 2.1(a), we may assume without loss of generality that $X$ is affine.

1. Denote by $O_x$ the local ring of $X$ at $x$, by $m_x$ its maximal ideal, and by $p_{D_i}$ the ideal of $D_i$ in $O_x$.

To construct $u_1, \ldots, u_l$, note that the group $H$ acts on $O_x$, the ideals $m_x$ and $p_{D_i}$ are $H$-invariant subspaces in $O_x$, and the $H$-representation $(p_{D_i} + m_x^2)/m_x^2 \cong p_{D_i}/(p_{D_i} \cap m_x^2)$ is one-dimensional; let $\xi_i$ be the corresponding character of $H$. The $H$-linear epimorphism $p_{D_i} \rightarrow p_{D_i}/(p_{D_i} \cap m_x)$ splits;
this yields a generator \( u_i \in \mathfrak{p}_{D_i} \) — a local equation of \( D_i \) — on which \( H \) acts by the character \( \xi_i \).

To construct \( u_{i+1}, \ldots, u_n \), consider the \( H \)-linear epimorphism
\[
m_x \twoheadrightarrow m_x/(\mathfrak{p}_{D_1} + \cdots + \mathfrak{p}_{D_i} + m_x^2) \cong \frac{m_x/m_x^2}{\sum_{i=1}^t (\mathfrak{p}_{D_i} + m_x^2)/m_x^2};
\]
its splitting yields the elements \( u_{i+1}, \ldots, u_n \in m_x \) such that \( H \) acts on each of them by a character and the images of \( u_1, \ldots, u_n \in m_x/m_x^2 \) form a basis there. It follows that \( u_1, \ldots, u_n \) form a regular system of parameters at \( x \) that satisfies properties (1)(i) and (1)(ii).

According to the Luna Slice Theorem \([PV, \text{Corollary to Theorem 6.4}]\), \( X^H \) is given in a neighborhood of \( x \) by the local equations \( u_{i_1} = \cdots = u_{i_t} = 0 \), where \( \{i_1, \ldots, i_t\} \) is the set of all subscripts \( i \) for which the character \( \xi_i \) is nontrivial. This proves part (1)(iii).

(2) Let \( U \) be a small enough affine neighborhood of \( x \) in \( X \) so that \( u_1, \ldots, u_n \) form a local coordinate system (i.e., their differentials are linearly independent) everywhere on \( U \). The blown-up variety \( X' \) in a neighborhood of \( \pi^{-1}(x) \) is covered by the charts \( U_i \), \( 1 \leq i \leq s \), where \( U_i \) is the complement of \( \pi^{-1}(U) \) of the strict transform of the subvariety \( u_{j_i} = 0 \); the local coordinates in \( U_i \) are \( v_{j_i} = \pi^* u_{j_i}, v_{j'_i} = \pi^* u_{j'_i}/\pi^* u_{j_i} \) for \( i' \neq i \), and \( v_j = \pi^* u_j \) for \( j \not\in \{j_1, \ldots, j_s\} \). The exceptional divisor in \( U_i \) is given by the local equation \( v_{j_i} = 0 \), and \( D_j \) is given by the local equation \( v_j = 0 \) in case \( j \neq j_i \), and is empty in case \( j = j_i \). Since the local equations of \( D_1', \ldots, D_l \) and of the exceptional divisor are elements of the same local coordinate system, they are transverse; this proves (2)(i).

The local equations in \( U_i \) of the preimage \( \pi^{-1}(x) \) are \( v_j = 0 \) for \( j \not\in \{j_1, \ldots, j_s\} \), and \( v_{j_i} = 0 \); hence, \( \pi^{-1}(x) \) is contained in \( \widetilde{D}_j \) if \( j \not\in \{j_1, \ldots, j_s\} \), so that in this case \( \widetilde{D}_j \cap \pi^{-1}(x) = \pi^{-1}(x) \) as claimed in (2)(ii). If \( j \in \{j_1, \ldots, j_s\} \) then \( D_j \supset Z \) and \( \widetilde{D}_j \cap \pi^{-1}(x) \) can be identified with \( \mathbb{P}(T_z D_i/T_z Z) \); here \( L_i = T_z D_i/T_z Z \) is an \( H \)-invariant subspace of \( T_z X/T_z Z \) of codimension 1 as claimed. This completes the proof of (2)(ii).

\begin{lemma}
Let \( G \) be a finite group, \( H \) be an abelian subgroup of \( G \), \( X \) be a smooth quasiprojective \( G \)-variety and \( x \in X^H \). Then there exists a sequence of blowups
\end{lemma}
\[
(9.1) \quad f: X_i \twoheadrightarrow \cdots \twoheadrightarrow X_0 = X
\]
with smooth \( G \)-invariant centers, a point \( y \in X_i^H \) satisfying \( f(y) = x \), and smooth \( G \)-invariant hypersurfaces \( D_1, \ldots, D_l \) meeting transversely, such that locally at \( y \), the fixed point set \( X_i^H \) coincides with \( D_1 \cap \cdots \cap D_l \).

\begin{proof}
By Theorem 2.5, we may assume that \( \text{Stab}(x) \) is commutative.

The statement of the Lemma follows from the case \( H = \text{Stab}(x) \). Indeed, if \( X_{i}^\text{Stab}(y) = D_1 \cap \cdots \cap D_l \) locally at \( y \), then by Luna’s slice theorem (see
[PV, Theorem 6.4], \( X^H_i \) is the intersection of those of \( D_j \) for which the action of \( H \) on the normal space \( T_y(X_i)/T_y(D_j) \) is nontrivial; thus, \( X^H_i \) is the intersection of some \( D_j \), as claimed. (Note that the Luna Slice Theorem can be applied to the \( G \)-action on \( X_i \) because \( X_i \) is quasiprojective and, hence, every point of \( X_i \) has an open affine \( G \)-invariant neighborhood; see Lemma 2.1.)

From now on we assume that \( \text{Stab}(x) = H \). Let \( X' = X - G \cdot \bigcup_{H \subsetneq H'} X^{H'} \); this is a \( G \)-invariant open dense quasiprojective subvariety of \( X \) containing \( x \) and not containing any point whose stabilizer is strictly larger than \( H \). If we can find a sequence of blowups (9.1) for \( X' \), then we can extend it to a similar sequence for \( X \) by extending each blowup center in \( X' \) to its closure in \( X \) and equivariantly resolving it before blowing it up, in case it is not smooth; for equivariant resolution of singularities, see either [V, Theorem 7.6.1] or [BM, Theorem 13.2].

Thus, we may assume that \( X \) does not contain points with stabilizers strictly containing \( H \); this implies that the subvarieties \( X^gH^g \) for different \( g \in G \) are disjoint unless they coincide. Each of these subvarieties is smooth by Luna’s slice theorem (see [PV, Corollary to Theorem 6.4]), and hence, their union \( GX^H = \bigcup_{g \in G} X^gH^g \) is smooth.

Let \( \dim X = n \), and let

\[
X_n \xrightarrow{\pi_n} \ldots \xrightarrow{\pi_{i+1}} X_i \xrightarrow{\pi_i} \ldots \xrightarrow{\pi_1} X_0 = X
\]

be the sequence of blowups centered at \( Z_i = GX^H_i \subset X_i \); each blowup \( \pi_i \) is \( G \)-equivariant. Inductively, each \( X_i \) has no points whose stabilizers strictly contain \( H \); together with the fact that \( X_i \) is smooth, this implies that \( Z_i \) is smooth, and hence, \( X_{i+1} \) is smooth, so that \( X_i \) and \( Z_i \) are smooth for every \( i \).

Let \( E_i = D_{i1} \cup \cdots \cup D_{ii} \) be the exceptional divisor in \( X_i \), where \( D_{ij} \subset X_i \) is the strict transform (in \( X_i \)) of the exceptional divisor of \( \pi_j : X_j \to X_{j-1} \). We claim that \( E_i \) is a normal crossing divisor. The proof is by induction on \( i \). The base case, \( i = 0 \), is obvious, since \( E_0 = \) the empty divisor, is normal crossing. For the inductive step, we assume that \( E_i \) is a normal crossing divisor. Then \( E_{i+1} \) is also a normal crossing divisor by Lemma 9.1(2)(i). This completes the proof of the claim.

To obtain the required point \( y \in X^H_m \), we start with \( x_0 = x \) and inductively construct \( x_i \in X_i \) satisfying \( \pi_i(x_i) = x_{i-1} \) and \( x_i \in X^H_i \cap D_{i1} \cap \cdots \cap D_{ii} \), until we get a point \( y = x_m \) with the desired properties.

Suppose \( x_i \) has been constructed for some \( i \geq 0 \). Note that near \( x_i \), the center \( Z_i = GX^H_i \) coincides with \( X^H_i \).
Case 1. The germ of \( X_i^H \) at \( x_i \) does not contain the germ of \( S = \bigcap_{j=1}^{i} D_{ij} \).

Let

\[
W = \pi_{i+1}^{-1}(x_i) \cap \bigcap_{j=1}^{i+1} D_{i+1,j} \subset X_{i+1}.
\]

We claim \( W \neq \emptyset \). Indeed, since \( x_i \in X_i^H \), we have \( \pi_{i+1}^{-1}(x_i) \subset \pi_{i+1}^{-1}(Z_i) = D_{i+1,i+1} \), and thus

\[
W = \pi_{i+1}^{-1}(x_i) \cap \bigcap_{j=1}^{i} D_{i+1,j}.
\]

Since \( D_{i+1,j} \) is the strict transform of \( D_{ij} \), \( \bigcap_{j=1}^{i} D_{i+1,j} \) contains the strict transform of \( S = \bigcap_{j=1}^{i} D_{ij} \). Thus \( W \) contains the intersection of the strict transform of \( S \) with \( \pi_{i+1}^{-1}(x_i) \). As the germ of \( S \) at \( x_i \) is not contained in the germ of the blowup center \( Z_i \), i.e., of \( X_i^H \), the strict transform of \( S \) is nonempty and intersects \( \pi_{i+1}^{-1}(x_i) \). Consequently, \( W \) is nonempty, as claimed.

We now identify \( \pi_{i+1}^{-1}(x_i) \) with \( \mathbb{P}(T_x X_i/T_x Z_i) \) in the usual manner. Note that this identification is \( H \)-equivariant. Then by Lemma 9.1(2)(ii), \( \pi_{i+1}^{-1}(x_i) \cap D_{i+1,j} \) is identified with \( \mathbb{P}(L_j) \), where \( L_j \) is an \( H \)-invariant subspace of the normal space \( T_x X_i/T_x Z_i \). Thus \( W \) is identified with \( \mathbb{P}(L) \), where \( L = L_1 \cap \cdots \cap L_i \); see (9.2). Note that \( L \neq (0) \) because \( W \neq \emptyset \); moreover, \( H \) acts linearly on \( L \). Since \( H \) is diagonalizable, it has an eigenvector in \( L \), i.e., a fixed point in \( W \cong \mathbb{P}(L) \). This fixed point is our new point \( x_{i+1} \). By our construction, \( \pi_{i+1}(x_{i+1}) = x_i \) and \( x_{i+1} \in X_{i+1}^H \cap \bigcap_{j=1}^{i+1} D_{i+1,j} \), as desired.

Case 2. \( X_i^H \) at \( x_i \) contains \( \bigcap_{j=1}^{i} D_{ij} \) at \( x_i \). Note that this is necessarily the case if \( i = n \). By Lemma 9.1(1), \( X_i^H \) at \( x_i \) coincides with the intersection of those of \( D_{ij} \) for which the action of \( H \) on \( T_{x_i} X_i/T_{x_i} D_{ij} \) is nontrivial; in particular, \( X_i^H \) at \( x_i \) is an intersection of smooth \( G \)-invariant hypersurfaces meeting transversely, as required.

Thus, we see that for some \( i \leq n \), Case 2 occurs and we get a point \( x_i \in X_i \) with the required properties. \( \square \)

**Resolving the action on the tangent space.** In this subsection we prove the following result.

**Theorem 9.3.** Let \( G \) be a finite group, let \( X \) be a smooth quasiprojective \( G \)-variety, let \( H_1, \ldots, H_n \) be abelian subgroups of \( G \) and \( x_1, \ldots, x_n \) be points of \( X \) such that \( x_i \) is fixed by \( H_i \). Denote the rank of \( H_i \) by \( r_i \). Then there is a sequence of blowups \( \pi: X_m \longrightarrow \cdots \longrightarrow X_0 = X \) with smooth \( G \)-invariant centers, and points \( y_1, \ldots, y_s \in X_m \), such that \( \pi(y_i) = x_i \), \( y_i \in X_m^{H_i} \), and \( X_m^{H_i} \) has codimension \( r_i \) at \( y_i \).
Our proof relies on the following two simple lemmas.

**Lemma 9.4.** Let $H$ be a finite abelian group, let $X$ be an $H$-variety and $x \in X^H$ be a smooth point of $X$. If $\pi: \tilde{X} \rightarrow X$ is a blowup with a smooth $H$-invariant center $Z \subset X$ then there exists a point $\tilde{x} \in \pi^{-1}(x)$ fixed by $H$ such that $\dim_{\tilde{x}}(\tilde{X}^H) \geq \dim_x(X^H)$.

**Proof.** Replacing $X$ by the set of its smooth points (which is clearly $H$-invariant), we may assume that $X$ is smooth. We claim that $\pi(\tilde{X}^H) = X^H$. The inclusion $\pi(\tilde{X}^H) \subset X^H$ is obvious. To prove the opposite inclusion, note that since $\pi$ is an isomorphism over $X \setminus Z$, every $y \in X^H - Z$ lies in $\pi(\tilde{X}^H)$. On the other hand, if $y \in Z^H$ then $\pi^{-1}(y)$ can be identified with $\mathbb{P}(T_y\tilde{X}/T_yZ)$ as $H$-varieties, and the (linear) action of $H$ on $\mathbb{P}(T_y\tilde{X}/T_yZ)$ has a fixed point $\tilde{y}$; then $y = \pi(\tilde{y}) \in \pi(\tilde{X}^H)$. This proves the claim, and the lemma follows. \qed

**Remark 9.5.** Lemma 9.4 remains true under the more general assumption that $H$ is Levi-commutative rather than finite abelian; see Definition 2.3 and Lemma 2.4(iv). The version we stated is sufficient for our application.

**Lemma 9.6.** Let $A$ be a finite abelian group of rank $r$. An elementary operation on $A^s$ is one of the form

$$(\xi_1, \ldots, \xi_i, \ldots, \xi_s) \mapsto (\xi_1, \ldots, \xi_i - \xi_j, \ldots, \xi_s)$$

for some $1 \leq i, j \leq s$, where $i \neq j$.

Assume $s \geq r$. Then any $\xi = (\xi_1, \ldots, \xi_s) \in A^s$ can be transformed, by a finite sequence of elementary operations, into an $s$-tuple with at least $s - r$ zeros. (Here by a zero, we mean the identity element of $A$.)

**Proof.** First we note that it does no harm to permute the components of $\xi$. In other words, we may as well consider an operation of the form $(\xi_1, \ldots, \xi_s) \mapsto (\xi_{\sigma(1)}, \ldots, \xi_{\sigma(s)})$ with $\sigma \in S_n$, as another type of elementary operation. The assertion of the lemma is then equivalent to saying that any $\xi \in A^s$ can be transformed, by these two types of elementary operations, into an $s$-tuple

$$\lambda = (\lambda_1, \ldots, \lambda_r, 0_A, \ldots, 0_A),$$

where $0_A$ is the identity element of $A$.

We will prove this assertion by induction on $r$. Suppose $r = 1$, i.e., $A = \mathbb{Z}/n\mathbb{Z}$ for some $n \geq 1$. We can use elementary operations to perform the Euclidean algorithm on $\xi_1$ and $\xi_2$. After interchanging them if necessary, we may assume $\xi_2 = 0$. (The new value of $\xi_1$ is the greatest common divisor of the old values of $\xi_1$ and $\xi_2$.) Applying the same procedure to $\xi_1$ and $\xi_3$, then $\xi_1$ and $\xi_4$, etc., we reduce the original $s$-tuple to $(\xi_1, 0, \ldots, 0)$, as claimed.

For the induction step, write $A = B \times C$, where $B$ has rank $r - 1$ and $C$ is cyclic. Set $\xi_i = (\beta_i, \gamma_i)$, where $\beta_i \in B$ and $\gamma_i \in C$. As we saw above, after performing a sequence of elementary operations, we may assume $\gamma_2 = \cdots =
\( \gamma_s = 0_C \). By the induction assumption, there exists a sequence of elementary operations in \( B^{s-1} \) which reduces \( (\beta_2, \ldots, \beta_s) \) to \( (\lambda_2, \ldots, \lambda_r, 0_B, \ldots, 0_B) \). (Note that since \( r \leq s \), \( \text{rank}(B) = r - 1 \leq s - 1 \), so that we may, indeed, use the induction assumption.) Applying the same sequence to \( (\xi_2, \ldots, \xi_s) \), we reduce \( (\xi_1, \ldots, \xi_s) \) to

\[
(\xi_1, (\lambda_2, 0_C), \ldots, (\lambda_r, 0_C), 0_A, \ldots, 0_A) \in A^s.
\]

This completes the proof of the lemma. \( \square \)

**Proof of Theorem 9.3.** By (8.4), we have

\[
(9.3) \quad \text{codim}_{x_i}(X^{H_i}) \geq r_i
\]

for any \( i \). We want to modify \( X \) by a sequence of blowups so as to decrease \( \text{codim}_{x_i} X^{H_i} \) to \( r_i \) for each \( i \). (Of course, after each blowup \( \tilde{X} \longrightarrow X \) we replace \( X \) by \( \tilde{X} \) and \( x_i \) by \( \tilde{x}_i \), as in Lemma 9.4.) We claim that we may do this for one \( i \) at a time; in other words, we may assume \( s = 1 \). Indeed, suppose we have reduced to the case where \( \text{codim}_{x_1}(X^{H_1}) = r_1 \). If we now perform a further blowup \( \tilde{X} \longrightarrow X \) and choose \( \tilde{x}_1 \) above \( x_1 \) as in Lemma 9.4, then Lemma 9.4 and (9.3) tell us that \( \text{codim}_{x_1}(\tilde{X}^{H_1}) = r_1 \). Thus we are free to perform another sequence of blowups that would give us the desired equality for \( i = 2 \), then \( i = 3 \), etc.

We will thus assume \( s = 1 \) and set \( x = x_1 \), \( H = H_1 \), \( r = r_1 \).

After performing a sequence of blowups given by Lemma 9.2, we may assume that there exist \( G \)-invariant divisors \( D_1, \ldots, D_c \) such that \( X^H = D_1 \cap \cdots \cap D_c \) in a neighborhood of \( x \), \( c = \text{codim}_x X^H \) and \( D_1, \ldots, D_c \) intersect at \( x \) transversely.

Note that \( T_x(X)/T_x(X^H) \cong \bigoplus_{j=1}^c T_x(D_j)/T_x(D_j) \). Here \( H \) acts on each one-dimensional space \( T_x(X)/T_x(D_j) \) by a character \( \xi_j \in H^* \) which is nontrivial by Lemma 9.1(1)(iii). In other words, the linear action of \( H \) on the tangent space \( T_x(X) \) decomposes as a direct sum of \( c \) nontrivial characters \( \xi_1, \ldots, \xi_c \) and \( n - c = \dim X^H \) trivial characters.

Recall that by (9.3), \( c \geq r \). We would like to modify \( X \) by a sequence of blowups to arrive at the situation where \( c = r \). In other words, if \( c > r \), we want to perform a sequence of blowups that would lower the value of \( c \).

With this goal in mind, we would like to know how the characters \( \xi_i \) change after one blowup. Specifically, we will consider the blowup \( \pi: \tilde{X} \longrightarrow X \) with center \( Z = D_i \cap D_j \), where \( 1 \leq i, j \leq c \), \( i \neq j \). Since \( Z \) is of codimension 2 in \( X \), \( \pi^{-1}(x) \) is isomorphic to \( \mathbb{P}^1 \). Let \( \tilde{x} \) be the (unique) point of \( \pi^{-1}(x) \) that lies in the strict transform of \( D_j \), take \( \tilde{D}_l \) to be the strict transform of \( D_l \) for \( l = 1, \ldots, j, \ldots, c \), and let \( \tilde{D}_j = \pi^{-1}(Z) \) be the exceptional divisor of \( \pi \). Then the action of \( H \) in \( T_{\tilde{x}} \tilde{X} \) is given by the direct sum of the characters \( \xi_l = \xi_l \) if \( l \neq i \), \( \xi_i = \xi_i \xi_j^{-1} \), and \( c \) trivial characters. In other words, the new characters \( \xi_1, \ldots, \xi_c \in H^* \) are obtained
from the old characters \( \xi_1, \ldots, \xi_c \in H^* \) by an elementary operation, as in Lemma 9.6. (Note that our group \( H^* \) is written multiplicatively, whereas the group \( A \) in Lemma 9.6 is written additively.)

Now Lemma 9.6 tells us that there is a sequence of elementary operations which transforms \((\xi_1, \ldots, \xi_c)\) to \((\lambda_1, \ldots, \lambda_r, 1_H, \ldots, 1_H^*)\) for some \( \lambda_1, \ldots, \lambda_r \in H^* \). Recall that initially our characters \( \xi_i \) are all nontrivial. We want to follow the above sequence of elementary operations until we create the first trivial character. Each one of these operations is given by a blowup of a codimension 2 subvariety, as described above. When the first trivial character appears, \( \dim(X^H) \) goes up by one. At that point, we reduce \( c \) by 1 and repeat the above procedure, until \( c \) becomes equal to \( r \).

\[ \square \]

A weak equivariant Bertini theorem.

**Theorem 9.7.** Let \( G \) be a finite group, \( X \) a primitive smooth projective \( G \)-variety, \( x_1, \ldots, x_s \) points of \( X \) with stabilizers \( H_i = \text{Stab}(x_i) \) for \( i = 1, \ldots, s \). Suppose that each \( x_i \) is not an isolated point of \( X^{H_i} \) and \( \dim(X) \geq 2 \). Then:

(a) There exists a smooth closed \( G \)-invariant primitive hypersurface \( W \subset X \) passing through \( x_1, \ldots, x_s \).

(b) Moreover, if \( X \) is a generically free \( G \)-variety then we can choose \( W \) so that it is also generically free.

We shall need the following variant of Bertini’s theorem; for lack of a reference we will supply a proof.

**Lemma 9.8.** Let \( x_1, \ldots, x_s \) be points in \( \mathbb{P}^n \) and let \( V_d \) be the space of homogeneous polynomials of degree \( d \geq s + 1 \) in \( \mathbb{P}^n \) that vanish at \( x_1, \ldots, x_s \).

(a) Suppose \( Y \) is a locally closed subvariety of \( \mathbb{P}^n \), \( y \) is a smooth point of \( Y \) and \( V_{d,y} = \{ P \in V_d \mid P|_Y \text{ has zero of order } > 1 \text{ at } y \} \). Then the codimension of \( V_{d,y} \) in \( V_d \) is given by

\[
\text{codim}(V_{d,y}) = \begin{cases} 
\dim_y(Y) + 1 & \text{if } y \notin \{x_1, \ldots, x_s\}, \\
\dim_y(Y) & \text{if } y \in \{x_1, \ldots, x_s\}.
\end{cases}
\]

(b) Suppose \( Y_1, \ldots, Y_l \) are smooth locally closed subvarieties of \( \mathbb{P}^n \) such that \( x_j \) is not an isolated point of \( Y_i \) for any \( i, j \).

Then a generic hypersurface of degree \( d \geq s + 1 \) in \( \mathbb{P}^n \) that passes through \( x_1, \ldots, x_s \), is transverse to \( Y_1, \ldots, Y_l \).

Note that assumption (9.4) in part (b) is necessary, since otherwise no hypersurface passing through \( x_j \) can be transverse to \( Y_i \). (By definition, a hypersurface \( W \) is transverse to a one-point set \( \{x_j\} \) iff \( W \) does not pass through \( x_j \).)
Proof. (a) Choose an affine subset \( \mathbb{A}^n = \text{Spec } k[z_1, \ldots, z_n] \subset \mathbb{P}^n \) that contains \( x_1, \ldots, x_s \) and \( y \); then \( V_d \) may be identified with the space of all polynomials in \( z = (z_1, \ldots, z_n) \) of degree \( \leq d \) (not necessarily homogeneous) that vanish at the points \( x_1, \ldots, x_s \).

Consider the linear map \( \phi_y : V_d \rightarrow \mathcal{O}_y/m_y^2 \), where \( \mathcal{O}_y/m_y^2 \) is the space of 1-jets of regular functions on \( \mathbb{A}^n \) at \( y \); \( \phi_y \) sends each polynomial \( P \in V_d \) into its 1-jet at \( y \) (cf. [Ha, Proof of Theorem 8.18]). Then \( V_{d,y} = \phi_y^{-1}(N_y) \) where \( N_y \) is the subspace of \( \mathcal{O}_y/m_y^2 \) consisting of the jets of all functions that vanish on \( Y \); \( N_y \) may be identified with the conormal space to \( Y \) at \( y \), and hence, it is a linear subspace of \( \mathcal{O}_y/m_y^2 \) of dimension \( \dim(N_y) = \dim \mathcal{O}_y(Y) \).

Assume \( y \notin \{x_1, \ldots, x_s\} \). Let \( l(z) \) be a linear combination of \( z_1, \ldots, z_n \) whose value at \( y \) is different from its values at \( x_1, \ldots, x_s \). Then the degree \( s \) polynomial \( P(z) = \left( l(z) - l(x_1) \right) \cdots \left( l(z) - l(x_s) \right) \) vanishes at \( x_1, \ldots, x_s \) and does not vanish at \( y \). This means that \( \phi_y(P(z)) \) and \( \phi_y(z_j P(z)) \) (where \( j = 1, \ldots, n \)) span \( \mathcal{O}_y/m_y^2 \) as a \( k \)-vector space. Hence, \( \phi_y \) is onto, and

\[
\text{codim}(V_{d,y}) = \text{codim}(\phi_y^{-1}(N_y)) = \dim(\mathcal{O}_y/m_y^2) - \dim(N_y) = n + 1 - \text{codim}_y(Y) = \dim_y(Y) + 1
\]

as claimed.

Now suppose \( y \in \{x_1, \ldots, x_s\} \), say, \( y = x_1 \). In this case \( \phi_y(V_d) \) is clearly contained in \( m_y/m_y^2 \); we will show that, in fact, equality holds. Indeed, we may assume without loss of generality that \( y = x_1 = (0, \ldots, 0) \in \mathbb{A}^n \). Let \( l(z) \) be a linear combination of \( z_1, \ldots, z_n \) such that \( l(x_2), \ldots, l(x_s) \neq l(x_1) = 0 \) and let \( Q(z) = \left( l(z) - l(x_2) \right) \cdots \left( l(z) - l(x_s) \right) \). Given any linear function \( a(z) \) in \( z_1, \ldots, z_n \), let \( R(z) = a(z)Q(z) \). Note that \( R(z) \) vanishes at \( x_1, \ldots, x_s \) and has degree \( s \); hence, \( R(z) \in V_d \). On the other hand, \( \phi_y(R) \) equals \( R(z) \) modulo the terms of degree \( \geq 2 \) in \( z_1, \ldots, z_n \), i.e., \( \phi_y(R) = Q(0)a(z) \bmod m_y^2 \), where \( Q(0) \neq 0 \) is a nonzero element of \( k \). This means that \( \phi_y(V_d) \) contains \( a(z) \), thus proving that \( \phi_y(V_d) = m_y/m_y^2 \), as desired. Now

\[
\text{codim}(V_{d,y}) = \text{codim}(\phi_y^{-1}(N_y)) = \dim(m_y/m_y^2) - \dim(N_y) = n - \text{codim}_y(Y) = \dim_y(Y).
\]

(b) After replacing each \( Y_i \) by the collection of its irreducible components, we may assume each \( Y_i \) is irreducible. Moreover, we may assume without loss of generality that \( l = 1 \); we shall denote \( Y_1 \) by \( Y \).

Let \( X \) be the algebraic subvariety of \( V_d \times Y \) given by

\[
X = \left\{ (P,y) \mid P|_Y \text{ has a zero of order } > 1 \text{ at } y \right\}.
\]

Denote the natural projections of \( X \) to \( V_d \) and \( Y \) by \( \pi_1 \) and \( \pi_2 \). We want to show that \( \dim(\pi_1(X)) < \dim(V_d) \). It is thus enough to prove that \( \dim(X) <
dim($V_d$). Write $X = X_1 \cup X_2$, where $X_1 = \pi_2^{-1}(Y - \{x_1, \ldots, x_s\})$ and $X_2 = \pi_2^{-1}(Y \cap \{x_1, \ldots, x_s\})$. It is enough to show that $\dim(X_i) < \dim(V_d)$ for $i = 1, 2$. The fibers of $\pi_2$ are precisely the sets $V_{d,y}$ we considered in part (a). Since $\dim(V_{d,y}) = \dim(V_d) - \dim(Y) - 1$ for every $y \in Y - \{x_1, \ldots, x_s\}$, we conclude that $\dim(X_1) \leq \dim(V_d) - 1$.

It remains to show that $\dim(X_2) \leq \dim(V_d) - 1$. If $Y \cap \{x_1, \ldots, x_s\} = \emptyset$ then $X_2 = \emptyset$, and there is nothing to prove. On the other hand, if $Y \cap \{x_1, \ldots, x_s\} \neq \emptyset$ then assumption (9.4) says that $\dim(Y) \geq 1$. Thus by part (a),

$$\dim(X_2) \leq \dim(V_d) - \dim(Y) \leq \dim(V_d) - 1,$$

as claimed.

□

Proof of Theorem 9.7. We begin with three simple observations. First of all, we may assume without loss of generality that the orbits $Gx_i$ are disjoint. Indeed, if $W$ passes through $x_i$ then it will pass through every point of $Gx_i$. Thus if, say, $x_j$ happens to lie in $Gx_i$ then we can simply remove $x_j$ from our finite collection of points and proceed to construct $W$ for the smaller collection.

Secondly, part (b) is an immediate consequence of part (a). Indeed, since $G$ is a finite group, generically free $G$-varieties are precisely faithful $G$-varieties, i.e., $G$-varieties, where every nonidentity element of $G$ acts nontrivially. The set $X_0 = \{x \in X \mid \text{Stab}(x) = \{1\}\}$ is open and dense in $X$; in order to ensure that $W$ is generically free, it is enough to construct $W$ so that $W \cap X_0 \neq \emptyset$. This is accomplished by applying part (a) to the collection $\{x_0, x_1, \ldots, x_s\}$, where $x_0 \in X_0$. Therefore, it is enough to prove part (a).

Thirdly, since $G$ is a finite group and $X$ is projective, there exists a (finite) geometric quotient morphism $\psi: X \longrightarrow X//G$ with $X//G$ projective; see Lemma 2.1. Since $X$ is a primitive $G$-variety, $X//G$ is irreducible. (Recall that the geometric quotient $X//G$ is a birational model for the rational quotient $X/G$ which is irreducible since $X$ is primitive.) Note that $X$ is partitioned into a union of nonintersecting smooth locally closed subsets $\widetilde{X^H} = \{x \in X \mid \text{Stab}(x) = H\}$, where $H$ ranges over the set of subgroups of $G$. By the Luna Slice Theorem [PV, Theorem 6.1] the morphism

$$\psi|_{\widetilde{X^H}}: \widetilde{X^H} \longrightarrow \psi(\widetilde{X^H})$$

is étale, and hence, the sets $\psi(\widetilde{X^H})$ are also smooth. (Note that Luna’s theorem can be applied to the $G$-action in a neighborhood of any point of $X$ by Lemma 2.1(a).) Two subvarieties $\psi(\widetilde{X^H})$ and $\psi(\widetilde{X^{H'}})$ coincide if the subgroups $H$ and $H'$ are conjugate, and are disjoint otherwise. In other
words, $X//G$ is partitioned into a union of nonintersecting smooth locally closed subsets $\psi(\widetilde{X^H})$, one for each conjugacy class of subgroups in $G$. Every $y \in X//G$ lies in $\psi(\widetilde{X^H})$, where $H = \text{Stab}(x)$ and $x \in \psi^{-1}(y)$ (the conjugacy class of $H$ does not depend on the choice of $x$ in $\psi^{-1}(y)$).

We are now ready to proceed with the proof of part (a). Since $X//G$ is projective, we can embed it in $\mathbb{P}^N$ for some $N$. Let $U$ be a generically chosen hypersurface of degree $s + 1$ in $\mathbb{P}^N$ which passes through $\psi(x_1), \ldots, \psi(x_s)$, and let $\overline{W} = X//G \cap U$. We claim that

$$W = \pi^{-1}(\overline{W}) \subset X$$

satisfies the conditions of part (a). By our construction $W$ is a closed $G$-invariant hypersurface in $X$ passing through $x_1, \ldots, x_s$; thus we only need to show that $W$ is smooth and primitive.

Since each $x_i$ lies in $X^H_i$ and is not its isolated point, each $\psi(x_i)$ lies in $\psi(\widetilde{X^H})$ and is not its isolated point; it follows that $\psi(x_i)$ is not an isolated point of any $\psi(\widetilde{X^H})$. By Lemma 9.8(b), $U$ intersects every subvariety $\psi(\widetilde{X^H})$ transversely.

Let $x \in W$ and $H = \text{Stab}(x)$; then $x \in \widetilde{X^H}$ and $\psi(x) \in \psi(\widetilde{X^H})$. Let $f$ be a local equation of $U$ at $\psi(x)$; then $\psi^*(f)$ is a local equation of $W$ in $X$ at $x$. Since $U$ is transverse to $\psi(\widetilde{X^H})$, the restriction $f|_{\psi(\widetilde{X^H})}$ is nondegenerate at $\psi(x)$. As the morphism (9.5) is étale, $\psi^* \left( f|_{\psi(\widetilde{X^H})} \right) = \psi^*(f)|_{\widetilde{X^H}}$ is a nondegenerate function on $\widetilde{X^H}$ at $x$. Hence, $\psi^*(f)$ is nondegenerate at $x$; in other words, $W$ is smooth at $x$.

It remains to prove that $W$ is primitive or, equivalently, $\overline{W}$ is irreducible. By [Ha, Corollary III.7.9], $\overline{W} = X//G \cap U$ is connected. On the other hand, since $W$ is smooth, $\overline{W} = W//G$ is normal. We conclude that $\overline{W}$ is irreducible, as claimed.

Proof of Proposition 8.6. If $r = 0$ then $H_1 = \cdots = H_s = \{1\}$, and we can take $Y$ to be a set of $|G|$ points with a transitive $G$-action and $y_1, \ldots, y_s$ to be any $s$ points in $Y$ (not necessarily distinct). From now on we shall assume $r \geq 1$.

Let $V$ be a generically free linear representation of $G$. By Proposition 2.2 and Theorem 2.5, there exists a smooth projective $G$-variety $X$ such that $X \simeq V$ (as $G$-varieties) and Stab$(x)$ is commutative for any $x \in X$. (Note that a Levi-commutative finite group is commutative.) Every $H_i$ has a smooth fixed point in $V$, namely the origin. Applying the Going Down Theorem 2.6 to the birational isomorphism $V \xrightarrow{\sim} X$, we conclude that $X^{H_i} \neq \emptyset$ for every $i$. The resulting smooth projective irreducible generically free $G$-variety $X$ is the starting point for our construction.

After birationally modifying $X$ by a sequence of blowups with smooth $G$-equivariant centers, we may assume that there are points $x_1, \ldots, x_s \in X$
such that each $x_i$ is fixed by $H_i$ and the codimension of $X^{H_i}$ at $x_i$ is $r_i$; see Theorem 9.3.

If $\dim X > r = \max_i r_i$ then $\dim X > r_i = \dim_{x_i} X^{H_i}$ and hence, $x_i$ is not an isolated fixed point of $H_i$ for each $i$. In addition, $\dim X > r \geq 1$ implies $\dim X \geq 2$. Then Theorem 9.7(b) yields a smooth closed generically free $G$-invariant primitive hypersurface $W$ in $X$ passing through $x_1, \ldots, x_s$. Replacing $X$ by this hypersurface reduces $\dim X$ by one. Applying this procedure $\dim X - r$ times, we obtain a smooth $G$-invariant primitive subvariety $Y$ of dimension $r$ passing through $x_1, \ldots, x_s$, and hence, having points fixed by $H_1, \ldots, H_s$. This completes the proof of Proposition 8.6 and thus of Theorem 1.4. □

**Remark 9.9.** A closer examination of the proof of Proposition 8.6 shows that the $G$-variety $Y$ can, in fact, be constructed over $\mathbb{Q}$. Thus the division algebra $D$ is Theorem 1.4 can be assumed to be defined over $\mathbb{Q}$. This means that there exists a finitely generated field extension $F/\mathbb{Q}$ and a division algebra $D_0$ with center $F$ such that $\text{trdeg}_{\mathbb{Q}}(F) = 6$ and $D = D_0 \otimes_F K$.

**Remark 9.10.** Our argument can be modified to prove the following stronger form of Theorem 1.4: For any integer $e \geq 0$ there exists a division algebra $D$ with center $K$ such that

(a) $K$ is a finitely generated extension of $k$ of transcendence degree $6 + 2e$

and

(b) any extension of $D$ of degree $s$ is not a crossed product, provided that $p^{e+1} \nmid s$.

**References**


Received September 16, 1999 and revised July 6, 2000. Z. Reichstein was partially supported by NSF grant DMS-9801675.

Department of Mathematics
Oregon State University
Corvallis OR 97331
E-mail address: zincy@math.orst.edu

Department of Mathematics and Computer Science
University of the Negev
Be’er Sheva’
Israel

Hashofar 26/3
Ma’ale Adumim
Israel
E-mail address: youssin@math.bgu.ac.il
THE ASYMPTOTIC EXPANSION OF SPHERICAL FUNCTIONS ON SYMMETRIC CONES

P. Sawyer

In this paper, we compute the second order Taylor expansion of spherical functions on symmetric cones. In order to compute the expansion, we use an expression for the spherical function in terms of a spherical function on a symmetric cone of lower rank.

1. Introduction.

In [7], Genkai Zhang gives the asymptotic expansion for the spherical functions on symmetric cones. This is done to prove a central limit theorem for these spaces. The work of Zhang is a natural continuation of the work of Audrey Terras [6] (the case of the positive definite matrices of rank 2) and of the work of Donald St.P. Richards [3] (the case of the positive definite matrices of all ranks). In each case, the focus is to find the expansion of the spherical function $h_\lambda(e^H)$ of order 2 both in $H$ and $\lambda$. Zhang uses a generalized binomial expansion to obtain the first order terms of the expansion, and a recursion formula for the product of spherical polynomials from [8] in order to obtain the second order terms of the expansion.

We should point out the work of Piotr Graczyk in [1, 2] who also investigates the central limit theorem and the expansion of the spherical functions on symmetric matrices and, in particular, on the space of positive definite matrices.

In this paper, we prove the same result using a recurrence formula for the spherical functions on symmetric cones we obtained in [4]. The interest of this approach is its straightforwardness and the possibilities it opens, as a new method, for other symmetric spaces. In particular, we do not require a product formula to obtain the second order terms in the expansion.

In Section 2, we recall the nature of the problem and some of the notation of [7]. In Section 3, we recall our result of [4] and explain how it can be used to compute the expansion of the spherical functions. Finally, in Section 4, we find recurrence relations which describe the coefficients of our expansion. Solving these recurrence relations is straightforward.
2. The asymptotic expansion.

We have followed the notation of [7] as closely as possible.

In what follows, given a maximal orthogonal system of idempotents in the underlying Jordan algebra, \( H = \sum_{i=1}^r h_i M_{e_i} \) where \( M_x \) corresponds to multiplication by \( x \) in the Jordan algebra and \( \lambda(H) = \sum_{i=1}^r \lambda_i h_i \). We know that when \( H \) is close to zero then

\[
(1) \quad h_\lambda(e^H) = 1 + a(\lambda) \sum_{i=1}^r h_i + b(\lambda) \sum_{i=1}^r h_i^2 + P(\lambda) \sum_{i<j} h_i h_j + O(\|h\|^3).
\]

In order to prove the central limit theorem for symmetric cones, we need to compute \( a(\lambda) \) and \( b(\lambda) \). We note that the terms \( a(\lambda), b(\lambda) \) and \( P(\lambda) \) are symmetric in \( \lambda \). For convenience, we will write \( f(h) \sim g(h) \) whenever \( f(h) = g(h) + O(\|h\|^3) \) for \( H \) close to 0.

3. A recursion formula.

Taking into account the different notation, we show in [5, Theorem 5.3] that

\[
(2) \quad h_\lambda(e^H) = e^{(\lambda_1 + a(\lambda - 1/4) \sum_{i=1}^r h_i \Gamma(r a/2)(\Gamma(a/2))^{r-1})} \int_\sigma h_{\tilde{\lambda}}(e^{\tilde{\xi}})(\beta_1 \cdots \beta_r)^{a/2-1} d\beta
\]

where \( a \) is the common multiplicity of the roots, \( \sigma = \{ t = (\beta_1, \ldots, \beta_r) : 0 \leq \beta_i \leq 1, \sum_{i=1}^r \beta_i = 1 \} \). The \( \xi_i \)'s, \( 1 \leq i \leq r-1 \), are determined (modulo their order) by the relation

\[
(3) \quad \sum_{i_1 < \cdots < i_k} e^{\xi_{i_1} + \cdots + \xi_{i_k}} = \sum_{p=1}^r \beta_p \sum_{i_1 < \cdots < i_k} e^{h_{i_1} + \cdots + h_{i_k}}
\]

for \( 1 \leq k \leq r-1 \) and \( \tilde{\lambda}(\xi) = \sum_{i=1}^{r-1} (\lambda_i - \lambda_r - a r/4) \xi_i \). It is understood that when \( r = 1, h_\lambda(e^H) = e^{\lambda_1 h_1} \) and therefore that \( a(\lambda) = a(\lambda_1) = \lambda_1 \) and \( b(\lambda) = b(\lambda_1) = \lambda_1^2/2 \).

Note that computing Harish-Chandra \( c \)-function using this formulation is straightforward.

In view of (2), it is natural to point out that (1) is equivalent to

\[
(4) \quad h_\lambda(e^H) \sim 1 + (2 b(\lambda) - P(\lambda)) \sum_{i=1}^r (e^{h_i} - 1)
\]

\[
+ (b(\lambda) - a(\lambda)/2) \left( \sum_{i=1}^r (e^{h_i} - 1) \right)^2
\]

\[
+ (a(\lambda) - 2 b(\lambda) + P(\lambda)) \left( e^{\sum_{i=1}^r h_i} - 1 \right).
\]
Note that the relation $e^x - 1 = x + x^2/2 + O(x^3)$ (when $x$ is close to 0) will be used repeatedly.

**Remark 3.1.** It is clear that any Taylor expansion of $h_\lambda(e^H)$ which would be in terms of symmetric polynomials of the $h_i$’s can be expressed as an expansion of the elementary symmetric polynomials of the $e^{h_i}$’s as in (4).

Before we use (4) with $\hat{\lambda}$ and $\xi$, we must make sure that the integration in (2) preserves the relation $\sim$. It suffices to refer to [4, Corollary 2.9] to see that the relation (3) implies that the $\xi_i$’s are squeezed between $h_i$’s and hence that $O(\|\xi\|^3) = O(\|h\|^3)$.

The following result, whose proof is straightforward, will allow us to make the necessary computations.

**Lemma 3.2.** For any integers $k_i \geq 0$, $1 \leq i \leq r$, we have
$$
\frac{\Gamma(r a/2)}{(\Gamma(a/2))^r} \int_\sigma \prod_{i=1}^r \beta_i^k_i (\beta_1 \cdots \beta_r)^{a/2-1} d\beta = \frac{\Pi'_{i=1}(a/2)k_i}{(ar/2)^{\sum_i k_i}}
$$
where $(\alpha)_k = \prod_{i=1}^k (\alpha + i - 1)$ (the empty product is equal to 1).

### 4. Computing the expansion.

The idea is simple: We are to use Formula (2) with $h_\lambda(e^\xi)$ replaced by its expansion (such as the one in (4)) and then use (3) and Lemma 3.2 to do the rest.

Once we use (4) with $\hat{\lambda}$ and $\xi$, the integrand in (2) becomes
$$
h_\lambda(e^\xi) \sim 1 + (2b(\hat{\lambda}) - P(\hat{\lambda})) \sum_{p=1}^r \beta_p \sum_{i \neq p} (e^{h_i} - 1)
$$
$$
+ (b(\hat{\lambda}) - a(\hat{\lambda})/2) \left( \sum_{p=1}^r \beta_p \sum_{i \neq p} (e^{h_i} - 1) \right)^2
$$
$$
+ (a(\hat{\lambda}) - 2b(\hat{\lambda}) + P(\hat{\lambda})) \sum_{p=1}^r \beta_p \left( e^{\sum_{i \neq p} h_i} - 1 \right)
$$
$$
\sim 1 + (2b(\hat{\lambda}) - P(\hat{\lambda})) \sum_{p=1}^r \beta_p \left( \sum_{i \neq p} h_i + \sum_{i \neq p} h_i^2/2 \right)
$$
$$
+ (b(\hat{\lambda}) - a(\hat{\lambda})/2) \left( \sum_{p=1}^r \left( \sum_{i \neq p} \beta_i \right)^2 h_p^2 \right)
$$
$$
+ 2 \sum_{p < q} \left( \sum_{i \neq p} \beta_i \right) \left( \sum_{j \neq q} \beta_j \right) h_p h_q \)
\[ + (a(\hat{\lambda}) - 2b(\hat{\lambda})) \]
\[ + P(\hat{\lambda}) \sum_{p=1}^{r} \beta_p \left( \sum_{i \neq p} h_i + \sum_{i \neq p} h_i^2/2 + \sum_{i < j, i \neq p, j \neq p} h_i h_j \right) \]

for \( t = (\beta_1, \ldots, \beta_r) \in \sigma \), and by taking (3) into account. With Lemma 3.2, integrating this last expression is straightforward. It is important to remember that \( \sum_{p=1}^{r} \beta_p = 1 \). From (2),

\[ h_{\lambda}(e^H) \sim \left( 1 + (\lambda_r + a(r-1)/4) \sum_{i=1}^{r} h_i + (\lambda_r + a(r-1)/4)^2 \sum_{i=1}^{r} h_i^2/2 \right. \]
\[ + (\lambda_r + a(r-1)/4)^2 \sum_{i<j} h_i h_j \]
\[ \cdot \left[ 1 + (2b(\hat{\lambda}) - P(\hat{\lambda})) \sum_{p=1}^{r} \frac{1}{r} \left( \sum_{i \neq p} h_i + \sum_{i \neq p} h_i^2/2 \right) \right. \]
\[ + (b(\hat{\lambda}) - a(\hat{\lambda})/2) \left( \frac{(r-1)(a(r-1)+2)}{r(ar+2)} \sum_{p=1}^{r} h_p^2 \right. \]
\[ + 2 \frac{a(r-1)^2 + 2(r-2)}{r(ar+2)} \sum_{p<q} h_p h_q \]
\[ + (a(\hat{\lambda}) - 2b(\hat{\lambda}) + P(\hat{\lambda})) \sum_{p=1}^{r} \frac{1}{r} \left( \sum_{i \neq p} (h_i + h_i^2/2) + \sum_{i < j, i \neq p, j \neq p} h_i h_j \right) \]
\[ \left. \right] \]
\[ \sim 1 + a(\lambda) \sum_{i=1}^{r} h_i + b(\lambda) \sum_{i=1}^{r} h_i^2 + P(\lambda) \sum_{i<j} h_i h_j. \]

This gives us
\[ (5) \quad a(\lambda) = \frac{r-1}{r} a(\hat{\lambda}) + \lambda_r + a(r-1)/4, \]
\[ (6) \quad b(\lambda) = \left[ \frac{a/2}{ar+2} + \lambda_r + a(r-1)/4 \right] \frac{r-1}{r} a(\hat{\lambda}) \]
\[ + b(\hat{\lambda}) \frac{(r-1)(a(r-1)+2)}{r(ar+2)} + \frac{(\lambda_r + a(r-1)/4)^2}{2}, \]
\[ P(\lambda) = \frac{r-2}{r} P(\hat{\lambda}) + (\lambda_r + a(r-1)/4)^2 + (2b(\hat{\lambda}) - a(\hat{\lambda})) \frac{a}{r(ar+2)} \]
\[ + 2 (\lambda_r + a(r-1)/4) \frac{r-1}{r} a(\hat{\lambda}), \]
which leads us to the main result of the paper.

**Theorem 4.1.** When \( r = 1 \), \( a(\lambda) = a(\lambda_1) = \lambda_1 \) and \( b(\lambda) = b(\lambda_1) = \lambda_1^2/2 \) while \( P(\lambda) = 0 \). For \( r \geq 2 \),

\[
a(\lambda) = \frac{1}{r} \sum_{i=1}^{r} \lambda_i,
\]

\[
b(\lambda) = \frac{2 + a}{2r(2 + ar)} \sum_{i=1}^{r} \lambda_i^2 + \frac{a}{r(2 + ar)} \sum_{i < j} \lambda_i \lambda_j - \frac{a^2 (r^2 - 1)}{48 (2 + ar)},
\]

\[
P(\lambda) = \frac{a}{r(2 + ar)} \sum_{i=1}^{r} \lambda_i^2 + \frac{2 (a (r - 1) + 2)}{r (r - 1) (2 + ar)} \sum_{i < j} \lambda_i \lambda_j + \frac{a^2 (r + 1)}{24 (ar + 2)}.
\]

**Proof.** Using induction and (5), we find easily that \( a(\lambda) = (\sum_{i=1}^{r} \lambda_i) / r \). From (6), we clearly have \( b(\lambda) = a_r + \sum_{i=1}^{r} \lambda_i + \delta_r \sum_{i=1}^{r} \lambda_i^2 + \gamma_r \sum_{i < j} \lambda_i \lambda_j \) with \( \alpha_1 = \beta_1 = \gamma_1 = 0 \) and \( \delta_1 = 1/2 \). By considering the coefficients of \( \lambda_i^2 \), \( \lambda_1, \lambda_1 \lambda_2 \) and the constant term in (6), we derive the following recurrence relations:

\[
\delta_r = \frac{(r - 1) (a (r - 1) + 2)}{r (ar + 2)} \delta_{r-1},
\]

\[
\gamma_r = \frac{(r - 1) (a (r - 1) + 2)}{r (ar + 2)} \gamma_{r-1},
\]

\[
\beta_r = \left[ \frac{a/2}{ar + 2} + a (r - 1)/4 \right] \frac{1}{r}
\]

\[
+ \frac{(r - 1) (a (r - 1) + 2)}{r (ar + 2)} \left[ \beta_{r-1} - a r \delta_{r-1}/2 - (r - 2) a r \gamma_{r-1}/4 \right],
\]

\[
\alpha_r = -\frac{a}{4} (r - 1) \left[ \frac{a/2}{ar + 2} + a (r - 1)/4 \right]
\]

\[
+ \frac{(r - 1) (a (r - 1) + 2)}{r (ar + 2)} \left[ \alpha_{r-1} - a r (r - 1)/4 \beta_{r-1} \right]
\]

\[
+ \frac{a^2 (r - 1)^2}{16} \delta_{r-1} + \frac{a^2 r^2 (r - 1) (r - 2)}{32} \gamma_{r-1} + a^2 (r - 1)^2/32.
\]

All these equations are valid for \( r \geq 2 \) except the second one which is valid only for \( r \geq 3 \). One can easily find \( \delta_r \) and \( \gamma_r \). This in turn brings us to

\[
\beta_r = \frac{(r - 1) (a (r - 1) + 2)}{r (ar + 2)} \beta_{r-1}
\]

for \( r \geq 2 \) which means that \( \beta_r = 0 \) for all \( r \). Finally, we have

\[
\alpha_r = \frac{1}{16} \frac{(r - 1) (a (r - 1) + 2)}{r (ar + 2)} \alpha_{r-1} - \frac{1}{16} \frac{a^2 r (r - 1)}{r (ar + 2)}
\]
whose solution is straightforward. We compute \( P(\lambda) \) in much the same way.

\[ \Box \]

**Remark 4.2.** There is a minor computation error in the last term of [7, Page 573] which is why we obtain a different value for \( \alpha_r \) and therefore for \( b(\lambda) \). The term \( P(\lambda) \) is not computed in [3, 6, 7]. It is given in [1] in the case of the real positive definite matrices \( (a = 1) \) and in [2] in the case of the hermitian positive definite matrices \( (a = 2) \).

**Conclusion.**

This approach suggests, that once we express a spherical function on a symmetric space in terms of spherical functions on a symmetric space of the same type but lower rank, we become able to compute its expansion. This approach could be applied to the other symmetric spaces which correspond to the classical Lie algebras.

**References**


Received October 29, 1999. This research was supported by a grant from NSERC.

**Laurentian University**  
**Sudbury, Ontario**  
**Canada P3E 5C6**  
**E-mail address:** sawyer@cs.laurentian.ca
Guidelines for Authors

Authors may submit manuscripts at pjm.math.berkeley.edu/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use \LaTeX, but papers in other varieties of \TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as \LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of Bib\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.

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.
Multipliers of operator spaces, and the injective envelope
David P. Blecher and Vern I. Paulsen

Existence and nonexistence of interior-peaked solution for a nonlinear
Neumann problem
Daomin Cao, E.S. Noussair and Shusen Yan

Alexander and Thurston norms of fibered 3-manifolds
Nathan M. Dunfield

Sur les opérateurs nilpotents à images des itérés fermées dans un espace de
Banach
Abdelkalek Faouzi

Critical-exponent Sobolev norms and the Slice Theorem for the quotient
space of connections
Paul M.N. Feehan

Techniques for approaching the dual Ramsey property in the projective
hierarchy
Lorenz Halbeisen and Benedict Löwe

Elliptic surfaces and ample vector bundles
Antonio Lanteri and Hidetoshi Maeda

Quasihomoeomorphisms and univalent harmonic mappings onto punctured
bounded convex domains
Abdallah Lyzzaik

Vassiliev invariants of knots in a spatial graph
Yoshiyuki Ohryama and Kouki Taniyama

Splitting fields of G-varieties
Zinovy Reichstein and Boris Youssin

The asymptotic expansion of spherical functions on symmetric cones
P. Sawyer