JOURNAL OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 1, Number 3, July 1988

REGULARITY PROPERTIES OF THE $\overline{\partial}_b$ EQUATION ON WEAKLY PSEUDOCONVEX CR MANIFOLDS OF DIMENSION 3

MICHAEL CHRIST

To the memory of José Luis Rubio de Francia

1. INTRODUCTION AND STATEMENT OF RESULTS

Let M be a compact, C^∞ CR manifold of dimension 3 over $\mathbf{R}$. Associated to the CR structure is a first-order differential operator, $\overline{\partial}_b$, on M. We study the regularity properties, in terms of L^p Sobolev and Hölder norms, of the equation $\overline{\partial}_b u = f$.

M is said to be CR if there is given a C^∞ sub-bundle, denoted $T^{1,0}M$, of the complexified tangent bundle TM, such that each fiber $T_x^{1,0}M$ is of dimension 1 over $\mathbf{C}$, and $T_x^{1,0}M \cap \overline{T_x^{1,0}M} = \{0\}$ for all $x \in M$. Define $T^{0,1}M = \overline{T^{1,0}M}$ and let $B^{0,1}M$ denote its dual bundle. Then for any C^∞ function u on M, $\overline{\partial}_b u$ is the smooth section of $B^{0,1}M$ obtained by restricting du to $T^{0,1}M$. In other words if $\overline{Z} \in T_x^{0,1}M$ then $(\overline{\partial}_b u)(\overline{Z}) = (\overline{Z}u)(x)$.

The boundary M of any smoothly bounded relatively compact open set $\Omega \subset \mathbf{C}^2$ carries a natural CR structure: $T_x^{1,0}M$ is the subspace of the tangent space of $\mathbf{C}^2$ at x consisting of all holomorphic tangent vectors, that is, linear combinations of $\frac{\partial}{\partial z_1}, \frac{\partial}{\partial z_2}$, which are tangent to M. A prime motivation for the study of $\overline{\partial}_b$ is its connection with complex analysis on Ω, in the case $M = \partial\Omega$. In order to construct holomorphic functions in Ω one often needs to solve $\overline{\partial} u = \alpha$ in Ω, where α is a given $(0,1)$ form satisfying the necessary condition $\overline{\partial}\alpha = 0$. It is desirable to have as much control over the regularity of some solution u as possible. Currently much more is known in terms of L^2 and L^2 Sobolev norms than L^p, L^p Sobolev and Hölder norms. In particular the algebra of functions holomorphic on Ω and continuous on $\overline{\Omega}$ is of interest, so one seeks conditions on α which guarantee the existence of a continuous solution u. J. J. Kohn has pointed out [K3] that if it were proved that

Received by the editors October 28, 1987.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 35H05, 32F30; Secondary 35N15, 35F99.

Key words and phrases. CR manifold, finite type, $\overline{\partial}_b$, subellipticity, Szegö projection, free nilpotent group, parametrix, microlocal analysis, singular integral operator.

Research supported in part by the National Science Foundation and the Mathematical Sciences Research Institute. Alfred P. Sloan fellow.

for any continuous section g of $B^{0,1}M$ belonging to the range of $\overline{\partial}_b$, there existed a continuous (respectively Hölder continuous) solution v of $\overline{\partial}_b v = g$, then it would follow from standard elliptic theory that for any $\overline{\partial}$-closed $(0,1)$ form α continuous on $\overline{\Omega}$, there exists a solution u continuous (respectively Hölder continuous) on $\overline{\Omega}$. For pseudoconvex domains Ω of finite type we will establish Hölder regularity for certain canonical solutions of $\overline{\partial}_b$ in this paper.

$\overline{\partial}_b$ is also of some interest in its own right as a partial differential equation, for it is in general neither locally solvable nor hypoelliptic and has an infinite-dimensional kernel. In case $M = \partial\Omega \subset \mathbf{C}^2$ this kernel is again of significance for complex analysis, since (sufficiently regular) functions in the kernel are the boundary values of holomorphic functions on Ω. A tool for analysis of the kernel is the Szegö projection, the orthogonal projection of $L^2(M)$ onto the kernel, analogous to the Cauchy projection from L^2 onto H^2 on the unit circle in $\mathbf{C}^1$. Therefore we shall also study the Szegö projection.

We say that $\overline{\partial}_b$ has closed range if whenever f is an L^2 section of $B^{0,1}$ such that there exists a sequence $u_n \in L^2$ such that $\overline{\partial}_b u_n \in L^2$ for all u and $\|\overline{\partial}_b u_n - f\|_{L^2} \to 0$, then there exists $u \in L^2$ satisfying $\overline{\partial}_b u = f$ and $\|u\|_{L^2} \le C\|f\|_{L^2}$, where C is a constant independent of f. By the range of $\overline{\partial}_b$ we shall mean the space of all such $f \in L^2$. Let $\mathscr{H}_b = \{u \in L^2 : \overline{\partial}_b u = 0\}$ be the kernel of $\overline{\partial}_b$ in L^2. Fix a positive measure on M, given in local coordinates by a C^∞, nonvanishing density. Henceforth the L^2 norm will be defined with respect to that measure. When $M = \partial\Omega \subset \mathbf{C}^2$ and Ω is pseudoconvex, $\overline{\partial}_b$ necessarily has closed range [K1, BS].

Since the complex bundle $B^{0,1}$ has dimension one, in local coordinates in a neighborhood of some point x_0 in M we may regard $\overline{\partial}_b u$ as a function, and we may then write $\overline{\partial}_b u = (X + iY)u$ where X, Y are real, C^∞ vector fields, linearly independent at every point. Choose a third C^∞ real vector field T, so that at each point near x_0, $\{X, Y, T\}$ forms a basis for the tangent space. We say that M is pseudoconvex if for each $x_0 \in M$ and any choice of such a frame near x_0, $[X, Y](x) = \lambda(x)T + a(x)X + b(x)Y$ where λ does not change sign in some neighborhood of x_0, in other words that $\lambda \ge 0$ for all x or $\lambda \le 0$ for all x. M is said to be of finite type if X, Y satisfy the condition of Hörmander [H1] and Kohn [K4], that is, that they together with their iterated commutators $[X, Y], [X, [X, Y]], \ldots$ of all orders span TM at each point. Both these conditions are independent of the choice of local coordinate system and of T. M is called strongly pseudoconvex if λ is never zero. Consider a point in the coordinate chart, and consider the commutator of X, Y with the fewest factors such that X, Y and it span the tangent space to $\mathbf{R}^3$ at that point. Let m be the supremum, over all points in M, of the number of factors in such minimal-length commutators. We say that M is of type m. Thus $m = 2$ in the strongly pseudoconvex case.

In general since $\overline{\partial}_b$ is assumed to have closed range, for each L^2 function f in the range there exists a unique $u \in L^2$ satisfying

$$\begin{cases} \overline{\partial}_b u = f \\ u \perp \mathscr{H}_b . \end{cases} \tag{0}$$

Our main result is

Theorem A. *Let M be pseudoconvex of finite type and let $\overline{\partial}_b$ have closed range. Then for any $f \in \mathrm{Range}(\overline{\partial}_b)$, if $u \in L^2$ satisfies (0), then for any $p \in (1, \infty)$ and $s \geq 0$, for any open set U such that $f \in L^p_s(U)$, u belongs to $L^p_{s+m^{-1}}$ on a neighborhood of every compact subset K of U.*

For $s \in [0, \infty)$, L^p_s denotes the Sobolev space of all L^p functions which are differentiable of order s in L^p. Because of the form of the hypotheses and conclusions in the theorem, we may assume that U has a smooth boundary, so no technical difficulties arise in defining $L^p_s(U)$.

It follows at once that if $f \in L^\infty(U)$ then u is Hölder continuous of order $m^{-1} - \varepsilon$ on every compact subset of U, for all $\varepsilon > 0$. In a subsequent paper [C3] we show that this holds with $\varepsilon = 0$ as well.

In closely related work Fefferman and Kohn [FK] have proved that $f \in \Lambda_s(U) \Rightarrow u \in \Lambda_{s+m^{-1}}(K)$ (provided $s + m^{-1}$ is not an integer) where Λ_s denotes the space of functions Hölder continuous of order s (note that taking p very large in our result yields $u \in \Lambda_{s+\varepsilon}(K)$ for all ε strictly less than m^{-1}). It may be that the sharp order of Hölder regularity can also be obtained by our method, but we have not verified this. The first step in both approaches is an idea of Kohn which involves microlocal analysis and a reduction to the study of a different operator A, which is in some respects better behaved than $\overline{\partial}_b$. But their work differs substantially from ours in the particular operator A studied—one has some flexibility in choosing A—and in the technique used to analyze it. It should be pointed out that as it stands, our technique does not appear to be applicable to the analysis of $\overline{\partial}_b$ in higher dimensions, while that of [FK] may have more promise.

Denote by S the Szegö projection. Let $G : \mathrm{Range}(\overline{\partial}_b) \to L^2$ be the operator which to each f assigns the solution u of (0). Then $S = I - G \circ \overline{\partial}_b$, where I denotes the identity operator. Thus properties of S are closely linked to the regularity of (0). S is bounded on L^2 since it is an orthogonal projection. Our second main result, Theorem C, formulated in §12, gives pointwise upper bounds for the Szegö kernel, the distribution-kernel for S, which together with the machinery of Calderón-Zygmund analysis on spaces of homogeneous type imply

Theorem B. *Let M be pseudoconvex of finite type and let $\overline{\partial}_b$ have closed range. Then the Szegö projection S extends to an operator bounded on $L^p(M)$ for all $p \in (1, \infty)$.*

The L^2 Sobolev regularity conclusions of Theorem A are due to Kohn [K1]. Detailed results in the strongly pseudoconvex case have been obtained by Greiner and Stein [GS]. Our Theorems B and C are generalizations of results by several authors for special cases. In the strongly pseudoconvex case, Fefferman [F] and Boutet de Monvel and Sjöstrand [BMS] obtained asymptotics for the Szegö kernel. See also Greiner and Stein [GS]. For higher-dimensional strongly pseudoconvex CR manifolds the analogous result was obtained by Folland and Stein [FSt1] (it is implicit in their estimates for the fundamental solution of the Kohn Laplacian). In certain special weakly pseudoconvex cases pointwise bounds for the Szegö kernel and its derivatives have been obtained by Nagel, Stein, and Wainger [N], Machedon [M1], and Nagel, Rosay, Stein, and Wainger [NRSW1]. The general case of a finite, pseudoconvex boundary in $\mathbf{C}^2$ has more recently been treated by the latter group of authors [NRSW2]. Even more recently analogous estimates have been obtained in certain higher-dimensional cases by Machedon [M2]. In all these cases the pointwise bounds imply that S is a "singular integral" operator in a natural sense, and the L^p boundedness follows from an extension of the Calderón-Zygmund theory.

The $\overline{\partial}_b$ equation presents immediately three difficulties. First, it is not even locally solvable [L, H3]. Therefore we impose the global condition that f belongs to the range. Second, $\overline{\partial}_b$ will typically have an infinite-dimensional kernel, for in the case $M = \partial\Omega$, $\Omega \subset \mathbf{C}^2$, $\overline{\partial}_b$ annihilates restrictions to M of arbitrary holomorphic functions. Hence there can be no regularity results for arbitrary solutions u. Consequently following Kohn, we impose the global condition that $u \perp \mathscr{H}_b$ to single out a canonical solution. Third, the $\overline{\partial}_b$ equation is at best subelliptic, rather than elliptic. In local coordinates with $\overline{\partial}_b u = Xu + iYu$, even if one had control over Xu and Yu separately, one would still lack direct control over Tu. The finite type hypothesis is needed to guarantee that X and Y indirectly do control T; without it one could not hope to gain derivatives as in Theorem A.

The analysis consists of three principal steps. The first is the idea of Kohn [K3] on which everything rests. It is well understood how subelliptic operators like the subLaplacian $\triangle = -(X^2 + Y^2)$ may be inverted [RS, NSW, FS, S] modulo a smoothing, compact error. For $\overline{\partial}_b$ we need to find the relative inverse G, which only inverts $\overline{\partial}_b$ modulo the infinite-dimensional kernel $\mathscr{H}_b$. Kohn showed how a microlocal analysis could be used to determine G from the inverse of any of a family of related operators $A = \triangle + B$. In this paper we construct and study a parametrix for one A in this family, inverting A modulo a smoothing error term. The advantage of A is that unlike $\overline{\partial}_b$, it is (in a sense) hypoelliptic and has a finite-dimensional kernel. Unfortunately this requires some analysis to prove, for the extra term B is rather poorly behaved. In local coordinates it is related to $|T|$—the absolute value of the operator T, defined via Fourier transformation in coordinates in which T is a coordinate vector field. This operator is not even pseudolocal, that is, applying it to a

function which happens to be C^∞ on some open set need not produce another function C^∞ on that set. In fact the distribution-kernel for $|T|$ is a singular measure away from the diagonal. Moreover any parametrix must necessarily reflect the singularities present in B.

The overall strategy for constructing the parametrix is taken from the work of Folland, Rothschild, and Stein [FSt1, RS]. Step number 2 is to study a model operator $\widehat{A}$, a convolution operator on a certain free nilpotent Lie group. A general result on the convolution kernels for inverses of certain classes of singular integral operators on such groups, obtained in [C2], may after some manipulation be applied to $\widehat{A}$ and yields just enough control on the regularity of the kernel for $\widehat{A}^{-1}$ for our purpose.

In the third step we follow the technique developed by Folland, Rothschild, and Stein. To construct a parametrix for $-(X^2+Y^2)$, where X, Y satisfy the finite type hypothesis, they first associate to it a left-invariant, homogeneous differential operator $-(\widehat{X}^2+\widehat{Y}^2)$ on an appropriate homogeneous nilpotent Lie group. Understanding how to invert this model operator, they use its inverse and a freezing-of-coefficients procedure to produce a parametrix for $-(X^2+Y^2)$. We use the convolution kernel for the inverse of $\widehat{A}$ to invert A, in exactly the same spirit. However because of the singularities present in our problem the procedure of [RS] does not apply in the usual way. We are led to introduce a technical modification which leads to unfortunate complication.

The plan of the paper is as follows. In §2 we carry out Kohn's reduction of the problem to the study of A. We show how Theorems A and B follow, once A is known to have a parametrix with certain specific mapping properties on L^p and Sobolev spaces. In fact we obtain a sharper version of Theorem A which asserts that u is actually one full derivative smoother than f in the "good" directions X and Y. In §3 we discuss the absolute value of any smooth vector field, an operator which figures prominently in our analysis. We prove a lemma showing how somewhat more general operators may be related to vector fields. In §4 we show how the lifting procedure of [RS] may be applied to our problem. Then step 2 is carried out in §5 using the result of [C2]. $\widehat{A}$ is not precisely the type of operator considered there, so this section contains the necessary reduction and the translation of the general information obtained there into appropriate control of the convolution kernel for $\widehat{A}^{-1}$.

Our parametrix for A is defined in §6. Its smoothing properties are analyzed in §§7 through 10. In §§11 and 12 we state and prove the pointwise bounds on the Szegö kernel.

There is a large literature relevant to these matters. See the bibliography and introduction of [FK].

A word on notation. We write "$A \prec B$" to indicate that A is majorized by some constant times B. Such constants are permitted to depend on M, but frequently A and B will depend on other parameters, and the constant in the inequality is to be independent of these parameters. $\|f\|_q$ denotes the

L^q norm. Constants C may change in value from one occurrence to the next, while those denoted C_j retain a fixed value for a short part of the argument but may later be redefined.

This research began as joint work with J. J. Kohn, whom I thank for encouragement and his generosity. I am also indebted to C. Fefferman and A. Nagel for encouraging conversations.

2. Reduction to A

Introduce a smoothly varying family of inner product structures on the fibers of the bundle $B^{0,1}$. Let $\overline{\partial}_b^*$ be the adjoint of $\overline{\partial}_b$, determined uniquely by this inner product structure on $B^{0,1}$ and our fixed measure on M. $\overline{\partial}_b^*$ maps smooth sections of $B^{0,1}$ to functions. In local coordinates in which

$$\overline{\partial}_b = X + iY,$$

$\overline{\partial}_b^*$ is $X - iY + c$, where c denotes multiplication by some C^∞ function. Since $\overline{\partial}_b$ has closed range, $\overline{\partial}_b^*$ does also, and for each $u \in L^2$ orthogonal to $\mathscr{H}_b$ there exists an L^2 section v of $B^{0,1}$ such that $u = \overline{\partial}_b^* v$ and $\|v\|_2 \le C\|u\|_2$. Thus following Kohn, we may write the equation as

$$\overline{\partial}_b \overline{\partial}_b^* v = f$$

where it is given that $\|v\|_2 + \|\overline{\partial}_b^* v\|_2 \le C\|f\|_2$, remembering always that it is $\overline{\partial}_b^* v$, rather than v itself, whose regularity is to be studied.

It is a theorem of Rothschild and Stein [RS] that if u, Xu, Yu belong to L_s^p on some open set, then u belongs to $L_{s+m^{-1}}^p$ on any compact subset, where m is the type as defined in §1. Therefore Theorem A follows from

Theorem A′. *Let M be pseudoconvex of finite type and let $\overline{\partial}_b$ have closed range. Let U be any coordinate chart as above. Suppose that $f \in \operatorname{Range}(\overline{\partial}_b)$, that $u \in L^2$ satisfies* (0), *and that $f \in L_s^p(U)$ for some $p \in (1, \infty)$ and $s \ge 0$. Then Xu and Yu belong to L_s^p on every compact subset of U.*

In this section we introduce an operator A and show how Theorem A′ would follow if a suitable parametrix for A could be found.

Working in a coordinate system of the type described we shall write $\overline{L}$ for $X + iY = \overline{\partial}_b$. $\overline{L}^*$ will denote $X - iY + c = \overline{\partial}_b^*$. Fix such a coordinate system about some $x_0 \in M$ in which $x_0 = 0, \overline{\partial}_b = X + iY = \overline{L}, X(0) = \frac{\partial}{\partial x_1}, Y(0) = \frac{\partial}{\partial x_2}, T(x) \equiv \frac{\partial}{\partial x_3}$ at all points x near 0, and

$$[X, Y](x) = \lambda(x)T + a_1(x)X + a_2(x)Y$$

where $\lambda \ge 0$. We shall make use of three classes of microlocalizing pseudodifferential operators, elements of which classes are denoted P^0, P^-, and P^+. Each such P will be a composition, the first factor being multiplication by a C^∞ function φ supported in a small neighborhood of 0, and the second a Fourier multiplier with symbol $\sigma(\xi), \xi$ denoting the variable dual to

x. Each σ will be C^∞ and will be homogeneous of degree 0 for $|\xi| \geq C > 0$. σ^0, the symbol for the second factor of P^0, is to be supported in $\{\xi : |\xi| \leq C_0 \text{ or } |\xi_3| \leq C_1|\xi'|\}$ where $\xi' = (\xi_1, \xi_2)$. σ^+ is to be supported in $\{|\xi| \geq C_2, \xi_3 > 0 \text{ and } \xi_3 \geq C_3|\xi|\}$ and σ^- in $\{|\xi| \geq C_4, \xi_3 < 0 \text{ and } -\xi_3 \geq C_5|\xi|\}$, for some $0 < C_0, C_1, C_2, C_3, C_4, C_5 < \infty$. These C_i are permitted to depend on the particular operators P^0, P^-, P^+, and the support of φ must be sufficiently small relative to them, but only finitely many such operators will be required in the proof. Let us permit a symbol such as φ to denote either a smooth function or the multiplication operator $f \to \varphi f$. Then it is possible to choose P^0, P^-, P^+ which sum to φ, where $\varphi \equiv 1$ in some neighborhood of 0.

All of these symbols belong to the standard symbol class $S^0_{1,0}$, and of course X and Y belong to $S^1_{1,0}$. Thus the standard symbolic calculus of pseudodifferential operators may be freely applied. Given an operator in one of our classes, we shall systematically denote by $\widetilde{P}$ another operator in the same class, constructed so that $\tilde{\varphi}\varphi \equiv \varphi$ and $\tilde{\sigma}\sigma \equiv \sigma$. By the symbolic calculus, $P = P \circ \widetilde{P} + S^\infty$ where S^∞ will denote an operator smoothing of infinite order.

Suppose $\overline{L}\,\overline{L}^* v = f$ in a neighborhood of x_0, with $v, u = \overline{L}^* v \in L^2$. We analyze the regularity of $P^0 u, P^+ u, P^- u$ by separate arguments. For P^0 the reasoning is simple. It is based on the identity

$$\overline{L}\,\overline{L}^* P^0 v = P^0 f + O(X\widetilde{P}^0 v, Y\widetilde{P}^0 v, \widetilde{P}^0 v) + S^\infty v. \tag{2.1}$$

We write "$f = O(g)$" if for any $p \in (1, \infty)$, $\|f\|_p \prec \|g\|_p$. To derive (2.1) write

$$\begin{aligned}\overline{L}\,\overline{L}^* P^0 v &= \overline{L} P^0 \overline{L}^* v + \overline{L}[\overline{L}^*, P^0]v \\ &= P^0 \overline{L}\,\overline{L}^* v + [\overline{L}, P^0]\overline{L}^* v + \overline{L}[\overline{L}^*, P^0]v \\ &= P^0 f + [\overline{L}, P^0]\overline{L}^* v + \overline{L}[\overline{L}^*, P^0]v.\end{aligned}$$

The symbolic calculus implies that

$$[\overline{L}, P^0] = [\overline{L}, P^0] \circ \widetilde{P} + S^\infty = O(\widetilde{P}^0) + S^\infty$$

and

$$\begin{aligned}\overline{L}[\overline{L}^*, P^0] &= [\overline{L}^*, P^0]\overline{L} + [\overline{L}, [\overline{L}^*, P^0]] \\ &= [\overline{L}^*, P^0]\overline{L}\widetilde{P}^0 + [\overline{L}, [\overline{L}^*, P^0]] + S^\infty \\ &= O(\overline{L}\widetilde{P}^0) + O(\widetilde{P}^0) + S^\infty.\end{aligned}$$

The boundedness of $[\overline{L}^*, P^0]$ and $[\overline{L}, [\overline{L}^*, P^0]]$ on L^p_s, for all $p \in (1, \infty)$ and $s \in \mathbf{R}$, follows from the symbolic calculus, since any operator with symbol in $S^0_{1,0}$ is bounded on L^p for all $p \in (1, \infty)$. (2.1) now follows.

Since

$$\overline{L}\,\overline{L}^* = (X + iY)(X - iY + \text{order } 0) = X^2 + Y^2 + \text{order one},$$

its symbol at $x = 0$ is $-\xi_1^2 - \xi_2^2 = -|\xi'|^2$ modulo a first-order term. On $\{\xi : \sigma^0(\xi) \neq 0\}$ this symbol is elliptic, since $1 + |\xi'| \succ 1 + |\xi|$ there, so we may construct a second-order elliptic pseudodifferential operator A' whose symbol is in $S^2_{1,0}$ and agrees with that of $\overline{L}\overline{L}^*$ on the microlocal support of P^0. Thus $A' \circ P^0 = \overline{L}\overline{L}^* P^0 + S^\infty$. Since A' is elliptic there exists $\mathscr{P}'$ such that $\mathscr{P}'A' = I + S^\infty$ on functions supported in a small neighborhood of 0, with $\mathscr{P}'$ smoothing of order 2, that is, its composition with any differential operator of order two is bounded on L^p_s. Here and elsewhere in the paper I denotes the identity operator.

Now (2.1) gives

$$P^0 v = \mathscr{P}' P^0 f + \mathscr{P}'(O(X\widetilde{P}^0 v, Y\widetilde{P}^0 v, \widetilde{P}^0 v)) + S^\infty v. \tag{2.2}$$

We have applied the identity $\mathscr{P}'A' = I + S^\infty$, which holds for functions supported in a small neighborhood of 0, to $P^0 v$, which satisfies no such support restriction. However because of the cutoff function φ in the definition of P^0, $P^0 v$ will be C^∞ outside a small neighborhood of 0. Therefore the identity $\mathscr{P}'A' = I + S^\infty$ remains valid when both sides are applied to $P^0 v$, with an additional term smoothing of infinite order added to the error term S^∞. We find that $P^0 v \in L^2_1$, since $\mathscr{P}' \circ (X$ or Y or $I)$ is smoothing of order one and v is already known to be in L^2. Therefore $\widetilde{P}^0 v \in L^2_1$, since P^0 could have been chosen to be any element of its class, and a second application of (2.2) establishes that $P^0 v \in L^2_2$. Thus we have elliptic regularity microlocally on the support of P^0; arguing in the same way we find that $f \in L^p_s \Rightarrow P^0 v \in L^p_{s+2} \Rightarrow P^0 u \in L^p_{s+1}$ as desired.

For $P^\pm u$ it cannot be possible to obtain elliptic regularity, and we aim only for a small gain in the order of differentiability. Define

$$|T| f(x) = \int e^{ix\cdot\xi} |\xi_3| \hat{f}(\xi)\, d\xi$$

for any f. Define further

$$\Delta = -(X^2 + Y^2)$$

and

$$A = \Delta + \lambda(x)|T|.$$

As usual $\lambda(x)$ denotes the multiplication operator $f \to \lambda f$. We shall demonstrate that $P^\pm u$ may be controlled, so that Theorem A′ follows, if a parametrix $\mathscr{P}$ for A can be found with certain mapping properties.

We say that an operator B is smoothing if there exists $\delta > 0$ such that B is bounded from L^p to L^p_δ whenever $p \in (1, \infty)$. The parametrix $\mathscr{P}$ will be required to satisfy

$$\mathscr{P}A = I - \mathscr{E}$$

on functions supported in a small neighborhood of $x_0 = 0$, where

$$\mathscr{P}, X\mathscr{P}, Y\mathscr{P}, \mathscr{P}X, \mathscr{P}Y, \mathscr{E}, X\mathscr{E}, Y\mathscr{E} \tag{2.3}$$

are all smoothing and

$$(X \text{ or } Y) \circ \mathscr{P} \circ (X \text{ or } Y), \quad \mathscr{P} \circ (X \text{ or } Y) \circ (X \text{ or } Y) \tag{2.4}$$

are bounded on L^p for all $p \in (1, \infty)$. Moreover $\mathscr{P}$ and $\mathscr{E}$ should map C^∞ to itself boundedly. These conditions are enough to permit us to conclude that u is δ derivatives smoother than f in Theorem A, for some $\delta > 0$. In order to show that δ may be taken to be the sharp value, m^{-1}, we need

$$(X \text{ or } Y) \circ (X \text{ or } Y) \circ \mathscr{P} \tag{2.4$'$}$$

to be bounded on L^p for all $p \in (1, \infty)$ and

$$(X \text{ or } Y) \circ (X \text{ or } Y) \circ \mathscr{E} \tag{2.4$''$}$$

to be bounded from L^p_ε to L^p for all $p \in (1, \infty)$ and $\varepsilon > 0$.

For $P^0 u$ the regularity promised in Theorem A$'$ is already established, when $s = 0$. To analyze $P^- u$ using a parametrix satisfying (2.3) and (2.4), it is not necessary to invoke the representation $u = \overline{L}^* v$. The fundamental identity for P^- is

$$\begin{aligned} P^- u = &-\mathscr{P}\overline{L}^* P^- f - \mathscr{E} P^- u \\ &+ \mathscr{P} X O(\widetilde{P}^- u) + \mathscr{P} Y O(\widetilde{P}^- u) + \mathscr{P} O(\widetilde{P}^- u) + \mathscr{P} S^\infty u. \end{aligned} \tag{2.5}$$

To derive it write

$$\begin{aligned} \overline{L}^* \overline{L} &= \tfrac{1}{2}(\overline{L}^* \overline{L} + \overline{L}\,\overline{L}^*) + \tfrac{1}{2}[\overline{L}^*, \overline{L}] \\ &= -\triangle - i[X, Y] + O(X, Y, I) \\ &= -\triangle - i\lambda(x) T + O(X, Y, I). \end{aligned}$$

On the support of the symbol of P^-, the symbol of $T = \frac{\partial}{\partial x_3}$ is $i\xi_3 = -i|\xi_3|$, so $T \circ P^- \equiv -i|T| \circ P^-$, recalling that P^- is a composition in which the second factor acting is a Fourier multiplier operator. Thus

$$\overline{L}\,\overline{L}^* \circ P^- = -A \circ P^- + O(X, Y, I) \circ P^-. \tag{2.6}$$

On the other hand

$$\begin{aligned} \overline{L}^* \overline{L} P^- u &= \overline{L}^* P^- \overline{L} u + \overline{L}^* [\overline{L}, P^-] u \\ &= \overline{L}^* P^- f + \overline{L}^* O(\widetilde{P}^- u) + S^\infty w \\ &= \overline{L}^* P^- f + O(X, Y, I) \widetilde{P}^- u + S^\infty u. \end{aligned}$$

The term $O(X, Y, I)\widetilde{P}^- u$ may be rewritten as $XO(\widetilde{P}^- u) + YO(\widetilde{P}^- u) + O(\widetilde{P}^- u)$, again using commutators, so we have by (2.6)

$$-AP^- u = \overline{L}^* P^- f + XO(\widetilde{P}^- u) + YO(\widetilde{P}^- u) + O(\widetilde{P}^- u) + S^\infty u$$

from which (2.5) follows since $P^- = (\mathscr{P} A + \mathscr{E}) P^-$.

(2.5) plus (2.3) imply that $P^- u \in L^2_\delta$ for some $\delta > 0$, so by the Sobolev embedding theorem $P^- u \in L^r$ for some $r > 2$. Iterating proves that if $f \in L^p$ near 0, for some $p \in (2, \infty)$, then $P^- u \in L^p$ near 0 as well. One more iteration gives $P^- u \in L^p_\delta$ near 0. To obtain the full gain $\delta = m^{-1}$, differentiate (2.5) to derive

$$(X \text{ or } Y)P^- u = -(X \text{ or } Y)\mathscr{P}\overline{L}^* P^- f - (X \text{ or } Y)\mathscr{E} P^- u \\ + O[(X \text{ or } Y \text{ or } I)\mathscr{P}(X \text{ or } Y \text{ or } I)O(\widetilde{P}^- u)] + S^\infty u.$$

Since $\widetilde{P}^- u \in L^p$ and $P^- f \in L^p$, (2.4) and (2.3) imply that $(X \text{ or } Y)P^- u \in L^p$.

There remains $P^+\overline{L}^* v$. We first analyze $P^+ v$. Kohn observes that

$$\overline{L}\,\overline{L}^* P^+ = AP^+ + O(X, Y, I)P^+,$$

derived as (2.6) was. Note that we consider $\overline{L}\,\overline{L}^*$, rather than $\overline{L}^*\overline{L}$ as we did for P^-. Since

$$\overline{L}\,\overline{L}^* P^+ = P^+\overline{L}\,\overline{L}^* + \overline{L}[\overline{L}^*, P^+]\widetilde{P}^+ + [\overline{L}, P^+]\overline{L}^*\widetilde{P}^+ + S^\infty$$

we obtain

$$P^+ v = \mathscr{P} P^+ f + \mathscr{E} P^+ v + \mathscr{P} O(X, Y, I)\widetilde{P}^+ v + S^\infty v. \tag{2.7}$$

If $f \in L^p$ near x_0 then the usual bootstrapping argument, using (2.3) and (2.7), gives $P^+ v \in L^p_\varepsilon$. Since ε will typically be much less than one, this does not suffice to give any control on $P^+\overline{L}^* v$. Instead note that (2.7) implies

$$P^+ Xv = X\mathscr{P} P^+ f + X\mathscr{E} P^+ v + X\mathscr{P} O(X, Y, I)\widetilde{P}^+ v + O(\widetilde{P}^+ v) + S^\infty v, \tag{2.8}$$

and there is a similar expression for $P^+ Yv$. Now invoke (2.4). The term $O(X\widetilde{P}^+ v)$ may be rewritten as $X(O(\widetilde{P}^+ v)) + O(\widetilde{P}^+ v)$ using commutators, so that $X\mathscr{P} O(X\widetilde{P}^+ v) = (X\mathscr{P} X)O(\widetilde{P}^+ v) + X\mathscr{P} O(\widetilde{P}^+ v) = O(\widetilde{P}^+ v)$ in any L^p norm, $p \in (1, \infty)$. Similar terms involving Y or both X and Y may be treated in the same way. Therefore another bootstrap yields $P^+ Xv, P^+ Yv \in L^p$, assuming still that $f \in L^p$ near 0. Once $P^+(X \text{ or } Y)v$ is known to be in L^p, a term such as $X\mathscr{P} O(Y\widetilde{P}^+ v)$ may be viewed as a smoothing operator, $X\mathscr{P}$, applied to an L^p function, $O(Y\widetilde{P}^+ v)$. Therefore one more bootstrap yields $\widetilde{P}^+(X \text{ or } Y \text{ or } I)v \in L^p_\delta$ for some $\delta > 0$, so that in particular $\widetilde{P}^+\overline{L}^* v \in L^p_\delta$. In order to obtain the full gain $\delta = m^{-1}$ for $P^+ u$, X by $\overline{L}^*$ in (2.8) and then differentiate to obtain

$$(X \text{ or } Y)P^+\overline{L}^* v = (X \text{ or } Y)\overline{L}^*\mathscr{P} P^+ f + (X \text{ or } Y)\overline{L}^*\mathscr{E} P^+ v \\ + (X \text{ or } Y)\mathscr{P} O(X, Y, I)\widetilde{P}^+ v + (X \text{ or } Y)O(\widetilde{P}^+ v) + S^\infty v.$$

Invoking $(2.4')$ and $(2.4'')$, we see that each term is in L^p. In fact this argument shows that if $f \in L^p$ near 0 then $(X \text{ or } Y) \circ (X \text{ or } Y)P^+ v \in L^p$ near 0. This concludes the proof of the L^p Sobolev results of Theorems A and A$'$, in the special case $s = 0$.

Observe that it is the closed range property which permits the localization to a coordinate patch—what is needed is not the global condition $u \perp \mathscr{H}_b$, but the local condition $u = \overline{\partial}_b^* v$, $v \in L^2$.

To prove Theorem A$'$ for $s > 0$, let $f \in L^2$ globally and consider a small coordinate patch U on which $f \in L_s^p$. By a bootstrapping argument we may assume it to be already known that $u \in L_s^p$ on every compact subset of U. For $\sigma \geq 0$ let D^σ be an operator in the class $S_{1,0}^\sigma$ which, like the microlocalizing operators P, is the composition of two factors, the first being pointwise multiplication by a $C_0^\infty(U)$ function φ, and the second being the Fourier multiplier with symbol $(1+|\xi|^2)^{\sigma/2}$. $\tilde{D}^\sigma$ will denote an operator of the same type, with a function $\tilde{\varphi}$ which is identically one on the support of φ.

To prove that $P^- u \in L_{s+m^{-1}}^p$, it suffices to show that $(X \text{ or } Y)D^s P^- u \in L^p$. Just as above,

$$\begin{aligned}\overline{L}D^s P^- u &= D^s P^- f + [\overline{L}, D^s P^-]u \\ &= D^s P^- f + O(\tilde{D}^s \tilde{P}^- u) + S^\infty u\end{aligned}$$

and

$$\overline{L}^* \overline{L} D^s P^- u = \overline{L}^* D^s P^- f + \overline{L}^* O(\tilde{D}^s \tilde{P}^- u) + S^\infty u$$

so that

$$D^s P^- u = -\mathscr{P}\overline{L}^* D^s P^- f + \mathscr{E} D^s P^- u + \mathscr{P}\overline{L}^* O(\tilde{D}^s \tilde{P}^- u) + S^\infty u.$$

It follows that $(X \text{ or } Y)D^s P^- u$ is in L^p as desired. $P^0 u$ and $P^+ u$ may be treated by modifying the arguments for the case $s = 0$ in the same way.

Theorem B may be derived directly from the arguments of this section. Since it also follows from Theorem C of §12 we omit the details.

3. The absolute value of a vector field

The operator

$$\lambda(x)|T| = \lambda(x)\left|\frac{\partial}{\partial x_3}\right|$$

is awkward to deal with because, in addition to being rather singular, it is not expressed very directly in terms of the basic vector fields X and Y—one could write $T = \lambda(x)^{-1}[X, Y] + O(X, Y)$, but such an expression would be disastrous since in the weakly pseudoconvex case, λ will vanish somewhere. The goal of this section is to show that essentially $\lambda(x)|T|$ equals $|[X, Y]|$ modulo a lower order operator. However we shall not formulate this conclusion precisely, but will rest content with the technical Lemma 3.4. Observe that there is some difficulty in even defining $|[X, Y]|$. A reasonable definition for $|V|$, if V is a nowhere-vanishing vector field, could be obtained by changing coordinates so that V becomes a coordinate vector field, then using Fourier transformation. This will not do in the weakly pseudoconvex case, when $[X, Y]$ will vanish somewhere. There arises the question of how to define, and manipulate, the

absolute value of an arbitrary real, smooth vector field. Rather than attempting to apply spectral theory, we shall introduce a geometric approach which is suited to the calculations we shall encounter in studying the parametrix $\mathscr{P}$.

To begin consider d/dx on $\mathbf{R}^1$. Let $\mathscr{A}$ be the tempered distribution on $\mathbf{R}^1$ whose Fourier transform is $|\xi|$. Then $\mathscr{A}$ is real, even, C^∞ away from 0, and homogeneous of degree -2. For every $g \in \mathscr{S}$, the Schwartz class,

$$\left|\frac{d}{dx}\right| g(0) = \langle \mathscr{A}, g \rangle .$$

This also equals $\langle \mathscr{A}, \tilde{g} \rangle$ where $\tilde{g}(x) \equiv g(-x)$. The symbol $\mathscr{A}$ will be ubiquitous and will retain this meaning throughout the paper.

Given a smooth vector field V on an open set U, denote by $\exp(V)(x) = \exp(V)x$ the point $\psi(1)$, where $\psi : [0,1] \to U$ satisfies the ordinary differential equation

$$\begin{cases} \dfrac{d\psi}{dt} = V(\psi(t)), \\ \psi(0) = x, \end{cases}$$

provided the solution ψ exists on $[0,1]$. On any compact subset K of U there exists $\delta = \delta(V, K) > 0$ such that $\exp(rV)x$ is defined for all $r \le \delta$ and all $x \in K$.

Let $\varphi \in C_0^\infty(\mathbf{R})$ be identically one in some neighborhood of 0. If V is an arbitrary smooth real vector field on U, and $K \subset U$ is a compact subset, choose φ to be supported on $[-\delta, \delta]$ where δ is as above.

Definition 3.1. *For $f \in C^\infty(U)$ and $x \in K$,*

$$|V|_0 f(x) = \langle \mathscr{A}, \varphi g \rangle$$

where $g \in C^\infty([-\delta, \delta])$ is defined by $g(t) = f(\exp(tV)x)$.

The definition depends on the choice of φ. If φ were identically one on $\mathbf{R}$, this would agree with the natural definition of $|V|$ in the case where $U = \mathbf{R}^n$ and V is a coordinate vector field. In that case $|V|_0 - |V|$ is clearly bounded on L^1 and on $L^\infty(\mathbf{R}^n)$, since $(1-\varphi)\mathscr{A} \in L^1(\mathbf{R})$, so that $|V|_0 = |V|$ plus a lower-order error. In certain cases the cutoff φ is unnecessary, and we define $|V| f(x) = \langle \mathscr{A}, g \rangle$.

More generally consider a C^∞ map γ from a neighborhood of $K \times \{0\} \subset U \times \mathbf{R}$ to U, satisfying $\gamma(x, 0) \equiv x$.

Definition 3.2. *For $x \in K$ and $f \in C^\infty(U)$,*

$$|D_\gamma| f(x) = \langle \mathscr{A}, \varphi g \rangle$$

where $g(t) = f(\gamma(x, -t))$.

When $\gamma(x,t) \equiv \exp(tV)x$, $|D_\gamma| \equiv |V|_0$, but of course a generic family of curves $\gamma(x,t)$ will not be of this simple form.

Now consider a small neighborhood U of $0 \in \mathbf{R}^3$, let $T = \partial/\partial x_3$ and let $\lambda \in C^\infty(U)$ be an arbitrary nonnegative smooth function on U. We wish to

relate $\lambda(x)|T|_0$ to $|\lambda T|_0$. By $\lambda(x)|T|_0$ we mean the composition $|T|_0$ followed by pointwise multiplication by λ, while $|\lambda T|_0$ stands for $|V|_0$ where $V = \lambda T$. Define $\gamma(x,t)$ to be $(x_1, x_2, x_3 + t\lambda(x)) = x + t\lambda(x)e_3$.

Lemma 3.3. *For each compact set $K \subset U$ and each $p \in (1,\infty]$, $\lambda(x)|T|_0 - |D_\gamma|$ is bounded on L^p for all $p \in (1,\infty]$.*

It is assumed implicitly that the auxiliary functions φ appearing in the definitions of the two operators are supported sufficiently close to the origin, so that $\exp(tT)x \in U$ and $\gamma(x,t) \in U$ for all $x \in K$ and all $t \in \operatorname{support}(\varphi)$. It is not necessary that the two φ's be the same.

Proof. Introduce

$$\mathscr{M} f(x) = \sup_{r>0} r^{-1} \int_{|t|\le r} |f(x + te_3)|\, dt.$$

$\mathscr{M}$ is bounded on L^p for all $p \in (1,\infty]$ by the Hardy-Littlewood maximal theorem. If $\lambda > 0$ then $\lambda \int_{|t|\ge C\lambda} |f(x+te_3)|t^{-2}\, dt \prec \mathscr{M} f(x)$, uniformly in λ and in x. Let us adopt the convention that the action of the distribution $\mathscr{A}$ on a test function will be written as $\int_{\mathbf{R}} \mathscr{A}(t)g(t)\, dt$. This convention will hold throughout the paper, and will apply to other distributions as well. Now if $\lambda(x) \ne 0$, for any $f \in C^\infty$

$$\begin{aligned}
\lambda(x)|T|_0 f(x) &= \lambda(x) \int \mathscr{A}(t)\varphi(t) f(x + te_3)\, dt \\
&= \int \mathscr{A}(s)\varphi(\lambda(x)s) f(x + s\lambda(x)e_3)\, ds \\
&= \int \mathscr{A}(s)\varphi(s) f(\gamma(x,s))\, ds \\
&\quad + \int \mathscr{A}(s)[\varphi(\lambda(x)s) - \varphi(s)] f(x + s\lambda(x)e_3)\, ds.
\end{aligned}$$

The first term is $|D_\gamma| f(x)$. Reversing the change of variables, the second is majorized by

$$C\lambda(x) \int_{|t|\ge C\lambda(x)} |t|^{-2} |f(x + te_3)|\, dt \prec \mathscr{M} f(x).$$

If $\lambda(x) = 0$ then $\lambda(x)|T|_0 f(x) = 0$, while

$$\begin{aligned}
|D_\gamma| f(x) &= \int \mathscr{A}(t)\varphi(t) f(\gamma(x,t))\, dt \\
&\equiv f(x) \int \mathscr{A}(t)\varphi(t)\, dt \\
&= c f(x)
\end{aligned}$$

since $\gamma(x,t) \equiv x$. Thus in either case $\lambda(x)|T|_0 - |D_\gamma|$ is majorized pointwise by a fixed multiple of $\mathscr{M}$.

Lemma 3.4. *There exist smooth functions a_j, with $a_1 \equiv 1$, so that if $V_j = a_j \lambda T$ then for all $M \in \mathbf{Z}^+$ and all x, t*

$$x + t\lambda(x)e_3 = \exp\left(\sum_{j=1}^{M} t^j V_j\right)x + O(t^{M+1}). \tag{3.1}$$

That there should exist V_j satisfying (3.1) is an idea of Nagel, Stein, and Wainger [CNSW]; such an expansion exists for any smooth $\gamma(x,t)$ with $\gamma(x,0) \equiv x$, and the V_j are unique. What must be proved is that in our special case, they are all multiples of λT. First let us review the construction in [CNSW]. By taking $M = 1$ we find at once that V_1 must equal λT if (3.1) is to hold. The rest of the V_j are constructed by induction: once $V_1, \dots, V_k$ are found, set $\gamma_k(x,t) = \exp(-\sum_{j=1}^{k} t^j V_j)(\gamma(x,t))$. Then $\gamma_k(x,t) = x + O(t^{k+1})$ by the induction hypothesis. Hence the operator $f \to \partial^{k+1}/\partial t^{k+1} f(\gamma_k(x,t))|_{t=0}$ is a *first*-order differential operator which annihilates constants, in other words a vector field. Call it $(k+1)!\,V_{k+1}$. Taking f to be each coordinate function in turn, we find that $\gamma_k(x,t) = \exp(t^{k+1} V_{k+1})x + O(t^{k+2})$. Solving the relation $\exp(-\sum_{j=1}^{k} t^j V_j)(\gamma(x,t)) = \gamma_k(x,t)$ for γ and applying the Campbell-Hausdorff formula yields

$$\gamma(x,t) = \exp\left(\sum_{j=1}^{k+1} t^j V_j\right) x + O(t^{k+2})$$

as desired.

To show that each V_j is a smooth multiple of λT, suppose by induction that $V_j = a_j \lambda T$ for $1 \le j \le k$, $a_j \in C^\infty$. $\gamma_k(x,t) = x + b(x,t)e_3$ where b is C^∞. Expand b in a Taylor series $b \sim \sum_{i \ge k+1} (1/i!)\, t^i c_i(x)$ about $t = 0$. Of course the c_i depend also on k. We claim that each c_i is a smooth multiple of λ. Since $V_{k+1}(x)$ is a constant times $c_{k+1}(x)$ times $\partial/\partial x_3$, the claim suffices to conclude the proof.

Begin with a simpler computation, the Taylor expansion in t of $\exp(tV_1)y$ about $t = 0$. Write $\varphi(t) = \exp(tV_1)y = \exp(t\lambda T)y$. We have $\varphi(t) = (y_1, y_2, \psi(t))$ where $y = (y_1, y_2, y_3)$; y_1 and y_2 are parameters on which ψ depends, but they play no essential role and the problem is really one-dimensional. Write $\lambda(s)$ for $\lambda(y_1, y_2, s)$. Then ψ satisfies the ODE $\psi'(t) = \lambda(\psi(t))$, $\psi(0) = y_3$. Expand $\psi(t) \sim y_3 + \sum_{j=1}^{\infty} d_j t^j$ where the d_j depend implicitly on y, so that by the ODE,

$$\begin{aligned}\psi'(t) = \lambda(\psi(t)) &\sim \lambda\left(y_3 + \sum_{j=1}^{\infty} d_j t^j\right)\\ &\sim \lambda(y_3) + t\lambda'(y_3)d_1 + t^2\left[\lambda'(y_3)d_2 + \tfrac{1}{2}\lambda''(y_3)\,d_1^2\right] + \cdots.\end{aligned}$$

Differentiation of the original Taylor series gives the alternate expansion

$$\psi'(t) \sim \sum_{j=1}^{\infty} j d_j t^{j-1}.$$

Compare coefficients of t^j to obtain

$$d_1 = \lambda, \qquad 2d_2 = d_1\lambda', \qquad 3d_3 = \lambda' d_2 + \tfrac{1}{2}\lambda'' d_1^2$$

and so on. For each j, d_j is expressed as a polynomial in $\{d_i : i < j\}$ and in derivatives of λ. Each monomial involves at least one factor of some d_i. Substituting λ for d_1 in the expression for d_2 gives $d_2 = (1/2)\lambda\lambda'$, then substituting for both d_1 and d_2 gives $3d_3 = (1/2)\lambda(\lambda')^2 + (1/2)\lambda''\lambda^2$, and by induction on j we see that each d_j is a multiple of λ.

Next observe that in the Taylor expansion of $\exp(tV_1)(x + t\lambda(x)e_3)$, as a function of t, again all coefficients are multiples of $\lambda(x)$. For we may apply the last computation with $y = x + t\lambda(x)e_3$, and then expand each resulting coefficient $d_j(y) = d_j(x + t\lambda(x)e_3)$ in Taylor series about $t = 0$. Again the result is that

$$\exp(tV_1)(\gamma(x,t)) = \exp(tV_1)(x + t\lambda(x)e_3) \sim x + \sum_{j=1}^{\infty} b_j(x)t^j,$$

where each b_j is a polynomial function of $\lambda, \lambda', \lambda'', \ldots,$ with at least one factor of λ.

Now we are given that $V_j = a_j\lambda T$ for $j \le k$, and we might examine the Taylor expansion of

$$\gamma_k(x,t) = \exp\left(-\sum_{j=1}^{k} t^j V_j\right)(\gamma(x,t)) = \exp\left(-\sum_{j=1}^{k} t^j V_j\right)(x + t\lambda(x)e_3).$$

Use the Campbell-Hausdorff formula to write for any y

$$\exp\left(-\sum_{j=1}^{k} t^j V_j\right)(y) = \exp(t^{k+1}W_{k+1})\exp(t^k W_k)\cdots\exp(tW_1)(y) + O(t^{k+2})$$

where each of the W_j is a finite linear combination of the V_i and their iterated commutators, $1 \le i \le j$ ($i \le k$ when $j = k+1$). The commutator of any two smooth multiples of λT is another, so each W_j is a smooth multiple of λT. Apply the above computation to $\exp(tW_1)\gamma(x,t)$ (in fact $W_1 = V_1$). Then apply the same reasoning to $\exp(t^2 W_2)$ of the result, and continue by induction to treat the factors $\exp(t^j W_j)$ one by one. The conclusion is that c_{k+1} is a smooth multiple of λ (and the same for the other c_i, by the same reasoning).

4. The lifting procedure

Let g denote the free nilpotent Lie algebra of some step $m \ge 2$ on two generators $\hat{X}$, $\hat{Y}$. See [RS] for details. Let N be the dimension of g as

a vector space. Fix a basis $\{\widehat{Z}_\alpha\}$ for g as a vector space, each $\widehat{Z}_\alpha$ being a monomial $[(\widehat{X} \text{ or } \widehat{Y}), [(\widehat{X} \text{ or } \widehat{Y}), \dots, (\widehat{X} \text{ or } \widehat{Y})]\cdots]$.

Suppose that two smooth, real vector fields $\widetilde{X}$, $\widetilde{Y}$ are given in a neighborhood of $0 \in \mathbf{R}^N$. Define vector fields $\widetilde{Z}_\alpha$ by substituting $\widetilde{X}$, $\widetilde{Y}$ for $\widehat{X}$, $\widehat{Y}$ in the expressions defining the $\widehat{Z}_\alpha$. Let ϕ be the unique linear map from g to the tangent space to $\mathbf{R}^N$ at 0 satisfying $\phi(\widehat{Z}_\alpha) = \widetilde{Z}_\alpha$ for each α. We say that $\widetilde{X}$, $\widetilde{Y}$ are free to step m (at 0) if ϕ is an isomorphism of vector spaces.

Suppose next that X, Y are real, smooth vector fields in a neighborhood of $0 \in \mathbf{R}^3$. Suppose they satisfy the Hörmander/finite type condition, and let m be the order of the shortest commutator $Z = [(X \text{ or } Y), [(X \text{ or } Y), \dots]\cdots]$ such that X, Y, Z span the tangent space to $\mathbf{R}^3$ at 0; the order of $[X, Y]$ is 2, that of $[X, [X, Y]]$ is 3, and so on. Then consider the free nilpotent Lie algebra on two generators, with this particular step m. Let N be its dimension as a vector space, as above.

Our parametrix construction depends on a lemma of Rothschild and Stein [RS]:

Lemma 4.1. *There exist smooth, real vector fields $\widetilde{X}$, $\widetilde{Y}$ defined on a neighborhood $\widetilde{U}$ of $0 \in \mathbf{R}^N$ and a map π from $\widetilde{U}$ to a neighborhood U of $0 \in \mathbf{R}^3$ such that*

$$d\pi \text{ is surjective at every point in } \widetilde{U} \tag{4.1}$$

$$\pi \exp(t\widetilde{X})x \equiv \exp(tX)(\pi x) \quad \text{and} \quad \pi \exp(t\widetilde{Y})x \equiv \exp(tY)(\pi x) \tag{4.2}$$

$$\widetilde{X}, \widetilde{Y} \quad \text{are free to order } m. \tag{4.3}$$

Choose coordinates $x = (x', x'') \in \mathbf{R}^3 \times \mathbf{R}^{N-3}$ so that $\pi(x) \equiv x'$ in a neighborhood of $0 \in \mathbf{R}^N$. Then the lemma may be used as in [RS] to study $\Delta = -X^2 - Y^2$ as follows: Suppose $\Delta u = f$ in a neighborhood of $0 \in \mathbf{R}^3$. Fix $\zeta \in C_0^\infty(\mathbf{R}^{N-3})$, $\zeta \equiv 1$ near 0. Set $\widetilde{\Delta} = -\widetilde{X}^2 - \widetilde{Y}^2$, $\tilde{u}(x) = u(x')\zeta(x'')$, and $\tilde{f}(x) = f(x')\zeta(x'')$. $\tilde{f}$ and $\tilde{u}$ belong to precisely the same Sobolev and Hölder spaces near 0 as do f and u, respectively. By (4.2) the equation $\widetilde{\Delta}\tilde{u} = \tilde{f}$ holds when x' is near 0 on the interior of the region where $\zeta(x'') \equiv 1$, and more generally $\widetilde{\Delta}\tilde{u} = \tilde{f} + O(\widetilde{X}, \widetilde{Y}, I)u'$ where $u'(x) = u(x')\zeta'(x'')$, where ζ' is chosen so that $\zeta'\zeta \equiv \zeta$. Now suppose that P is a parametrix for $\widetilde{\Delta}$, so that $P\widetilde{\Delta} = I + E$ on functions supported near O, where E is in some sense a smoothing operator. Then

$$\tilde{u} = P\tilde{f} - E\tilde{u} + PO(\widetilde{X}, \widetilde{Y}, I)u'. \tag{4.4}$$

As in §2 we may equally well write $O(\widetilde{X}u') = \widetilde{X}\,O(u') + O(u')$ and the same for $\widetilde{Y}$. Any regularity for u, expressed for instance in terms of ordinary Sobolev spaces L_s^p, implies the same degree of regularity for $\tilde{u}$, and the same for $\tilde{f}$ relative to f. Then (4.4), together with knowledge of the degree to which

$P, P(\widetilde{X}$ or $\widetilde{Y})$, and E are smoothing, will imply an additional degree of regularity for $\tilde{u}$. To convert back to information on u, let $\eta \in C_0^\infty(\mathbf{R}^{N-3})$ be supported inside the region where $\zeta \equiv 1$, satisfying $\int \eta(x'')dx'' = 1$. Consider the operator $R\colon C^\infty(\widetilde{U}) \to C^\infty(U)$ defined by $Rg(x') = \int g(x', x'')\eta(x'')\, dx''$. Then near $0 \in \mathbf{R}^3$, $u \equiv R\tilde{u} = RP\tilde{f} - RE\tilde{u} + RPO(\widetilde{X}, \widetilde{Y}, I)u'$. R is bounded from any Sobolev space $L_s^p(\widetilde{U})$ to $L_s^p(U_0)$, where U_0 is open and $\overline{U}_0 \subset U$. A typical step in a bootstrapping procedure for the analysis of Δ would now go:

$$u \in L^p(U) \Rightarrow \tilde{u} \in L^p(\widetilde{U}_0) \Rightarrow \tilde{u} \in L_\delta^p(\widetilde{U}_0) \Rightarrow u \in L_\delta^p(U_0),$$

for some $\delta > 0$. In fact we have $\mathscr{P}\Delta = I + \mathscr{E}$ as operators from $L^p(U)$ to $L^p(U_0)$, where $\mathscr{P}h = RP\tilde{h}$ and $\mathscr{E}h = RE\tilde{h} - R\mathscr{P}O(\widetilde{X}, \widetilde{Y}, I)h'$.

Assume now that $[X, Y](x) = \lambda(x)\,\partial/\partial x_3 + O(X, Y)$, $0 \le \lambda \in C^\infty$. In order to carry out a similar analysis for $\Delta + \lambda(x)\,|\partial/\partial x_3|$, consider instead $\Delta + |D_\gamma|$ where $\gamma(x, t) = x + t\lambda(x)e_3$ as in §3. By Lemma 3.4 there exist C^∞ functions a_j, b_j, c_j such that for each $M \in \mathbf{Z}^+$,

$$\gamma(x, t) = \exp\left(\sum_{j=1}^{M} t^j\{a_j[X, Y] + b_jX + c_jY\}\right)x + O(t^{M+1}).$$

Lift X, Y to $\widetilde{X}$, $\widetilde{Y}$ on $\mathbf{R}^N$ as in Lemma 4.1. Lift a_j by setting $\tilde{a}_j(x', x'') = a_j(x')$ and similarly for b_j, c_j. Fix $M \ge m$ and observe that

$$\begin{aligned}\pi \exp\left(\sum_{j=1}^{M} t^j\{\tilde{a}_j[\widetilde{X}, \widetilde{Y}] + \tilde{b}_j\widetilde{X} + \tilde{c}_j\widetilde{Y}\}\right)x \\ \equiv \exp\left(\sum_{j=1}^{M} t^j\{a_j[X, Y] + b_jX + c_jY\}\right)(\pi x)\end{aligned}$$

since $d\pi(\tilde{a}_j[\widetilde{X}, \widetilde{Y}] + \tilde{b}_j\widetilde{X} + \tilde{c}_j\widetilde{Y}) \equiv a_j[X, Y] + b_jX + c_jY$. Therefore if we define $\tilde{\gamma}_0, \gamma', \gamma''$ by

$$\begin{aligned}\exp\left(\sum_{j=1}^{M} t^j\{a_j[\widetilde{X}, \widetilde{Y}] + b_j\widetilde{X} + c_j\widetilde{Y}\}\right)x &= \tilde{\gamma}_0(x, t) \\ &= (\gamma'(x, t), \gamma''(x, t)) \\ &\equiv (\gamma'(x', t), \gamma''(x, t)),\end{aligned}$$

we obtain

$$\pi\tilde{\gamma}_0(x, t) = \exp\left(\sum_{j=1}^{M} t^j\{a_j[X, Y] + b_jX + c_jY\}\right)(x') = \gamma(x', t) + O(t^{M+1}).$$

Define a lifted family of curves by $\tilde{\gamma}(x, t) = (\gamma(x', t), \gamma''(x, t))$. Then surely $\pi\tilde{\gamma}(x, t) \equiv \gamma(\pi x, t)$. But $\tilde{\gamma}(x, t) = \tilde{\gamma}_0(x, t) + O(t^{M+1})$. There exist unique vector fields $W_1, \ldots, W_M$ such that $\tilde{\gamma}(x, t) = \exp(\sum_{j=1}^{M} t^jW_j)x + O(t^{M+1})$ for

all x, t, and it is clear from the proof of Lemma 3.3 that the W_j depend only on the Taylor expansion of γ to order M about $t = 0$. Therefore $W_j = \tilde{a}_j[\tilde{X}, \tilde{Y}] + \tilde{b}_j\tilde{X} + \tilde{c}_j\tilde{Y}$. We have established

Lemma 4.2. *There exists a smooth family of curves $\tilde{\gamma}(x, t)$ on $\tilde{U}$ such that $\pi\tilde{\gamma}(x, t) \equiv \gamma(\pi x, t)$ and $\tilde{\gamma}(x, t) = \exp(\sum_{j=1}^{M} t^j(a'_j[\tilde{X}, \tilde{Y}] + b'_j\tilde{X} + c'_j\tilde{Y}))x + O(t^{M+1})$, where $a'_j, b'_j, c'_j \in C^\infty$ and $a'_1 \equiv 1$.*

To lift $|D_\gamma|$ define

$$\tilde{B}g(x) = \int \mathscr{A}(t)\varphi(t)g(\tilde{\gamma}(x, t))\, dt$$

on $\tilde{U}$, so $\tilde{B} = |D_{\tilde{\gamma}}|$. For f defined on U there was the relation $(\tilde{\Delta}\tilde{f})(x) = \zeta(x'')(\Delta f)(\pi x) + O(\tilde{X}, \tilde{Y}, I)f'$. For $\tilde{B}$ the situation is less simple:

$$\begin{aligned}\tilde{B}\tilde{f}(x) &= \int \mathscr{A}(t)\varphi(t)\zeta(\gamma''(x, t))f(\gamma(\pi x, t))\, dt \\ &= \zeta(x'')\int \mathscr{A}(t)\varphi(t)f(\gamma(\pi x, t))\, dt + \int \mathscr{A}(t)\varphi(t)\psi(x, t)f(\gamma(\pi x, t))\, dt \\ &= \zeta(x'')(|D_\gamma|f)(\pi x) + B'f'(x)\end{aligned}$$

where $\psi(x, t) = \zeta(\gamma''(x, t)) - \zeta(x'')$ satisfies $\psi(x, 0) \equiv 0$ and

$$(B'h)(x) = \int \mathscr{A}(t)\varphi(t)\psi(x, t)h(\tilde{\gamma}(x, t))\, dt.$$

Lemma 4.3. *For any $p \in (1, \infty)$ and any $\varepsilon > 0$, B' is bounded from $L^p_\varepsilon(\mathbf{R}^N)$ to $L^p(\tilde{U})$ and from $L^p(\mathbf{R}^N)$ to $L^p_{-\varepsilon}(\tilde{U})$.*

Proof. Fix $\eta \in C_0^\infty(\mathbf{R}^N)$, let $\eta_t(x) = t^{-N}\eta(t^{-1}x)$ and define $T_t g(x) = g * \eta_t(x)$. To establish the first conclusion it suffices to prove that for any such η, $B' \circ T_t$ is bounded from L^p to L^p with a bound of at worst $C(1 + \log t^{-1})$ as $t \to 0$. Note that T_t maps L^p to C^∞. The kernel $s \to \mathscr{A}(s)\varphi(s)\psi(x, s)$ formally equals $c\,\partial\psi/\partial s\,(x, 0)s^{-1} + O(s^0)$ as $s \to 0$, since $\mathscr{A}$ is homogeneous of degree -2 and $\psi(x, 0) \equiv 0$. For $g \in C^1$, $B'g(x)$ is equal to the principal-value integral $pv\int \mathscr{A}(s)\varphi(s)\psi(x, s)g(\tilde{\gamma}(x, s))\, ds$, which exists for each x. If $K_t(x, y)$ is the kernel satisfying $B' \circ T_t g(x) = \int K_t(x, y)g(y)\, dy$, then straightforward estimation gives

$$\sup_x \int |K_t(x, y)|\, dy + \sup_y \int |K_t(x, y)|\, dx \le C(1 + \log t^{-1}).$$

By the same reasoning B' is bounded from L^p to L^p_ε for all $\varepsilon < 0$.

Remark. The lifted family of curves $\tilde{\gamma}$ may be constructed so that the results of [C1] would imply that B' is actually bounded on all L^p, $1 < p < \infty$.

Suppose now that on a neighborhood of $0 \in \mathbf{R}^3$, $A_0 u = f$, where $A_0 = \Delta + \lambda(x)|T|_0$. Then

$$\begin{aligned}(\Delta + |D_\gamma|)u &= f + (|D_\gamma| - \lambda(x)|T|_0)u \\ &= f + Qu\end{aligned}$$

where Q is bounded on all L^p. Therefore if $\tilde{A} = \tilde{\Delta} + \tilde{B}$,

$$\tilde{A}\tilde{u} = \widetilde{\Delta u} + O(\tilde{X}, \tilde{Y}, I)u' + |D_\gamma|\tilde{u} + B'u' = \tilde{f} + O(\tilde{X}, \tilde{Y}, I)u' + \widetilde{Qu} + B'u'.$$

If we can find a parametrix $\widetilde{\mathscr{P}}$ for $\tilde{A}$ so that $\widetilde{\mathscr{P}}\tilde{A} = I - \tilde{\mathscr{E}}$, then we obtain

$$\tilde{u} = \tilde{\mathscr{E}}\tilde{u} + \widetilde{\mathscr{P}}\tilde{A}\tilde{u} = \tilde{\mathscr{E}}\tilde{u} + \widetilde{\mathscr{P}}\tilde{f} + \widetilde{\mathscr{P}}O(\tilde{X}, \tilde{Y}, I)u' + \widetilde{\mathscr{P}}\widetilde{Qu} + \widetilde{\mathscr{P}}B'u'.$$

Hence

$$u = R\tilde{u} = R\widetilde{\mathscr{P}}\tilde{f} + R\tilde{\mathscr{E}}\tilde{u} + R\widetilde{\mathscr{P}}O(\tilde{X}, \tilde{Y}, I)u' + R\widetilde{\mathscr{P}}\widetilde{Qu} + R\widetilde{\mathscr{P}}B'u'.$$

For derivatives of u we have for instance

$$Xu = R(\tilde{X}\tilde{u}) = R\tilde{X}\widetilde{\mathscr{P}}\tilde{f} + R\tilde{X}\tilde{\mathscr{E}}\tilde{u} + R\tilde{X}\widetilde{\mathscr{P}}O(\tilde{X}, \tilde{Y}, I)u' + R\tilde{X}\widetilde{\mathscr{P}}\widetilde{Qu} + R\tilde{X}\widetilde{\mathscr{P}}B'u'.$$

There is actually one more error term to take into account, for we are interested in the equation $(\Delta + \lambda(x)|T|)w = g$, so there remains $\lambda(x)(|T| - |T|_0)w$. In our application w will always be of the form $(P^+ \text{ or } P^- \text{ or } P^0)w'$, with $w' \in L^2$. Each P is a composition of two operators, the first being multiplication by a C_0^∞ function supported in a small neighborhood U_0 of 0. Applying the second factor results in a function which is C^∞ outside $\overline{U}_0$, and which together with all its derivatives is $O(|x|^{-3})$ there. Therefore $(|T| - |T|_0)w$ will be C^∞ on U, provided U is sufficiently small.

Now since Q is bounded and B' essentially so (Lemma 4.3), in order for (2.3) through (2.4″) to be justified it suffices to prove

Proposition 4.4. *There exists a parametrix $\widetilde{\mathscr{P}}$ for $\tilde{A}$ such that $\widetilde{\mathscr{P}}\tilde{A} = I + \tilde{\mathscr{E}}$ and all of the following operators are bounded on L^p for all $p \in (1, \infty)$:*

$$\begin{gathered}\widetilde{\mathscr{P}},\ \tilde{\mathscr{E}},\ (\tilde{X} \text{ or } \tilde{Y})\widetilde{\mathscr{P}},\ \widetilde{\mathscr{P}}(\tilde{X} \text{ or } \tilde{Y}),\ (\tilde{X} \text{ or } \tilde{Y})\tilde{\mathscr{E}} \\ (\tilde{X} \text{ or } \tilde{Y})\widetilde{\mathscr{P}}(\tilde{X} \text{ or } \tilde{Y}),\ (\tilde{X} \text{ or } \tilde{Y})(\tilde{X} \text{ or } \tilde{Y})\widetilde{\mathscr{P}}, \\ \widetilde{\mathscr{P}}(\tilde{X} \text{ or } \tilde{Y})(\tilde{X} \text{ or } \tilde{Y}),\ (\tilde{X} \text{ or } \tilde{Y})(\tilde{X} \text{ or } \tilde{Y})\tilde{\mathscr{E}}.\end{gathered}$$

It will not be necessary, until §11, to discuss further the original vector fields X, Y in $\mathbf{R}^3$. Until then we shall work entirely with the lifted vector fields $\tilde{X}$, $\tilde{Y}$. The reader should beware that the notation changes more than once; in the next section X, Y denote invariant vector fields on a certain Lie group, while in subsequent sections they denote those vector fields which have heretofore been called $\tilde{X}, \tilde{Y}$. Similarly what was denoted by $\widetilde{\mathscr{P}}$ above will be simply $\mathscr{P}$ in §§6–10.

5. Analysis on the free group: the model

Our goal is to construct the parametrix for the lifted operator $\tilde{A}$ of §4. In the present section we analyze a model operator which is invariant under the

translation and dilation structures of a certain nilpotent Lie group. We obtain a reasonably precise estimate on the degree of regularity of the convolution kernel for its inverse.

Let $\mathbf{g}$ be the free nilpotent Lie algebra of step m on two generators X, Y, and let G be the associated connected, simply connected nilpotent group. $G = \mathbf{R}^n$ topologically. Via the exponential map, X, Y may be identified with left-invariant vector fields on G. Throughout this section let

$$\Delta = -X^2 - Y^2 \quad \text{and} \quad A = \Delta + |[X, Y]|$$

where $|[X, Y]|$ is as defined in §3 with $\varphi \equiv 1$ on all of $\mathbf{R}$. Then A is a left-invariant operator on G. For each $r \in \mathbf{R}^+$ the map $X \to rX$, $Y \to rY$ extends uniquely to an automorphism of $\mathbf{g}$, and by the exponential map defines also an automorphism δ_r of G. The set of all δ_r forms a group, $\delta_r \delta_s = \delta_{rs}$, and they are called dilations. A is homogeneous, that is, if $f_r(x) = f(\delta_r(x))$ then $A(f_r) = r^2(Af)_r$ for any f in the Schwartz class $\mathscr{S}$. Note that both terms Δ and $|[X, Y]|$ have the same homogeneity. Thus as is usual in this theory (see [FS, RS]), $|[X, Y]|$ should be regarded as a second-order operator. Haar measure on G is simply Lebesgue measure in the coordinates defined by the exponential map. There is an exponent d, the homogeneous dimension of G, such that for any set E, $|\delta_r E| = r^d |E|$ for any $r \in \mathbf{R}^+$, where $|E|$ denotes here the Haar measure. We assume that the step m is at least two, so $d \geq 4$. A acts on Schwartz functions by convolution on the right with a distribution k_0 which is homogeneous of degree $-d-2$, that is, $k_0(\delta_r x) \equiv r^{-d-2} k_0(x)$, where this must be interpreted in the sense of distributions. The part of k_0 corresponding to Δ is simply Δ applied to the Dirac mass at 0, the group identity element, hence is supported at 0. The part corresponding to $|[X, Y]|$ is $\mathscr{A}(t)$ times one-dimensional Lebesgue measure on the line $t \to \exp(t[X, Y])0$. Thus away from 0, k_0 is a singular measure.

We will construct an operator $\mathscr{P}$ so that $\mathscr{P}A = A\mathscr{P} = I$ on Schwartz functions. The translation and dilation-invariance of A will be reflected in $\mathscr{P}$, for $\mathscr{P}f = f * \ell$ where ℓ is homogeneous of degree $-d+2$. Such a distribution can have no part supported at 0, and is determined by its behavior away from 0, that is, on the region $\|x\| \sim 1$. The issue for us is how smooth ℓ is away from 0. It follows easily from a theorem of [CG] that if ℓ were C^∞ away from 0, then k_0 must be also, a contradiction. The best one might hope for is that ℓ should be differentiable of some small finite order. In order for the transplantation of $\mathscr{P}$ to a parametrix for $\tilde{\Delta} + \tilde{B}$ on $\mathbf{R}^N$ to succeed, it will be vital to achieve a fairly sharp control on this order of differentiability.

The smoothness of ℓ will be analyzed in terms of the natural scale of Sobolev spaces on G. For $0 \leq \alpha \in \mathbf{R}$ and $p \in (1, \infty)$, $L^p_\alpha(G) = \{f \in L^p : \Delta^{\alpha/2} f \in L^p\}$. See [C2] for some discussion, and [Fo] for details. We make use of the following facts: For each $\alpha \in \mathbf{C}$ with $\mathrm{Re}(\alpha) > -d$ there exists a distribution I_α, homogeneous of degree $-d-\alpha$, so that $\Delta^{\alpha/2} f = f * I_\alpha$ for all $f \in \mathscr{S}$.

Whenever $\mathrm{Re}(\alpha)$, $\mathrm{Re}(\beta)$ and $\mathrm{Re}(\alpha+\beta)$ are all $> -d$, $I_\alpha * I_\beta$ is naturally defined and equals $I_{\alpha+\beta}$.

Let $\|\cdot\| : G \to (0,\infty)$ be a homogeneous norm on G, that is, $\|\delta_r x\| \equiv r\|x\|$, $\|x\| = 0 \Leftrightarrow x = 0$, and $\|\cdot\|$ is C^∞ on $G\backslash\{0\}$. Fix $\eta \in C_0^\infty(\mathbf{R}^+)$, not identically zero.

Definition. *A distribution k on G, homogeneous of any degree, is said to belong to L^p_β away from 0 if $\eta(\|\cdot\|)k \in L^p_\beta(G)$.*

The definition is independent of the choice of η.

The Calderón-Zygmund theory of singular integrals has the following extension to G: let $k \in \mathscr{S}'$ be homogeneous of degree $-d$, and suppose that $k \in L^q_\varepsilon$ away from 0 for some $q \in (1,\infty)$ and $\varepsilon > 0$. Then $Tf = f * k$ extends to a bounded operator on $L^p(G)$ for all $p \in (1,\infty)$. (In the abelian case $G = \mathbf{R}^n$, the assumption that k is a distribution implies the familiar condition that k has mean value zero on the unit sphere.) The result of [C2] [1] is:

Inversion Theorem. *Suppose $TF = f * k$ where k is homogeneous of degree $-d$, and $k \in L^q_\alpha$ away from 0, for some $q \in (1,\infty)$ and $\alpha > 0$. Suppose T is invertible as an operator on $L^2(G)$. Then there exists a unique $\ell \in \mathscr{S}'$ such that $T^{-1} f = f * \ell$ for all $f \in \mathscr{S}$. ℓ is homogeneous of degree $-d$, and $\ell \in L^q_\alpha$ away from 0.*

The principal conclusion is the last one; it is sharp in that there is no loss of regularity in the passage from k to ℓ. [2]

To reduce the analysis of A to an operator of order 0, factor (formally)

$$A = \Delta^{1/2}(I+S)\Delta^{1/2},$$

where $S = \Delta^{-1/2}|[X,Y]|\Delta^{-1/2}$. We shall use the inversion theorem to analyze $(I+S)^{-1}$, and then since $\Delta^{-1/2}$ is already well understood will be able to invert A by taking $A^{-1} = \Delta^{-1/2}(I+S)^{-1}\Delta^{-1/2}$. Of course these formal manipulations must be justified, and the regularity of the convolution kernel for S must be understood for the program to succeed.

For various technical purposes it is helpful to work with right-invariant derivations. By a left-invariant differential operator D on G we mean any constant-coefficient linear combination of monomials $(X \text{ or } Y)\circ(X \text{ or } Y)\circ\cdots\circ(X \text{ or } Y)$. There exist unique right-invariant vector fields $\widehat{X}$, $\widehat{Y}$ which agree with X, Y at 0, and a right-invariant differential operator is then any polynomial in $\widehat{X}$, $\widehat{Y}$, still with constant coefficients. The weight of a monomial is defined to be minus one times the number of factors of X or Y, or of $\widehat{X}$ or $\widehat{Y}$ respectively.

On $\mathbf{g}$ introduce a vector space basis $\{Z_j\}$, each Z_j a monomial $[X \text{ or } Y, [X \text{ or } Y, \ldots]]$, with w_j factors. Introduce coordinates $x \in \mathbf{R}^N$ on G, where

[1] A simpler proof may be obtained from the technique of [C5].

[2] The result is also valid for $\alpha = 0$ [C5], a question raised by A. Carbery.

$x = (x_1, \dots, x_N)$ is the image under the exponential map of $\sum x_j Z_j \in \mathbf{g}$. The homogeneous dimension d of G is then $\sum_j w_j$. Define the weight of a monomial $P(x) = \prod x_j^{\beta_j}$ to be $w = \sum \beta_j \cdot w_j$; P is homogeneous of weight w in the sense that $P(\delta_r x) \equiv r^w P(x)$. If D is either a left- or right-invariant differential monomial of weight α and P is a monomial homogeneous of weight w, we say that the operator $P(x)(Df)(x)$ has weight $\alpha + w$. Then any left-invariant differential monomial D of some weight α may be written as a finite sum $\sum P_i(x)\widehat{D}_i$ where each $\widehat{D}_i$ is right-invariant and $P_i \cdot \widehat{D}_i$ is homogeneous of weight α. Conversely every right-invariant $\widehat{D}$ admits such a representation in terms of polynomials and left-invariant operators. We shall have occasion to work with C^∞ functions on G which satisfy $|(Df)(x)| \le C_w(1 + \|x\|)^{a+w}$ for all left-invariant D of weight w, for all w, for some exponent a which depends on f. It follows that this is equivalent to requiring the same estimate for all right-invariant D.

We begin the analysis of $(I + S)^{-1}$.

Lemma 5.1. *S is well-defined on Schwartz functions. $Sf = f * k$ where k is homogeneous of degree $-d$. $\langle g, Sf\rangle = \langle Sg, f\rangle$ for all $f, g \in \mathscr{S}$, and $\langle f, Sf\rangle \ge 0$ for all $f \in \mathscr{S}$.*

$\langle f, g\rangle$ denotes $\int_G f\overline{g}\,dx$.

Proof. Write $|[X, Y]|g = g * \mu$. If $f \in \mathscr{S}$ then $\Delta^{-1/2} f = f * I_{-1}$ is C^∞ and satisfies $|f * I_{-1}|(x)| \le C(1 + \|x\|)^{-d+1}$ since $|I_1(x)| \prec \|x\|^{-d+1}$. Moreover any combination of right- and left-invariant partial derivatives of $f * I_{-1}$, of any order, satisfies the same decay estimate. Split $\mu = \mu^0 + \mu^\infty$ where μ^0 has compact support and μ^∞ vanishes near 0. Convolution with μ^0 maps smooth functions to smooth functions, and from left–invariance it follows that $(f * I_{-1}) * \mu^0$ satisfies the same pointwise estimates as does $(f * I_{-1})$. If we also split $I_{-1} = I_{-1}^0 + I_{-1}^\infty$ then $f * I_{-1} * \mu^0 * I_{-1}^0$ is certainly a well-defined smooth function. Since $d \ge 4$ the estimate $|f * I_{-1} * \mu^0(x)| \le C(1 + \|x\|)^{-d+1}$ implies $f * I_{-1} * \mu^0 \in L^2$, and also $I_{-1}^\infty \in L^2$, so $f * I_{-1} * \mu^0 * I_{-1}^\infty$ is a well-defined continuous function. Since μ^∞ is a finite measure, $(f * I_{-1}) * \mu^\infty \in L^2$ so again $f * I_{-1} * \mu^\infty * I_{-1}^\infty$ is defined and continuous. Finally $f * I_{-1} * \mu^\infty$ is bounded, and is even smooth (look at right-invariant derivatives), so $(f * I_{-1} * \mu^\infty) * I_{-1}^0$ is defined and smooth.

Each factor $\Delta^{-1/2}, |[X, Y]|, \Delta^{-1/2}$ satisfies $\langle g, \cdot f\rangle = \langle \cdot g, f\rangle$ for all $f, g \in \mathscr{S}$, so it is clear that the composition does also. Certainly our definition of Sf gives $Sf = f * k$ for some $k \in \mathscr{S}'$. k does not depend on the choice of the splittings $\mu = \mu^0 + \mu^\infty$ and $I_{-1} = I_{-1}^0 + I_{-1}^\infty$. Therefore since I_{-1}, μ are homogeneous, k must be also, and its degree is degree (I_{-1}) + degree μ + degree $(I_{-1}) + 2d = (-d+1) - d - 2 + (-d+1) + 2d = -d$. Finally it follows from

the argument used to construct k that $\langle g, Sf\rangle = \langle \Delta^{-1/2} g, |[X, Y]|\Delta^{-1/2} f\rangle$ for all $f, g \in \mathscr{S}$.

$\langle h, |[X, Y]|h\rangle \geq 0$ for all $h \in \mathscr{S}$. For certainly $\langle \varphi, |\partial/\partial x|\varphi\rangle \geq 0$ for all $\varphi \in \mathscr{S}(\mathbf{R})$. On G we may choose coordinates $(x_0, x') \in \mathbf{R} \times \mathbf{R}^{N-1}$ in which Haar measure is still Lebesgue measure, and $|[X, Y]|h(x) = \int \mathscr{A}(t)h(x_0 + t, x')\,dt$. It is then clear that $\langle h, |[X, Y]|h\rangle \geq 0$. Now by approximation by Schwartz functions it follows that $\langle \Delta^{-1/2} f, |[X, Y]|\Delta^{-1/2} f\rangle \geq 0$ for all $f \in \mathscr{S}$.

Lemma 5.2. *There exist $\gamma > 1$ and $q > 1$ such that $k \in L^q_\gamma$ away from 0.*

Proof. Fix $\zeta \in C_0^\infty(\mathbf{R}), \zeta \equiv 1$ near 0, and let $I^0_{-1} = I_{-1}(x)\zeta(\|x\|)$ and $\mu^0 = \zeta(\|x\|) \cdot \mu$. Recall that it is $\eta(\|x\|)k(x)$ which must be estimated, where η is supported in a compact subset of $(0, \infty)$. Choose the support of ζ so small that $\eta(\|x\|) \cdot (I^0_{-1} * \mu^0 * I^0_{-1})(x) \equiv 0$. $I_{-1} * \mu * I_{-1} = (I^0_{-1} + I^\infty_{-1}) * (\mu^0 + \mu^\infty) * (I^0_{-1} + I^\infty_{-1})$ is a sum of eight terms, of which seven remain to be considered. But all of the seven are C^∞, except $I^0_{-1} * \mu^\infty * I^0_{-1}$. Since I^0_{-1} has compact support, $\eta(\|x\|) \cdot (I^0_{-1} * \mu^\infty * I^0_{-1}) \equiv \eta(\|x\|) \cdot (I^0_{-1} * \bar{\mu} * I^0_{-1})$ where $\bar{\mu}$ is the restriction of μ^∞ to a large compact set. It then suffices to apply the next lemma.

Lemma 5.3. *$I^0_{-1} * \nu * I^0_{-1} \in L^q_\gamma$ for some $q \in (1, \infty)$ and $\gamma > 1$, for any compactly supported finite measure ν.*

Proof. Let $\hat{\Delta} = -\hat{X}^2 - \hat{Y}^2$ be the right-invariant sub-Laplacian. $\hat{\Delta}^{\alpha/2} f = I_\alpha * f$ for $f \in \mathscr{S}$, with the same I_α as for $\Delta^{\alpha/2}$. For any $0 \leq \beta < 1$, $I_\beta * I^0_{-1} \in L^q$ for some $q(\beta) > 1$. Indeed $|I_\beta(x)| \prec \|x\|^{-d-\beta}$ for $x \neq 0$, so $(I_\beta * I^0_{-1})(x) \prec \|x\|^{-d-\beta} \in L^1 \cap L^\infty$ on $\{\|x\| \geq 1\}$. For small x, $I_\beta * I^0_{-1} = I_\beta * I_{-1} - I_\beta * I^\infty_{-1} = I_{\beta-1} - I_\beta * I^\infty_{-1}$. The second term is C^∞. The first is $\prec \|x\|^{-d+1-\beta} \in L^q$ as $\|x\| \to 0$, if q is sufficiently close to 1. Therefore $I_\beta * (I^0_{-1} * \nu) = (I_\beta * I^0_{-1}) * \nu \in L^q$ as well, so $I^0_{-1} * \nu$ is a compactly supported function in the right-invariant Sobolev space $\hat{L}^q_\beta$, defined in the natural way. But it is a fundamental result of [RS] that $\hat{L}^q_\beta \subset L^q_{\beta/m}$ for compactly supported functions, where m is the step. Decomposing $I^0_{-1} = I_{-1} - I^\infty_{-1}$ again, it follows that convolution on the right with I^0_{-1} maps $L^q_{\beta/m}$ to L^q_γ, $\gamma = 1 + m^{-1}\beta > 1$.

Corollary 5.4. *$S = \Delta^{-1/2}|[X, Y]|\Delta^{-1/2}$ extends to a bounded operator on L^p for all $p \in (1, \infty)$, and $I + S$ is invertible on L^2.*

Proof. $Sf = f * k$ for $f \in \mathscr{S}$, where $k \in \mathscr{S}'$ is homogeneous of degree $-d$, and k is a locally integrable function away from 0. By Lemma 2.4 of [C2], this implies that k has mean value 0 on the unit sphere with respect to the appropriate measure, that is, $\int \eta(\|x\|)k(x)\,dx = 0$ for any $\eta \in C_0^\infty(\mathbf{R}^+)$. Moreover that lemma guarantees that as a distribution $k = c\delta + k'$ where δ

is the Dirac mass at 0, and k' is the principal-value distribution associated to the function $k(x), x \neq 0$.

Applying the Cotlar-Knapp-Stein almost-orthogonality lemma as in [KS] and using the regularity conclusion of Lemma 5.2 establishes the L^2 boundedness of S. The Calderón-Zygmund machinery then gives weak type (1,1) boundedness, again invoking Lemma 5.2. Interpolation and duality then give L^p boundedness. Finally $I + S$ is invertible on L^2 since S is selfadjoint and nonnegative.

From the inversion theorem follows

Corollary 5.5. *There exists a distribution ℓ homogeneous of degree $-d$, belonging to L^q_γ away from 0 for some $q > 1$ and $\gamma > 1$, such that $(I+S)^{-1} f = f * \ell$ for all $f \in \mathscr{S}$.*

Consider $L = I_{-1} * \ell * I_{-1}$, and set $Qf = f * L$. We show that L is well-defined and possesses a certain degree of regularity, and that $QA = AQ = I$. Again split $I_{-1} = I^0_{-1} + I^\infty_{-1}$. $I^0_{-1} \in L^p$ whenever $1 \leq p < d/(d-1)$, while $I^\infty_{-1} \in L^p$ for all $p > d/(d-1)$. Therefore $I_{-1} * \ell \in L^{p_1} + L^{p_2}$ whenever $p_1 < d/(d-1)$ and $p_2 > d/(d-1)$. $I^0_{-1} \in L^1$, so $I_{-1} * \ell * I^0_{-1}$ is also a well-defined function in $L^{p_1} + L^{p_2}$. Since $d \geq 4, d/(d-1) \leq 4/3 < 2$. If p_1 is any exponent less than $d/(d-1)$ and p_2 is chosen to be less than 2, I^∞_{-1} belongs to both the dual classes $L^{p_1'}$ and $L^{p_2'}$, so that $I_{-1} * \ell * I^\infty_{-1}$ is a well-defined bounded, continuous function. So L is well-defined. Clearly it is homogeneous of degree $-d+2$, and we have just shown that it belongs to L^p away from 0, for all $p < d/(d-1)$.

Lemma 5.6. *There exist $q > 1, \sigma > 1$, and $\tau > 2$ such that L has simultaneously τ left-invariant and σ right-invariant derivatives in L^q.*

In other words $(I \text{ or } \widehat{\Delta}^{\sigma/2}) \circ (I \text{ or } \Delta^{\tau/2})(\overline{\eta} L) \in L^q$, where $\overline{\eta}(x) = \eta(\|x\|)$ (and I is the identity operator). A property of these Sobolev spaces already used implicitly in the proof of Lemma 5.2 is that they are closed under multiplication by smooth, compactly supported functions.

Proof. Split $I_{-1} = I^0_{-1} + I^\infty_{-1}$ once more, and $\ell = \ell^0 + \ell^\infty$. As in the proof of Lemma 5.2 it suffices to examine $I^0_{-1} * \overline{\ell} * I^0_{-1}$ where

$$\overline{\ell}(x) = (\zeta(c\|x\|) - \zeta(\|x\|))\ell(x), \qquad \zeta \in C^\infty_0(\mathbf{R}),$$

$\zeta \equiv 1$ near 0 and c is a small number. We know $\overline{\ell} \in L^q_\gamma$ where $\gamma > 1$. Let $\gamma = 1 + 2\varepsilon$. Since $L^q_\varepsilon \subset \widehat{L}^q_{m^{-1}\varepsilon}$ for compactly supported functions, $\overline{\ell}$ has simultaneously $1 + \varepsilon$ left- and $m^{-1}\varepsilon$ right-invariant derivatives in L^q. As in the proof of Lemma 5.2, the two factors of I^0_{-1} each add an extra derivative, one left and one right.

Lemma 5.7. *$QA = AQ = I$ on Schwartz functions.*

Proof. First, $A = \Delta^{1/2}(I+S)\Delta^{1/2}$ on $\mathscr{S}$. For the composition is well-defined on $\mathscr{S}$, by the same reasoning we have already applied to

$$S = \Delta^{-1/2}|[X, Y]|\Delta^{-1/2}.$$

For $h_1, h_2 \in \mathscr{S}$, $\langle h_1, Sh_2\rangle = \langle \Delta^{-1/2} h_1, |[X, Y]|\Delta^{-1/2} h_2\rangle$ by definition of $Sh_2 = h_2 * (I_{-1} * \mu * I_{-1})$ as $(h_2 * I_{-1} * \mu) * I_{-1}$. Therefore approximating by Schwartz functions and passing to the limit gives $\langle \Delta^{1/2} g, S\Delta^{1/2} f\rangle = \langle g, |[X, Y]|f\rangle$ for all $f, g \in \mathscr{S}$, so $\langle \Delta^{1/2} g, (I+S)\Delta^{1/2} f\rangle = \langle g, Af\rangle$ and hence $A = \Delta^{1/2}(I+S)\Delta^{1/2}$.

QAf is defined for $f \in \mathscr{S}$, since $Af \in L^1 \cap L^\infty$ and the same for all its right-invariant derivatives, and because of our estimates on L.

$$QAf = (\Delta^{-1/2}(I+S)^{-1}\Delta^{-1/2})(\Delta^{1/2}(I+S)\Delta^{1/2})f.$$

Applying the associative law may be justified as above, so $QAf = f$.

The inversion theorem implies a variant of itself.

Corollary. *Let $q \in (1, \infty)$. Suppose $Tf = f * k$ for all $f \in \mathscr{S}$, where $k \in \mathscr{S}'$ is homogeneous of degree $-d$. Suppose that $0 < \sigma, \tau$ and that $\sigma + \tau < d$. Suppose further that k has simultaneously σ left-invariant and τ right-invariant derivatives in L^q away from 0. If T is invertible on $L^2(G)$ then there exists $\ell \in \mathscr{S}'$, homogeneous of degree $-d$, so that $T^{-1}f = f * \ell$ for all $f \in \mathscr{S}$, and ℓ has simultaneously σ left-invariant and $\tau - \delta$ right-invariant derivatives in L^q, for all $\delta > 0$.*

Proof. Consider, for $0 \le \rho < \tau$, $T_\rho = \Delta^{-\rho/2} \circ T \circ \Delta^{\rho/2}$. It may easily be verified that T_ρ is well-defined on $\mathscr{S}$ and is given by convolution with a distribution k_ρ homogeneous of degree $-d$, which belongs to L^q_s away from 0, where $s = \sigma + \rho$. In fact, this holds for $s = \sigma + \rho + m^{-1}(\tau - \rho)$, for k_ρ has simultaneously $\sigma + \rho$ left-invariant and $\tau - \rho$ right-invariant derivatives in L^q away from 0 and the latter may be converted to left-invariant derivatives as in the proof of Lemma 5.3. Let $\ell_\rho \in \mathscr{S}'$ be the convolution kernel for $\Delta^{-\rho/2} \circ T^{-1} \circ \Delta^{\rho/2}$, which is well-defined on $\mathscr{S}$ provided $\ell \in \widehat{L}^q_\rho$ away from 0.

By the inversion theorem, $\ell = \ell_0 \in L_{\sigma + m^{-1}\tau}$ away from 0. Therefore ℓ_0 has simultaneously σ left-invariant and $m^{-2}\tau$ right-invariant derivatives in L^q away from 0, so $\ell_\varrho \in L^q_{\sigma+\varrho}$ away from 0 where $\varrho = m^{-2}\tau$. Therefore $\Delta^{-\varrho/2} T^{-1} \Delta^{\varrho/2}$ extends to a bounded operator on L^2. Clearly this means that T_ϱ is invertible on L^2.

Now we iterate. Let $\rho \in (0, \tau)$ and suppose that T_ρ is invertible on L^2. Then by the inversion theorem $\ell_\rho \in L^q_{\sigma+\rho+m^{-1}(\tau-\rho)}$ away from 0. So ℓ_ρ has simultaneously $\sigma + \rho$ left- and $m^{-2}(\tau - \rho)$ right-invariant derivatives in L^q

away from 0. Therefore $T_{\rho+\varepsilon}$ is invertible on L^2, with $\varepsilon = m^{-2}(\tau - \rho)$, for its inverse is $\Delta^{-\varepsilon/2} T_\rho^{-1} \Delta^{\varepsilon/2}$, which has kernel in $L^q_{\sigma+\rho+\varepsilon}$ away from 0.

Therefore for each $\rho < \tau$, $\ell_\rho \in L^q_{\sigma+\rho}$ away from 0. This implies that ℓ, the convolution kernel for $T^{-1} = \Delta^{\rho/2} T_\rho^{-1} \Delta^{-\rho/2}$, satisfies the stated conclusion with $\tau - \delta = \rho$.

Actually the result remains valid when $\delta = 0$, but the weaker version will suffice.

As a consequence we can derive the main result of this section, a strengthened form of Lemma 5.6.

Proposition 5.6′. *For each $\varepsilon > 0$ there exist $q > 1$ and $\delta > 0$ such that the convolution kernel for A^{-1} has simultaneously $2 - \varepsilon$ left-invariant and $2 + \delta$ right-invariant derivatives in L^q, away from 0.*

Indeed by previous arguments the kernel for $I + S$ has simultaneously $1 - \varepsilon$ left- and $1+\delta$ right-invariant derivatives, so the strengthened inversion theorem may be combined with the proof of Lemma 5.6 to give 5.6′. Note that the roles of left- and right-regularity may equally well be reversed in the conclusion.

Lemma 5.8. *$I^0_\gamma * \mu \in L^1$ whenever $-d < \gamma < -2$.*

The hypothesis $\gamma > -d$ is surely unnecessary, but is made in order to avoid discussing $\Delta^{\gamma/2}$ for $\gamma \le -d$.

Proof. $I^0_\gamma * \mu^\infty \in L^1$ since $I^0_\gamma \in L^1$ and μ^∞ is a finite measure. $I^0_\gamma * \mu^0$ is compactly supported, so it suffices to show that $I^0_\gamma * \mu \in L^1_{\ell oc}$. Now write $I^0_\gamma * \mu = I_\gamma * \mu - I^\infty_\gamma * \mu$. The latter term is clearly bounded, so it remains to prove $I_\gamma * \mu \in L^1_{\ell oc}$. Now $I_\gamma * \mu$ is homogeneous of degree $-d - (2+\gamma) > -d$, so it suffices to see that $I_\gamma * \mu \in L^1$ on $\{\|x\| \sim 1\}$. We may split $I_\gamma = I^0_\gamma + I^\infty_\gamma$ and $\mu = \mu^0 + \mu^\infty$ so that $I^0_\gamma * \mu^0$ is identically zero on $\{\|x\| \sim 1\}$, so there remain the three terms $I^\infty_\gamma * \mu^0$, $I^0_\gamma * \mu^\infty$, $I^\infty_\gamma * \mu^\infty$. The first is C^∞, the second is in $L^1(G)$ and the third is C^∞.

Corollary 5.9. *$\|[X, Y]\|$ is well-defined on L^p_γ for all $\gamma > 2$ and $p \in (1, \infty)$, and it maps L^p_γ to L^p.*

We change notation: for the rest of this paper

$$\ell \text{ denotes the convolution kernel for } A^{-1}$$

and L will be assigned a new meaning. Let $\zeta \in C_0^\infty(\mathbf{R}^+)$ be an auxiliary function satisfying $\int_0^\infty \zeta(t)\, dt/t = 1$. On G set $\zeta^t(x) = \zeta(t^{-1}\|x\|) \in C_0^\infty$ and decompose

$$\ell = \int_0^\infty \ell_t \frac{dt}{t}$$

with $\ell_t = \zeta^t \ell$; note that $\ell_t(x) = t^{-d+2}\ell_1(\delta_t^{-1}x)$, where $\{\delta_t\}$ is the dilation group on G.

Lemma 5.10. *For each $\sigma < 2$ there exists $q > 1$ such that $\mu * \int_1^\infty \ell_t\, dt/t$ belongs to $L^q_\sigma(G)$.*

Proof. The involution $x \to x^{-1} = -x$ interchanges L^q_σ and $\widehat{L}^q_\sigma$. μ, ℓ_t are both symmetric about 0 if the norm $\|\cdot\|$ is chosen to be symmetric, so it is equivalent to show that $(\int_1^\infty \ell_t\, dt/t) * \mu \in \widehat{L}^q_\sigma$. By Corollary 5.10 and Proposition 5.6′, $\ell_1 * \mu \in L^q$ if q is sufficiently close to 1. In fact if q is chosen so that ℓ_1 has simultaneously σ right- and $2+\varepsilon$ left-invariant derivatives, then $\widehat{\Delta}^{\sigma/2}(\ell_1 * \mu) = (\widehat{\Delta}^{\sigma/2}\ell_1) * \mu$ with the factor $\widehat{\Delta}^{\sigma/2}\ell_1 \in L^q_{2+\varepsilon}$. Thus $\widehat{\Delta}^{\sigma/2}(\ell_1 * \mu) \in L^q$ by Corollary 5.9.

Now by homogeneity $(\ell_t * \mu)(x) \equiv t^{-d}(\ell_1 * \mu)(\delta_t^{-1}x)$. The L^q norm of $t^{-d}g(\delta_t^{-1}\cdot)$ is $t^{-d(1-q^{-1})}\|g\|_q$. Similarly

$$\|\widehat{\Delta}^{\sigma/2}t^{-d}g(\delta_t^{-1}\cdot)\|_q = t^{-d(1-q^{-1})-\sigma}\|\widehat{\Delta}^{\sigma/2}g\|_q.$$

Therefore

$$\int_1^\infty \|\ell_t * \mu\|_{\widehat{L}^q_\sigma}\, dt/t < \infty.$$

Fix an auxiliary function $\varphi \in C_0^\infty(G)$ satisfying $\varphi(x) \equiv \varphi(x^{-1})$ and $\int \varphi = 1$, and set $\varphi_t(x) = t^{-d}\varphi(\delta_t^{-1}x)$. We abuse notation by letting the subscript "t" have different meanings when attached to the different letters φ, ℓ, L, but we shall be entirely consistent in the usage of each of the three letters so that there need be no confusion.

Let $\omega > 0$ denote a very small constant. The calculations in §§6 through 14 will require that ω be chosen sufficiently small, depending only on m, which equals both the step of the free group G presently under consideration and the type of the pseudoconvex manifold M with which we began. Set

$$L_t = \ell_t * \varphi_{t^{1+\omega}} \tag{5.1}$$

and

$$L = \int_0^1 L_t \frac{dt}{t}. \tag{5.2}$$

Then it is illuminating to consider the rescaled kernels

$$L'_t(x) = t^{d-2}L_t(\delta_t x) = \ell_1 * \varphi_{t^\omega}(x).$$

Lemma 5.11. *For each $\sigma < 2$ there exist $\tau > 2$ and $q > 1$ such that L'_t has simultaneously τ right-invariant and σ left-invariant derivatives in $L^q(G)$, uniformly in $t \in (0, 1]$.*

Proof. Each $L'_t \in C_0^\infty$, so the uniformity is the issue. By Proposition 5.6$'$, it suffices to prove that

$$\|h * \varphi_s\|_{L^q_\sigma} \prec \|h\|_{L^q_\sigma} \ \forall\, h \in L^q_\sigma ,$$

for all $q \in (1, \infty)$ and $\sigma \geq 0$, uniformly in $s \in \mathbf{R}^+$. Certainly $\|h * \varphi_s\|_q \prec \|h\|_q$ since $\|\varphi_s\|_1$ is a finite constant. And $\Delta^{\sigma/2}(h * \varphi_s) = h * \varphi_s * I_\sigma = (h * I_\sigma) * (I_{-\sigma} * \varphi_s * I_\sigma)$, and since $h * I_\sigma \in L^q$, by rescaling it suffices to show that $I_{-\sigma} * \varphi * I_\sigma \in L^1$; the s-dependence cancels out. $I_{-\sigma} * \varphi * I_\sigma$ is C^∞, so the issue is its decay at infinity. Note that $I_{-\sigma} * I_\sigma$ is the Dirac mass δ at 0, hence vanishes identically away from 0. On the other hand straightforward computation yields the pointwise bound

$$|I_{-\sigma} * \varphi * I_\sigma(x) - I_{-\sigma} * I_\sigma(x)| \prec \|x\|^{-d-1}$$

as $\|x\| \to \infty$. For when $\|x\| \sim r$ we can make the change of variables $x = \delta_r y$ to reduce the inequality to

$$|I_{-\sigma} * (\varphi_{r^{-1}} - \delta) * I_\sigma(x)| \prec r^{-1} \quad \text{when } \|x\| \sim 1 ,$$

which is apparent.

Lemma 5.12. *For each $\sigma < 2$ there exist $\tau > 2$ and $q > 1$ such that $L'_t - \ell_1 = \ell_1 * (\varphi_{t^\omega} - \delta)$ has simultaneously τ right-invariant and σ left-invariant derivatives in L^q, with bound $O(t^\varepsilon)$ for some $\varepsilon(\omega, \sigma) > 0$, uniformly for $t \in (0, 1]$.*

More precisely the assertion is that

$$L'_t - \ell_1, \quad \Delta^{\sigma/2}(L'_t - \ell_1), \quad \widehat{\Delta}^{\tau/2}(L'_t - \ell_1), \quad \text{and} \quad \widehat{\Delta}^{\tau/2}\Delta^{\sigma/2}(L'_t - \ell_1)$$

all belong to L^q, with norms $0(t^\varepsilon)$. δ denotes here the Dirac mass at $0 \in G$.

Proof. Fix $\sigma' \in (\sigma, 2)$ and choose τ, q so that ℓ_1 has simultaneously σ' left- and τ right-invariant derivatives in L^q. It will suffice to prove that there exists $\varepsilon' > 0$ such that $\|\ell_1 * (\delta - \varphi_s)\|_q \prec s^{\varepsilon'}$ as $s \to 0$. For $\|\ell_1 * (\delta - \varphi_s)\|_{L^q_{\sigma'}} \prec 1$ by Lemma 5.11, and then an interpolation yields $\|\ell_1 * (\delta - \varphi_s)\|_{L^q_\sigma} \prec s^\varepsilon$ for some smaller $\varepsilon > 0$, since σ is between 0 and σ'. As usual it is no problem to introduce $\widehat{\Delta}^{\tau/2}$, which commutes with both $\Delta^{\sigma/2}$ and convolution on the right with $\delta - \varphi_s$.

To control $\|\ell_1 * (\delta - \varphi_s)\|_q$ it suffices to see that $\|I^0_{-\sigma} * (\delta - \varphi_s)\|_1 \prec s^{\varepsilon'}$. This follows from the facts that (1) $\delta - \varphi_s$ annihilates constants and is supported where $\|x\| \prec s$, and (2) $|I^0_{-\sigma}(x)| \prec \|x\|^{-d+\sigma}$ and $|(X \text{ or } Y) I^0_{-\sigma}(x)| \prec \|x\|^{-d+\sigma-1}$ and similarly for higher-order derivatives, and $I^0_{-\gamma}$ has compact support. We omit the details.

Lemma 5.13. $\|DL'_t\|_\infty \leq C_k t^{(-d-k)\omega}$ *for each homogeneous left-invariant differential operator D of weight $-k$, $k \geq 0$.*

Proof. $DL'_t = \ell_1 * (D\varphi_{t^\omega})$, $\ell_1 \in L^1$ and $\|D\varphi_{t^\omega}\|_\infty \prec t^{(-d-k)\omega}$ by homogeneity.

Remark. ℓ is actually C^∞ away from 0, off of the single line $\{\exp(t[X,Y])0 : t \in \mathbf{R}\}$; this follows from another version of the inversion theorem [C2] since k is C^∞ off of the line, away from 0. However Proposition 5.6′ is the only control we have been able to obtain over ℓ on that line. The situation is perhaps delicate, for although this line forms a homogeneous subgroup of G, it is an unstable subgroup, provided the step m is greater than two. To make this precise define the distance $r(x)$ from x to the line to be $\inf\{\|x^{-1}y\| : y \in \text{line}\}$. For $\varepsilon > 0$ let C_ε be the smallest constant such that wherever $1 \le \|x\|$, $\|u\| \le 2$ and $r(x), r(u) \le \varepsilon$, $r(xu) \le \varepsilon C_\varepsilon$. Then "unstable" means that C_ε blows up as $\varepsilon \to 0$. We make no use of the C^∞ smoothness of ℓ off the exceptional line.

6. Definition of the parametrix. Comments

In this section we introduce a new object and make a change in notation which will be in force through §10. The two basic left-invariant vector fields on the free nilpotent group G will be denoted $\widehat{X}$ and $\widehat{Y}$. X, Y will denote two C^∞ real vector fields in a neighborhood U of $0 \in \mathbf{R}^N$, free to order m, where m is the step of G and N its dimension. There will be a correspondence between certain objects on U and objects on G, and the latter will be distinguished by the superscript " $\widehat{\ }$ ". Thus $\Delta = -X^2 - Y^2$, $\widehat{\Delta} = -\widehat{X}^2 - \widehat{Y}^2$. $\widehat{B} = |[\widehat{X}, \widehat{Y}]|$ corresponds to

$$Bf(x) = \int \mathscr{A}(t)\zeta(t) f(\gamma(x,t))\, dt$$

where $\zeta \in C_0^\infty(\mathbf{R})$ is identically equal to one near 0, and

$$\gamma(x,t) = \exp\left(\sum_{j=1}^{m} t^j V_j\right)x + O(t^{m+1})$$

where $V_1 = [X,Y] + O(X,Y)$ and for $j > 1$, $V_j = O(X, Y, [X,Y])$, and where $V = O(X,Y)$ means that V is a linear combination of X, Y with C^∞ coefficients. $\widehat{A} = \widehat{\Delta} + \widehat{B}$, while $A = \Delta + B$. The term " $O(t^{m+1})$ " indicates that for any N, the C^N norm of $\gamma(x,t) - \exp(\cdots)$ is $O(N)$; the subtraction is carried out in some fixed Euclidean coordinate system on U.

Let $\widehat{U}$ denote a small neighborhood of $0 \in G$. Fix a homogeneous basis $\{\widehat{Z}_\alpha\}$ for the Lie algebra $\mathbf{g}$ of G. This means that each $\widehat{Z}_\alpha = [(\widehat{X} \text{ or } \widehat{Y}), [(\widehat{X} \text{ or } \widehat{Y}), \ldots]$ with $w(\alpha)$ factors, and $\{\widehat{Z}_\alpha\}$ is a basis for $\mathbf{g}$ as a vector space. Recall the linear map ϕ from $\mathbf{g}$ to the Lie algebra of vector fields on U generated by X, Y, defined by $\phi(\widehat{Z}_\alpha) = Z_\alpha$, where Z_α is obtained by replacing each factor of $\widehat{X}$ by X and of $\widehat{Y}$ by Y. In general ϕ will not be a Lie algebra homomorphism, but since X, Y are free up to order m it satisfies

$$\phi[\widehat{Z}_\alpha, \widehat{Z}_\beta] = [Z_\alpha, Z_\beta] \quad \text{whenever } w(\alpha) + w(\beta) \le m. \tag{6.1}$$

Define

$$\Theta : U \times \widehat{U} \to \mathbf{R}^N$$

by $\Theta(x,u) = \exp(\sum u_\alpha Z_\alpha)x = \exp(\phi(u\cdot\widehat{Z}))x$ where we write $u\cdot\widehat{Z} = \sum_\alpha u_\alpha \widehat{Z}_\alpha$ and $u = (u_\alpha) = \exp(u\cdot\widehat{Z})0$. To simplify the notation, $x \times u$ will frequently be written instead of $\Theta(x,u)$. $x \times u \times v$ is of course to be associated as $(x \times u) \times v$. By the hypothesis that X, Y are free to order m, the map $u \to x \times u$ is a diffeomorphism from $\widehat{U}$ into $\mathbf{R}^N$, for all x near $0 \in \mathbf{R}^N$. Let $\Theta^{-1}(x,\cdot)$ denote the inverse diffeomorphism, so that $\Theta^{-1} : U \times U \to G$ and $\Theta(x, \Theta^{-1}(x,y)) \equiv y$.

Θ may be thought of as defining an approximate group law, indeed an approximation to the group structure of G, on U. If $u, v \in G$ are close to 0 then $(x \times u) \times v$ is very nearly equal to $x \times (uv)$. To make this precise, define the weight of a smooth function $F : \widehat{U} \times \widehat{U} \to \mathbf{C}$ by expanding F in Taylor series about $(0,0)$, and taking the minimum of the weights of all monomials in (u,v) which occur with nonzero coefficient. The weight of a monomial $u^\beta v^\gamma$ is $\sum_\alpha \beta_\alpha w(\alpha) + \sum_\alpha \gamma_\alpha w(\alpha)$. Say that $F = O(u,v)^M$ if F is of weight $\geq M$. For functions on $\widehat{U} \times \widehat{U} \times \mathbf{R}$ we shall usually define the weight of $u^\beta v^\gamma t^n$ to be $\sum(\beta_\alpha + \gamma_\alpha)w(\alpha) + 2n$, and accordingly define the weight of $F : \widehat{U} \times \widehat{U} \times \mathbf{R} \to \mathbf{C}$. The same goes for $F : \widehat{U} \times \mathbf{R} \to \mathbf{C}$, and so on.

Since $u \to x \times u$ is a diffeomorphism there exists for each $x \in U$ a smooth

$$Q = Q_x : \widehat{U} \times \widehat{U} \to \mathbf{G}$$

such that $(x \times u) \times v \equiv x \times Q_x(u,v)$. Q_x depends smoothly on x, as well. The precise form of the approximate associative law is

Lemma 6.1. $Q(u,v) = uv + O(u,v)^{m+1}$.

Letting $(uv)_\alpha$ be the α-th component of the product uv, and $Q(u,v)_\alpha$ the α-th component of $Q(u,v)$, the assertion is that for each α, $Q(u,v)_\alpha - (uv)_\alpha = O(u,v)^{m+1}$. (On G there are canonical coordinates defined by the exponential map and the fixed basis $\{\widehat{Z}_\beta\}$ for g, so G is naturally identified with $\mathbf{R}^N$. The subtraction, and similar additions and subtractions elsewhere in §§6–10, are taken with respect to the usual Euclidean group structure on $\mathbf{R}^N$, in these coordinates.)

For the proof apply the Baker-Campbell-Hausdorff formula [RS, Appendix] to deduce that

$$\exp(v\cdot Z)\exp(u\cdot Z)x = \exp(\widetilde{Q}(u,v)\cdot Z)x + O(u,v)^{m+1}$$

where $\widetilde{Q}$ is an $\mathbf{R}^N$-valued polynomial, a sum of monomials of weights $\leq m$. Therefore $Q(u,v) = \widetilde{Q}(u,v)$ modulo $O(u,v)^{m+1}$. By the Baker-Campbell-Hausdorff formula, $\widetilde{Q}(u,v)\cdot Z$ is a finite linear combination of $u\cdot Z, v\cdot Z$ and their iterated commutators, and by (6.1) $\widetilde{Q}(u,v)\cdot Z \equiv (uv)\cdot Z$, so the lemma is correct.

Corollary 6.2 (Triangle inequality). $\|Q_x(u,v)\| \leq C(\|u\| + \|v\|)$ *uniformly for* $x \in U$, $u, v \in \widehat{U}$.

Here and elsewhere it is implicitly assumed that U and $\widehat{U}$ are sufficiently small, and they are permitted to shrink from one occurrence to the next.

Proof. Write $Q_x(u,v) = uv + R_x(u,v)$. By the triangle inequality on G,

$$\|Q_x(u,v)\| \prec (\|uv\| + \|(uv)^{-1}Q_x(u,v)\|) \prec \|u\| + \|v\| + \|(uv)^{-1}Q_x(u,v)\|.$$

But by the form of the group law on G, $(uv)^{-1}Q_x(u,v) = (uv)^{-1}\cdot(uv) = 0$ plus a polynomial in (uv) and $R_x(u,v)$, all of whose terms include at least one factor of some coordinate function $(R_x(u,v))_\alpha$. Thus $(uv)^{-1}Q_x(u,v) = O(u,v)^{m+1}$. For any homogeneous norm, $\|O(u,v)^{m+1}\| \prec (\|u\| + \|v\|)^{1+m^{-1}}$.

The method of Rothschild and Stein [RS] for constructing a parametrix for Δ out of the convolution kernel K for $\widehat{\Delta}^{-1}$ relies on the approximate group law Θ. Let η_0, η_1 be functions in $C_0^\infty(U)$, $\eta_0 \equiv 1$ in a neighborhood of 0, $\eta_1\eta_0 \equiv \eta_0$. [RS] define a parametrix for Δ by

$$Pf(x) = \eta_1(x)\int_G K(u)f(x\times u^{-1})\eta_0(x\times u^{-1})\,du. \tag{6.2}$$

Then $P(\Delta f)$ is calculated by integration by parts. The principal term in the outcome is

$$\eta_1(x)\int(\widetilde{\Delta}_x K)(u)f(x\times u^{-1})\eta_0(x\times u^{-1})\,du,$$

where $\widetilde{\Delta}_x$ is a differential operator on G which depends on x and equals $\widehat{\Delta}$ modulo error terms. Because of Lemma 6.1, the error terms are of weight ≥ -1; recall that $\widehat{\Delta}$ is of weight -2 and $\widehat{X}$ of weight -1 so that these are indeed lower-order terms, in the sense of weights. Since $\widehat{\Delta}K = \delta_{u=0}$, what must be proved is that the error terms, applied to K, give rise to smoothing kernels. A typical error term might be $u_1^2\widehat{W}$, where $\widehat{W}$ is one of the $\widehat{Z}_\alpha$, with $w(\alpha) = 3$, and u_1 is a coordinate function of weight one. K is known to satisfy $|\widehat{D}K(u)| \leq C_n\|u\|^{-d+2-n}$ for any left-invariant differential operator $\widehat{D}$ homogeneous of weight $-n$, so $|u_1^2\widehat{W}K(u)| \leq C\|u\|^{-d+2-3+2} = C\|u\|^{-d+1}$ Only small values of u come into play in (6.2), and the degree of homogeneity $-d+1$ is larger than the critical degree $-d$, so this is the sort of size estimate one would expect for a fractional integral operator, smoothing of order one. The factor $\widehat{W}$ is actually of higher order than $\widehat{\Delta}$, but it is adequately compensated for by u_1^2.

The natural approach to the construction of a parametrix for A is to $\widetilde{K}$ by ℓ, the kernel for $\widehat{A}^{-1}$, in (6.2). Unfortunately this fails badly. Again one is led to kernels such as $u_1^2\widehat{W}\ell(u)$. The kernel ℓ is not C^∞ away from 0, and so far as we know has only $2+\varepsilon$ derivatives (right-invariant) in some L^q; indeed if it had too many derivatives then a reversal of the reasoning of §5 would demonstrate that μ had some positive fraction of a derivative in L^q, a contradiction. Therefore since $\widehat{W}$ is of weight -3 it may be expected that $\widehat{W}\ell$

will not even be a function, away from 0, and so a straightforward application of the method of [RS] would appear to fail for A.

In actuality the situation is less clear. As mentioned in §5, ℓ is actually C^∞ off of the line $\{\exp(t[X, Y])0 : t \in \mathbf{R}\}$, so $\widehat{W}\ell$ will be C^∞ off this line and will blow up in some fashion at the line. The factor u_1^2 vanishes on the line, and might conceivably vanish more rapidly than $\widehat{W}\ell$ blows up. However we see no way to rule out the presence of error terms $u_3\widehat{W}$ where u_3 is the coordinate corresponding to $[X, Y]$, so that $u_3 \neq 0$ on the line of singularity.

However a modification of the procedure of [RS] does work. Replace ℓ by the mollification L defined in (5.1) and (5.2). Define

$$\mathscr{P}_0 f(x) = \eta_1(x) \int_G L(u) f(x \times u^{-1}) \eta_0(x \times u^{-1})\, du. \tag{6.3}$$

The obvious disadvantage of L is that it is not the kernel for $\widehat{A}^{-1}$. Therefore in showing that $\mathscr{P}_0 \circ A = I$ plus a smoothing term, one is required to show, for instance, that the kernel $\widehat{\Delta}(L - \ell)$ gives rise to a smoothing term, and to handle further new error terms. The advantage of L over ℓ is that it is smooth away from 0. To see how this might be used consider $u_1^2 \widehat{W} L(u)$, where $\widehat{W}$ has weight -3. By Lemma 5.13

$$\begin{aligned} |u_1^2 \widehat{W} L(u)| &\prec \|u\|^2 |\widehat{W} L(u)| \\ &\prec \|u\|^2 \|u\|^{-d+2-3-C\omega} \\ &= \|u\|^{-d+(1-C\omega)} = \|u\|^{-d+\varepsilon} \end{aligned}$$

where $\varepsilon > 0$, provided ω is chosen smaller than C^{-1}. This suggests that the error terms should still be smoothing, though not of one full degree. Of course $\mathscr{P}_0 \circ B$ must also be analyzed.

§§7 through 10 go as follows. In §7 we decompose $\mathscr{E}_0 = I - \mathscr{P}_0 A$ as a sum of a number of simpler error operators. By estimating each we show in §8 that the kernel associated to $\mathscr{E}_0$ satisfies the sort of size estimates one would hope for in a smoothing operator. In §9 we prove more delicate estimates which imply that $\mathscr{E}_0$ is smoothing of some positive order.

Unfortunately part of what was required for the bootstrapping arguments of §2 was that $\mathscr{E}$ should be smoothing of order one in the "good" directions X and Y, while it will be clear from the proofs that $\mathscr{E}_0$ can only be expected to be smoothing of some very small order. Therefore we introduce a better parametrix

$$\mathscr{P} = (I + \mathscr{E}_0 + \mathscr{E}_0^2 + \cdots + \mathscr{E}_0^M)\mathscr{P}_0$$

so that

$$\mathscr{P} A = I + \mathscr{E}, \quad \mathscr{E} = -\mathscr{E}_0^{M+1}$$

where M is a large number. If M is chosen sufficiently large then $\mathscr{E}_0^{M+1}$ will be smoothing of any preassigned degree.

In §10 we study the mapping properties of $\mathscr{P}_0$. The main point is that composing it with two factors of (X or Y) yields a type of Calderón-Zygmund singular integral operator. Finally we analyze the composition of $\mathscr{P}$ with two factors of (X or Y).

7. Computation of $\mathscr{E}_0$

The goal of §§7 and 8 is to establish

Proposition 7.1. *$\mathscr{E}_0 f(x) = \int k(x,y) f(y)\,dy$ where for all small $r > 0$,*

$$\sup_x \int_{\rho(x,y)\sim r} |k(x,y)|\,dy + \sup_y \int_{\rho(x,y)\sim r} |k(x,y)|\,dx \prec r^{\varepsilon} \tag{7.0}$$

for some $\varepsilon = \varepsilon(m) > 0$.

$\rho(x,y)$ is defined to be $\|u\|$, where $y = x \times u$. In the present section an explicit representation of $\mathscr{E}_0 = I - \mathscr{P}_0 A$ as a sum of simpler terms (7.2E, 7.3E, 7.4E, 7.8E, 7.10E, 7.11E and 7.12E) will be derived. Proposition 7.1 will be proved in §8.

To begin the analysis of $\mathscr{P}_0 \Delta f$ let $x \in U$ be arbitrary, write $g(u) = f(x \times u)$ and define a second order differential operator $\tilde{\Delta}$ on $\hat{U}$, depending smoothly on x, by

$$\tilde{\Delta} g(u) = (\Delta f)(x \times u).$$

Lemma 7.2. *$\tilde{\Delta} = \hat{\Delta} + \sum_\alpha b_\alpha(x,u)\hat{Z}_\alpha + \sum_{\beta,\gamma} b_{\beta\gamma}(x,u)\hat{Z}_\beta \hat{Z}_\gamma$ where each $b_\alpha(x,u)\hat{Z}_\alpha$ and each $b_{\beta\gamma}(x,u)\hat{Z}_\beta\hat{Z}_\gamma$ is of weight greater than or equal to -1, for all x.*

The weight of $b_\alpha(x,u)\hat{Z}_\alpha$ is the weight of $b_\alpha(x,\cdot)$ plus the weight $-w(\alpha)$ of $\hat{Z}_\alpha$, where the weight of $b_\alpha(x,\cdot)$ is the minimum of the weights of all monomials occurring with nonzero coefficients in its Taylor expansion about $u = 0$. The interpretation is that $\tilde{\Delta} - \hat{\Delta}$ is in a sense of order 1, while $\hat{\Delta}$ itself has order 2.

Proof. To X and Y associate $\tilde{X}$ and $\tilde{Y}$, in the same way that $\tilde{\Delta}$ is associated to Δ. It suffices to prove that $\tilde{X} - \hat{X}$ and $\tilde{Y} - \hat{Y}$ are of weight ≥ 0, for $\tilde{\Delta} = -\tilde{X}^2 - \tilde{Y}^2$ and the sought-after conclusion follows. Now

$$\begin{aligned}(Xf)(x \times u) &= \frac{d}{dt} f(\exp(tX)(x \times u))|_{t=0} = \frac{d}{dt} f(\exp(tX)\exp(u \cdot Z)x)|_{t=0} \\ &= \frac{d}{dt} f(\exp(Q(t,u) \cdot Z)x)|_{t=0} = \frac{d}{dt} g(Q(t,u))|_{t=0}\end{aligned}$$

where the second-to-last equality defines $Q(t,u)$. By the Baker-Campbell-Hausdorff formula it follows as in Lemma 6.1 that if $v = u \cdot \exp(t\hat{X})0$ then $Q(t,u) = v$ modulo $O(u,t)^{m+1}$, where t is assigned weight one. Therefore the chain rule gives $Xf(x \times u) = (\hat{X}g)(u) + \sum_\alpha b_\alpha(x,u)\hat{Z}_\alpha g(u)$ where each b_α is of weight $\geq m$, hence each $b_\alpha(x,u)\hat{Z}_\alpha$ has weight $\geq m - w(\alpha) \geq 0$.

By the definition (6.3) of $\mathscr{P}_0$,

$$\begin{aligned}\mathscr{P}_0\Delta f(x) &= \eta_1(x)\int L(-u)(\Delta f)(x\times u)\eta_0(x\times u)\,du\\ &= \eta_1(x)\int L(-u)(\widetilde{\Delta} g)(u)\eta_0(x\times u)\,du\,,\end{aligned}$$

recalling that $-u \equiv u^{-1}$ in our coordinates on G. Integrate by parts and assume that the manipulation is valid—this is in some doubt because L is singular at $u=0$. A basic identity is that if $\widehat{D}$ is any left-invariant differential operator on G and $\widehat{D}'$ its right-invariant analogue, then

$$\int_G h(-u)(\widehat{D}g)(u)\,du = \int_G (\widehat{D}'h)(-u)g(u)\,du$$

for any $g, h \in \mathscr{S}$. Therefore the main term resulting from the integration by parts is

$$\eta_1(x)\int(\widehat{\Delta}'L)(-u)f(x\times u)\eta_0(x\times u)\,du \tag{7.1}$$

where $\widehat{\Delta}'$ is the right-invariant analogue of $\widehat{\Delta}$. $\widehat{\Delta}'L$ is a distribution, and (7.1) is interpreted as described below. There arise two types of error terms, one from the discrepancy between $\widetilde{\Delta}$ and $\widehat{\Delta}$, and one from the falling of some of the u-derivatives on η_0 rather than on L. Both are of the form

$$\eta_1(x)\int(DL)(-u)f(x\times u)\,du \tag{7.2E}$$

where D has C^∞ coefficients in (x,u), and is of weight ≥ -1 for all x. This is immediate for the latter type of error and follows directly from Lemma 7.2 for the former. Moreover each term of D involves at most two factors $\widehat{Z}_\beta$, $\widehat{Z}_\gamma$. As functions of u, the coefficients of D have no larger support than does $\eta_0(x\times u)$.

To justify the integration by parts let $\rho > 0$, replace $L = \int_0^1 L_t\,dt/t$ by $\int_\rho^1 L_t\,dt/t \in C^\infty$ and consider $\eta_1(x)\int(\int_\rho^1 L_t(-u)\,dt/t)\,\Delta f(x\times u)\eta_0(x\times u)\,du$. If $f\in C^2$ the limit as $\rho\to 0$ exists and equals $\mathscr{P}_0\Delta f(x)$. The integration by parts is valid for $\rho>0$ and yields

$$\eta_1(x)\lim_{\rho\to 0}\int\Big(\int_\rho^1 \widehat{\Delta}'L_t(-u)\,dt/t\Big)f(x\times u)\eta_0(x\times u)\,du.$$

Lemma 7.3. $\lim_{\rho\to 0}\langle\int_\rho^1 \widehat{\Delta}'L_t(-u)\,dt/t\,,h\rangle$ *exists for* $h\in C^2$ *and equals* $(\widehat{\Delta}h)*L(0)$.

Proof.

$$\int_\rho^1 \widehat{\Delta}'L_t(-u)\,\frac{dt}{t} = \widehat{\Delta}'\int_\rho^1 \ell_t(-u)\,\frac{dt}{t} + \int_\rho^1 \widehat{\Delta}'(L_t-\ell_t)(-u)\,\frac{dt}{t}.$$

We will see below that $\|\widehat{\Delta}'(L_t - \ell_t)\|_{L^1} \prec t^\varepsilon$ for some $\varepsilon > 0$, so it remains to examine $\widehat{\Delta}' \int_\rho^1 \ell_t(-u)\,dt/t$. Certainly

$$\left\langle \int_\rho^1 \widehat{\Delta}' \ell_t(-u)\frac{dt}{t}, h \right\rangle = \int\int_\rho^1 \ell_t(-u)\frac{dt}{t}\widehat{\Delta}h(u)\,du$$
$$= \int\left(\int_0^1 \ell_t(-u)\frac{dt}{t}\right)\widehat{\Delta}h(u)\,du$$
$$- \int\left(\int_0^\rho \ell_t(-u)\frac{dt}{t}\right)\widehat{\Delta}h(u)\,du.$$

The ρ-dependent second term is $\langle \widehat{\Delta}' \int_0^\rho \ell_t(-u)\,dt/t, h\rangle = \langle k_\rho, h\rangle$, which must be interpreted in the sense of distributions. $k_\rho(u) \equiv \rho^{-d}k_1(\delta_\rho^{-1}u)$ since $\widehat{\Delta}'$ and ℓ are homogeneous. Moreover $k_1(u)$ is homogeneous of degree $-d$ as a function of u, for $\|u\|$ small. We know that $k_1 \in L^1$ away from 0, by Lemma 5.6, and k_1 annihilates constants since $\widehat{\Delta}$ does. Therefore it follows from standard reasoning about principal-value integrals that $\lim_{\rho\to 0}\langle k_\rho, h\rangle$ exists and equals zero for h in any Hölder class of positive order. That the original limit equals $(\widehat{\Delta}h) * L(0)$ is immediate since $\int_\rho^1 L_t\,dt/t \to L$ in L^1 norm.

The same type of reasoning based on truncation of the t-integral may be used to justify future manipulations; the details will henceforth be omitted.

Continue to examine the main term (7.1) by splitting it as an error,

$$\eta_1(x)\int\int_0^1 \widehat{\Delta}'(L_t - \ell_t)(-u)\frac{dt}{t}\, f(x\times u)\eta_0(x\times u)\,du\,, \tag{7.3E}$$

minus another error,

$$\eta_1(x)\int \widehat{\Delta}'\int_1^\infty \ell_t(-u)\frac{dt}{t}\, f(x\times u)\eta_0(x\times u)\,du\,, \tag{7.4E}$$

plus the principal term

$$\eta_1(x)\int (\widehat{\Delta}'\ell)(-u)\, f(x\times u)\eta_0(x\times u)\,du\,, \tag{7.5}$$

which integral must be interpreted in the distribution sense. Thus $\mathscr{P}_0\Delta f(x) =$ (7.5) plus error terms (7.2E), (7.3E), and (7.4E).

The next step is to decompose $\mathscr{P}_0 \circ B$ in an analogous way. Recall that

$$Bf(x\times u) = \int \mathscr{A}(s)\varphi(s)f(\gamma(x\times u, s))\,ds\,, \tag{7.6}$$

where $\varphi \in C_0^\infty(\mathbf{R})$, $\varphi \equiv 1$ near 0. Write $g(u) = f(x\times u)$, with x fixed. Then

$$f(\gamma(x\times u, s)) = g(Q_1(u,s))\,,$$

where $Q_1 : \widehat{U}\times(-\varepsilon,\varepsilon) \to G$ is C^∞ in x, u, s. Since $Q_1(u,0) \equiv u$, the map $u \to Q_1(u,s)$ is a diffeomorphism of $\widehat{U}$ into G for small s, and $Q_2(v,s)$

denotes its inverse. In what follows the variable s will be consistently used as in (7.6), and will be assigned weight 2. $Q_i(u,s)_\alpha$ denotes the α-th component of $Q_i(u,s)$ and more generally u_α denotes the α-th component of any $u \in G$. Define $\bar{s} \in G$ by $\bar{s} = \exp(s[X,Y])0$.

Lemma 7.4. *$Q_1(u,s)_\alpha = (u \cdot \bar{s})_\alpha + O(u,s)^{w(\alpha)+1}$ and $Q_2(v,s)_\alpha = (v \cdot \bar{s}^{-1})_\alpha + O(v,s)^{w(\alpha)+1}$ for all α.*

Proof. $\gamma(x \times u, s) = \exp(\sum_{j=1}^m s^j V_j)\exp(u \cdot Z)x + O(s)^{m+1}$ so it suffices to examine $\exp(\sum s^j V_j)\exp(u \cdot Z)x$. This is done using the Baker-Campbell-Hausdorff formula as in the proof of Lemma 6.1. The main points in the argument are

(i) $V_1 = [X,Y] + O(X,Y)$, while $V_j = O(X,Y,[X,Y])$ for $j > 1$,

(ii) $\exp(s[X,Y])\exp(u \cdot Z)x = \exp((u\bar{s} + O(u,s)^{m+1}) \cdot Z)x$,

(iii) For any vector fields W_i and C^∞ functions $a_i, [a_1 W_1, a_2 W_2] = O([W_1, W_2], W_1, W_2)$.

(iv) $s\widehat{X}, s\widehat{Y}$ are of weight $+1$, and for $j > 1$, $s^j[\widehat{X}, \widehat{Y}]$ has weight $2j-2 \geq 1$ while $s^j(\widehat{X}$ or $\widehat{Y})$ has weight $2j-1 \geq 1$ also.

This lets us write

$$\begin{aligned}\mathscr{P}_0 Bf(x) &= \eta_1(x) \int L(-u) \int \mathscr{A}_0(s) g(Q_1(u,s))\, ds\, \eta_0(x \times u)\, du \\ &= \eta_1(x) \int\int L(-Q_2(v,s))\mathscr{A}_0(s) g(v) \eta_0(x \times Q_2(v,s)) J(x,v,s)\, ds\, dv\end{aligned}$$

where $\mathscr{A}_0(s) = \mathscr{A}(s)\varphi(s)$ and $J \in C^\infty$ is the Jacobian determinant from the change of variable; $J(x,v,0) \equiv 1$ since $Q_2(v,0) \equiv v$. Write this as a main term

$$\eta_1(x) \int \left(\int \mathscr{A}_0(s) L(-Q_2(v,s)) ds \right) f(x \times v) \eta_0(x \times v)\, dv \tag{7.7}$$

plus the error term

$$\eta_1(x) \int \left(\int \mathscr{A}_0(s) L(-Q_2(v,s)) \{\cdot\}\, ds \right) f(x \times v)\, dv \tag{7.8E}$$

where $\{\cdot\} = J(x,v,s)\eta_0(x \times Q_2(v,s)) - 1 \cdot \eta_0(x \times v)$. The vanishing of $\{\cdot\}$ when $s = 0$ makes (7.8E) indeed a lower-order term. Further analyze (7.7) as a main part

$$\eta_1(x) \int \left(\int \mathscr{A}_0(s) L(\bar{s}v^{-1})\, ds \right) f(x \times v) \eta_0(x \times v)\, dv \tag{7.9}$$

plus an error,

$$\eta_1(x) \int \left(\int \mathscr{A}_0(s) \{L(-Q_2(v,s)) - L(\bar{s}v^{-1})\}\, ds \right) f(x \times v) \eta_0(x \times v)\, dv. \tag{7.10E}$$

Of course the integrations with respect to s must always be interpreted in the distribution sense. (7.10E) will turn out to be a smoothing term because of Lemma 7.4. (7.9) may be further unraveled as

(7.11E)
$$\eta_1(x)\int\left(\int \mathscr{A}_0(s)\left[\int_0^1 L_t(\bar{s}v^{-1})-\ell_t(\bar{s}v^{-1})\frac{dt}{t}\right]ds\right)f(x\times v)\eta_0(x\times v)\,dv$$

plus

$$-\eta_1(x)\int\left(\int \mathscr{A}_0(s)\left[\int_1^\infty \ell_t(\bar{s}v^{-1})\frac{dt}{t}\right]ds\right)f(x\times v)\eta_0(x\times v)\,dv \tag{7.12E}$$

plus the ultimate principal term

$$\eta_1(x)\int\left(\int \mathscr{A}_0(s)\ell(\bar{s}v^{-1})\,ds\right)f(x\times v)\eta_0(x\times v)\,dv. \tag{7.13}$$

Terms (7.11E) and (7.12E) are expected to be error terms for reasons already spelled out in §5. Observe that the integrals with respect to s in both simply represent convolutions on G.

The sum of the two final main terms, (7.5) and (7.13), is

$$\begin{aligned}
&\eta_1(x)[(\widehat{\Delta}g)*\ell](0)+\eta_1(x)\int\int \mathscr{A}_0(s)\ell(\bar{s}w)\,ds\,g(w^{-1})\,dw\\
&=\eta_1(x)\left[(\widehat{\Delta}g)*\ell(0)+\int \ell(v)\int \mathscr{A}_0(s)g(v^{-1}\bar{s}^{-1})\,ds\,dv\right]\\
&=\eta_1(x)[(\widehat{\Delta}+\widehat{B})g*\ell](0)\\
&=\eta_1(x)[(\widehat{A}g)*\ell](0)
\end{aligned}$$

where $g(u)\equiv f(x\times u)\eta_0(x\times u)$. Thus (7.5) plus (7.13) equals $\eta_1(x)f(x)\eta_0(x)=f(x)$ provided f is supported where $\eta_0\equiv 1$.

8. Estimation of size of $\mathscr{E}_0$

Recall the metric $\rho(x,y)=\|\Theta^{-1}(x,y)\|=\|\Theta^{-1}(y,x)\|=\rho(y,x)$ on U. For each x,

$$|\{y:\rho(x,y)<t\}|\sim t^d \tag{8.0}$$

as $t\to 0$, since $y\to\Theta^{-1}(x,y)$ is a diffeomorphism and $|\{u\in G:\|u\|<t\}|\equiv ct^d$.

We estimate the error terms of the last section one by one. Begin with (7.2E) $=\eta_1(x)\int(DL)(-u)f(x\times u)\,du$, where D is of weight ≥ -1 and involves 0, 1 or 2 factors $\widehat{Z}_\beta$. By Lemma 5.13, $|DL(-u)|\prec\|u\|^{-d+2-(2m+d)\omega-1}=\|u\|^{-d+\varepsilon}$ where $\varepsilon>0$ provided ω is chosen smaller than $(2m+d)^{-1}$. The kernel $k(x,y)$ satisfying $(7.2E)=\int k(x,y)f(y)\,dy$ is $\eta_1(x)(DL)(-u)J(x,y)$ where $J=|\partial u/\partial y|\in C^\infty$ and $u=\Theta^{-1}(x,y)$. So we have $|k(x,y)|\prec\rho(x,y)^{-d+\varepsilon}$, and therefore have $\int_{\rho(x,y)\sim r}|k(x,y)|\,dy\prec r^\varepsilon$ by (8.0). The same goes for $\int_{\rho(x,y)\sim r}|k(x,y)|\,dx$.

It will often be convenient to write the error operators in the form

$$\int \overline{k}(x,u) f(x \times u)\, du.$$

We will systematically let $u = \Theta^{-1}(x,y)$. Note that

$$\int_{\rho(x,y)\sim r} |k(x,y)|\, dy \sim \int_{\|u\|\sim r} |\overline{k}(x,u)|\, du$$

since $k(x,y) = \overline{k}(x,u)$ times a nonvanishing C^∞ Jacobian factor.

For (7.3E),

$$\overline{k}(x,u) = \eta_1(x)\eta_0(x \times u) \int_0^1 [\widehat{\Delta}'(L_t - \ell_t)](-u)\, \frac{dt}{t}.$$

$\widehat{\Delta}'(L_t - \ell_t)$ is supported where $\|u\| \sim t$, and $\|\widehat{\Delta}'(L_t - \ell_t)\|_1 \prec t^\varepsilon$ by Lemma 5.12. Thus $\overline{k}$ satisfies the first inequality in (7.0). For the second return to $k(x,y)$, which equals $\overline{k}(x, \Theta^{-1}(x,y))$ times a C^∞ Jacobian factor which may be disregarded in the present analysis. Write $x = y \times u^{-1}$ to obtain

$$k(x,y) \sim \overline{k}(y \times u^{-1}, u) = \eta_1(y \times u^{-1})\eta_0(y) \int_0^1 (\widehat{\Delta}'(L_t - \ell_t))(u))\, \frac{dt}{t}.$$

Thus the second half of (7.0) follows from the same estimate on $\widehat{\Delta}'(L_t - \ell_t)$.

It will generally happen that the two halves of (7.0) are essentially equivalent, for the same reason as in this case. Then we shall establish only the first half, and leave the second to the reader.

For (7.4E) the kernel is $\overline{k}(x,u) = \eta_1(x)\eta_0(x \times u) \cdot (\widehat{\Delta}' \int_1^\infty \ell_t\, dt/t)(-u)$ and the issue is the third factor. By an easier analogue of Lemma 5.10 it belongs to $L^q(G)$ for some $q > 1$. By (8.0) and Hölder's inequality, (7.0) follows with $\varepsilon = d(1 - q^{-1})$.

The more delicate error terms are those arising from $\mathscr{P}_0 \circ B$. There is no Leibnitz rule for B, so they take a different form than the errors for $\mathscr{P}_0\Delta$. More significantly, the distribution-kernel for B is no longer supported only on the diagonal, and is singular, so the failure of Θ to satisfy an exact associative law will come into play to a greater extent.

(7.8E) is comparatively simple and may be treated by the arguments which follow, so let us skip it and consider (7.10E). There

$$\begin{aligned}\overline{k}(x,u) &= \eta_1(x)\eta_0(x \times u) \int \mathscr{A}_0(s)\{L(-Q_2(u,s)) - L(\bar{s}u^{-1})\}\, ds \\ &= \eta_1(x)\eta_0(x \times u) \int \mathscr{A}_0(s)\{\cdot\}\, ds.\end{aligned}$$

Decompose

$$\{\cdot\} = \int_0^1 \{L_t(-Q_2(u,s)) - L_t(\bar{s}\bar{u}^{-1})\}\, \frac{dt}{t} = \int_0^1 \{\cdot\}_t\, \frac{dt}{t}.$$

To study $\int \mathscr{A}_0(s)\{\cdot\}_t\,ds$ fix a small $\omega_1 > 0$, set $\tau = t^{1-\omega_1}$, fix $\zeta_1 \in C_0^\infty(\mathbf{R})$ identically one in a neighborhood of 0, and decompose

$$\mathscr{A}_0(s) = \mathscr{A}_0(s)\zeta_1(\tau^{-2}s) + \mathscr{A}_0(s)[1 - \zeta_1(\tau^{-2}s)],$$

a local plus a global part.

To analyze $\int \mathscr{A}_0(s)\zeta_1(\tau^{-2}s)\{\cdot\}_t\,ds$ introduce the change of variables $s = \tau^2\sigma$, $u = \delta_\tau w$ and recall that $L_t'(v) = t^{d-2}L_t(\delta_t v) = (\ell_1 * \varphi_{t^\omega})(v)$ to write the integral as

(8.1)

$$\tau^{-2}t^{-d+2}\int \mathscr{A}(\sigma)\varphi(\tau^2\sigma)\zeta_1(\sigma)[L_t'(-\delta_t^{-1}Q_2(u,s)) - L_t'(-\delta_t^{-1}\bar{s}u^{-1})]\,d\sigma$$
$$= \tau^{-2}t^{-d+2}\int \mathscr{A}(\sigma)\varphi(\tau^2\sigma)\zeta_1(\sigma)[L_t'(-\delta_{\tau/t}Q_3(w,\sigma)) - L_t'(-\delta_{\tau/t}\bar{\sigma}w^{-1})]\,d\sigma$$

where $Q_2(u,s) \equiv \delta_\tau Q_3(w,\sigma)$. Recall the notation $\bar{\sigma} = \exp(\sigma[X,Y])0 \in G$. We must apologize to the reader for having used the letter φ to denote both the cutoff function used to truncate $\mathscr{A}$ to $\mathscr{A}_0$ and the auxiliary function used in (5.1) to define L_t in terms of ℓ_t. It will always be clear from the context which is intended.

As a function of u, $\int \mathscr{A}_0(s)\zeta_1(\tau^{-2}s)\{\cdot\}_t\,ds$ is supported where $\|u\| \prec \tau$. For

$$L_t(-Q_2(u,s))\cdot\zeta_1(\tau^{-2}s) = 0$$

unless $|s| \prec \tau^2$, which is equivalent to $\|\bar{s}\| \prec \tau$, and $\|Q_2(u,s)\| \prec t$. Therefore $u = Q_1(v,s)$ for some v,s satisfying $\|v\| \prec t$ and $\|\bar{s}\| \prec \tau$. By Lemma 7.4, $\|u\| \prec \tau$ if the contribution of $L_t(-Q_2(u,s))$ to the integral is nonzero. The same holds for the contribution of $L_t(\bar{s}u^{-1})$, by the triangle inequality on G. We shall repeatedly encounter this sort of argument, based on various analogues of Lemma 7.4; we shall omit the details in the future and simply cite the "triangle inequality" without being more precise.

It follows that the integrand in (8.1) is supported where $\|w\|, |\sigma| \prec 1$. Lemma 7.4 implies that on this region, $Q_3(w,\sigma) \equiv \delta_\tau^{-1}Q_2(\delta_\tau w, \tau^2\sigma) = \bar{\sigma}w^{-1}$ plus an error whose C^n norm in both variables is majorized by $C_n\tau$, for all n. The C^n norm of $z \to L_t'(-\delta_{\tau/t}z)$, on $\{\|z\| \prec 1\}$, is $\prec t^{-C\omega}(\tau/t)^C$, $C = C(n)$. Therefore the C^n norm of $L_t'(-\delta_{\tau/t}Q_3(w,\sigma)) - L_t'(-\delta_{\tau/t}\bar{\sigma}w^{-1})$ as a function of σ for $\|w\| \prec 1$, is $\prec t^{-C\omega}(\tau/t)^C\tau$ for some $C(n) < \infty$ independent of ω, ω_1; the crucial factor of τ^{+1} arises because $\|Q_3(w,\sigma) - \bar{\sigma}w^{-1}\|_{C^n} = O(\tau)$. Now

$$\left|\int \mathscr{A}_0(\sigma)\varphi(\tau^2\sigma)F(\sigma)\,d\sigma\right| \prec \|F\|_{C^2}$$

for any $F \in C^2(\mathbf{R})$ supported where $|\sigma| \prec 1$. Taking

$$F(\sigma) = L_t'(-\delta_{\tau/t}Q_3(w,\sigma)) - L_t'(-\delta_{\tau/t}\bar{\sigma}w^{-1})$$

results in

$$\left|\int \mathscr{A}_0(s)\zeta_1(\tau^{-2}s)\{\cdot\}_t\,ds\right| \prec \tau^{-2}t^{-d+2}t^{-C(\omega+\omega_1)}\tau \prec \tau^{-d+1-C(\omega+\omega_1)} = \tau^{-d+\varepsilon}$$

where $\varepsilon > 0$, so long as both ω, ω_1 are sufficiently small.

Integrating this pointwise bound with respect to dt/t gives

$$\left|\int_0^1 \int \mathscr{A}_0(s)\zeta_1(t^{-2(1-\omega_1)}s)\{\cdot\}_t\, ds\, \frac{dt}{t}\right| \le C \int_{Cr}^1 \tau^{-d+\varepsilon}\, \frac{d\tau}{\tau} \sim r^{-d+\varepsilon}$$

when $\rho(x,y) = \|u\| \sim r$. For since $\int \mathscr{A}_0(s)\zeta_1(\tau^{-2}s)\{\cdot\}_t\, ds$ is supported where $\|u\| \prec \tau$, only those t for which $t^{(1-\omega_1)} = \tau$ is $\succ r$ can make a nonzero contribution. (7.0) follows from (8.0) and the L^∞ bound

$$|k(x,y)| \prec \rho(x,y)^{-d+\varepsilon}.$$

To study $\int \mathscr{A}_0(s)(1-\zeta_1(\tau^{-2}s))\{\cdot\}_t\, ds$ observe that only small values of t need be considered. For $\{\cdot\}_t \in L^\infty$ and $\mathscr{A}_0(\cdot)(1-\zeta_1(\tau^{-2}\cdot)) \in L^1$ uniformly so long as t is bounded away from 0.

Once t is small, the triangle inequality implies $\int \mathscr{A}_0(s)(1-\zeta_1(\tau^{-2}s))\{\cdot\}_t\, ds$, a function of u, is supported where $\|u\| \succ \tau$. For suppose this fails for the contribution of the first part of $\{\cdot\}_t$. Then there exist u, s so that $\|u\| \le C_1\tau$, $\|Q_2(u,s)\| \le C_2 t$, and $\|\bar{s}\| \ge C_3\tau$. Since s must also belong to support$(\mathscr{A}_0) =$ support(φ) we may also assume that $|s| \le C_0$ where C_0 is independent of t but may be taken as small as desired. By Lemma 7.4 $(u\bar{s}^{-1})_\alpha = Q_2(u,s)_\alpha + O(u,s)^{w(\alpha)+1}$ for all components, so

$$\|u\bar{s}^{-1}\| \le C\|Q_2(u,s)\| + C(\|u\|^{1+m^{-1}} + \|\bar{s}\|^{1+m^{-1}}) \le Ct + CC_1\tau + C\|\bar{s}\|^{1+m^{-1}}.$$

But $\|\bar{s}\| \le C(\|u\| + \|u\bar{s}^{-1}\|) \le CC_1\tau + Ct + C\|\bar{s}\|^{1+m^{-1}}$, so that $\|\bar{s}\| \le C(t + C_1\tau)$. If C_1 is sufficiently small this contradicts the condition $\|\bar{s}\| \ge C_3\tau$. Therefore $\|u\| \succ \tau$. We have shown in addition that $\mathscr{A}_0(s)(1-\zeta_1(\tau^{-2}s))\{\cdot\}_t = 0$ unless $|s|$ is comparable to $\|u\|^2$. For the contribution of the second part of $\{\cdot\}_t$, namely $L_t(\bar{s}u^{-1})$, the argument is in the same vein but simpler since the triangle inequality is now needed only on G.

$$\{\cdot\}_t(1-\zeta_1(\tau^{-2}s)) = [L_t(-Q_2(u,s)) - L_t(\bar{s}u^{-1})](1-\zeta_1(\tau^{-2}s))$$

is a difference of two terms; it suffices to estimate them separately without exploiting any cancellation which might occur. $\|\{\cdot\}_t\|_{L^1(du)} \prec t^2$ uniformly in

s, and this holds for each term separately, since $\|L_t\|_1 \prec t^2$. Therefore

$$\int_{\|u\|\sim r} \left| \int\int \mathscr{A}_0(s)(1-\zeta_1(\tau^{-2}s))\{\cdot\}_t \, ds \, \frac{dt}{t} \right| du$$

$$\prec \int_{\|u\|\sim r} \int_{|s|\sim r^2} \int_{\tau\prec r} |s|^{-2} |\{\cdot\}_t| \frac{dt}{t} \, ds \, du$$

$$\prec \int_{|s|\sim r^2} \left(\int_{\tau\prec r} t^2 \frac{dt}{t} \right) |s|^{-2} \, ds$$

$$\prec r^{-2} r^{2/(1-\omega_1)} = r^{\varepsilon}$$

where $\varepsilon = 2/(1-\omega_1) - 2 > 0$. This completes the proof of the first half of (7.0) for (7.10E); the second half is very similar.

(7.11E) and (7.12E) are, disregarding the harmless factor $\eta_1(x)\eta_0(y)$, given by convolution on G. (7.11E) is $\mu * \int_0^1 (L_t - \ell_t)\, dt/t$, while (7.12E) is $\mu * \int_1^\infty \ell_t \, dt/t$. The latter is in $L^q(G)$ by Lemma 5.10, and Hölder's inequality then implies (7.0). For $0 < t \le 1$, $\mu * (L_t - \ell_t)(u) = t^{-d} \mu * \ell_1 * (\delta - \varphi_{t^\omega})(\delta_t^{-1} u)$. By Corollary 5.9 and Lemma 5.12, $\|\mu * \ell_1 * (\delta - \varphi_{t^\omega})\|_{L^q} \prec t^\varepsilon$ for some $\varepsilon > 0$. Therefore if p is chosen to be very close to but slightly larger than one, $\|\mu * (L_t - \ell_t)\|_{L^p} \prec t^{\varepsilon/2}$ uniformly as $t \to 0$. Integrating with respect to dt/t gives $\mu * \int_0^1 (L_t - \ell_t)\, dt/t \in L^p$, and again Hölder's inequality implies (7.0). Proposition 7.1 is completely proved.

9. $\mathscr{E}_0$ IS SMOOTHING

It will be convenient to express the smoothing property of $\mathscr{E}_0$ in terms of operators which are essentially powers of Δ on U. Recall the distributions I_α on G, satisfying $\widehat{\Delta}^{\alpha/2} f = f * I_\alpha$. $\widehat{\Delta}^{\alpha/2} f(0) = \langle f, I_\alpha \rangle$ since I_α is symmetric about 0. Define Λ_α, acting on functions defined on U, by

$$\Lambda_\alpha f(x) = \eta_2(x) \langle g, I_\alpha \rangle \tag{9.1}$$

where $g(u) = \eta_1(x \times u) f(x \times u)$ and $\eta_2 \in C_0^\infty(U)$ is chosen so that $\eta_2 \eta_1 \equiv \eta_1$.

A kernel $k(x, y)$ is said to be associated to an operator T, and vice versa, if $\int |k(x, y)|\, dy$ is finite for almost all x and if $Tf(x) = \int k(x, y) f(y)\, dy$ for $f \in C_0^\infty$.

Lemma 9.1. *For each $\alpha \in (0, d)$ $\Lambda_{-\alpha} \circ \Lambda_\alpha = \eta_1^2 I + e_\alpha$ where e_α is associated to a kernel k which is compactly supported in $U \times U$ and satisfies*

$$|D_x D_y k(x, y)| \prec \rho(x, y)^{-d+1-w}.$$

Here D_x denotes any operator $(X \text{ or } Y) \circ (X \text{ or } Y) \cdots$ acting in the x variable, D_y any operator of the same form acting in the y variable, and w

is the sum of the orders of D_x and D_y. Calculations like those in §8 and in the proof of Lemma 9.4 below, but simpler, establish Lemma 9.1.

Corollary 9.2. *Let $p \in (1, \infty)$. Let T be a linear operator mapping smooth functions to smooth functions and suppose that for some $\alpha > 0$, $\Lambda_\alpha \circ T$ is bounded on L^p. Then T is bounded from L^p to the (ordinary, Euclidean) Sobolev space L^p_ε for some $\varepsilon > 0$.*

By Lemma 9.1 it suffices to show that $\Lambda_{-\alpha}$ and e_α map L^p to L^p_ε. Since we do not care about the optimal value of ε this is straightforward, using the pointwise bounds on their kernels and derivatives thereof.

The main result of this section is

Proposition 9.3. *There exists $\varepsilon > 0$ such that for all $p \in (1, \infty)$, $\mathscr{E}_0$ is bounded from L^p to L^p_ε.*

By Lemma 9.2 this follows from

Lemma 9.4. *There exist $\varepsilon, \varepsilon_1 > 0$ such that for all $\gamma \in [0, \varepsilon_1]$, $\Lambda_\gamma \circ \mathscr{E}_0$ is associated to a kernel which is locally integrable in each variable separately and satisfies*

$$\sup_x \int_{\rho(x,y)\sim r} |k(x,y)|\,dy + \sup_y \int_{\rho(x,y)\sim r} |k(x,y)|\,dx \prec r^\varepsilon. \tag{9.2}$$

The proof is an amplification of that of Proposition 7.1, but there is a delicate point. The composition of Λ_γ with some of the error terms which sum up to $\mathscr{E}_0$ actually fails to satisfy the estimates desired, and fails rather badly. Examples are (7.11E) and (7.12E); it is the fundamental difficulty once again, the low order of differentiability of ℓ away from 0.

Fix γ. To decompose $\Lambda_\gamma \circ \mathscr{E}_0 = \Lambda_\gamma \circ \mathscr{P}_0 A - \Lambda_\gamma$ into simpler terms like the (7.nE), first write $g(u) = f(x \times u)\eta_1(x \times u)$ and

$$\begin{aligned} \Lambda_\gamma f(x) &= \eta_2(x)\langle I_\gamma, g\rangle = \eta_2(x)\langle I_\gamma, \widehat{A}^{-1}\widehat{A}g\rangle \\ &= \eta_2(x) \iint_{G\times G} I_\gamma(z)[(\widehat{\Delta} + \widehat{B})'\ell](wz^{-1}) f(x \times w^{-1})\eta_1(x \times w^{-1})\,dw\,dz \end{aligned} \tag{9.3}$$

where $\widehat{\Delta}'$, $\widehat{B}'$ denote the right-invariant analogues of $\widehat{\Delta}, \widehat{B}$. (9.3) will be compared to

$$\Lambda_\gamma \mathscr{P}_0 A f(x) = \eta_2(x) \int I_\gamma(z)(\mathscr{P}_0 A f)(x \times z^{-1})\eta_1(x \times z^{-1})\,dz.$$

The contributions of B and $\widehat{B}$ are more delicate than those of Δ and $\widehat{\Delta}$, so only the former will be discussed. Simpler versions of the arguments which follow apply to Δ and $\widehat{\Delta}$, as in §8.

By definition

$$\begin{aligned} \mathscr{P}_0 B f(x) &= \eta_1(x) \int L(u)(Bf)(x \times u^{-1})\eta_0(x \times u^{-1})\,du \\ &= \eta_1(x) \iint L(u)\mathscr{A}_0(s) f(\gamma(x \times u^{-1}, -s))\eta_0(x \times u^{-1})\,ds\,du \end{aligned}$$

so that

$$\Lambda_\gamma \mathscr{P}_0 Bf(x) = \eta_2(x) \int I(z)(\mathscr{P}_0 Bf)(x \times z^{-1})\eta_1(x \times z^{-1})\, dz$$

$$= \eta_2(x) \iiint I(z)\eta_1^2(x \times z^{-1})L(u)\mathscr{A}_0(s)f(\gamma(x \times z^{-1} \times u^{-1}, -s)) \cdot \eta_0(x \times z^{-1} \times u^{-1})\, ds\, du\, dz. \tag{9.4}$$

Replace u by a new variable v, defined by

$$x \times v = \gamma(x \times z^{-1} \times u^{-1}, -s) \tag{9.5}$$

to obtain

$$\eta_2(x) \iiint I_\gamma(z)\mathscr{A}_0(s)L(u)\eta_1^2(x \times z^{-1})\eta_0(x \times z^{-1} \times u^{-1})f(x \times v)J\, ds\, dv\, dz$$

where $J(x, v, s, z) \in C^\infty$ is the Jacobian determinant from the substitution (9.5), and where u is now a C^∞ function of x, s, v, z defined implicitly by (9.5). Set

$$K_t(x, y) = \iint I_\gamma(z)\mathscr{A}_0(s)L_t(u)\eta_1^2(x \times z^{-1})\eta_0(x \times z^{-1} \times u^{-1})J\, ds\, dz \tag{9.6}$$

where of course $y = x \times v$. The "kernel" for $\Lambda_\gamma \mathscr{P}_0 B$ is formally $\eta_2(x) \int_0^1 K_t(x, y)\, dt/t$.

Split K_t as $I_t + II_t + III_t + IV_t$ by writing

$$1 \equiv \{\zeta_1(t^{-1}\|z\|) + [1 - \zeta_1(t^{-1}\|z\|)]\} \cdot \{\zeta_1(\tau^{-2}s) + [1 - \zeta_1(\tau^{-2}s)]\}$$

in (9.6), where $\tau = t^{1-\omega_1}$, and expanding the resulting product into four terms. $\zeta_1 \in C_0^\infty(\mathbf{R})$ is even, and $\equiv 1$ on $[-C_0, C_0]$ where C_0 is a large constant. Let

$$\begin{aligned} I_t &\leftrightarrow \zeta_1(t^{-1}\|z\|)[1 - \zeta_1(\tau^{-2}s)] \\ II_t &\leftrightarrow \zeta_1(t^{-1}\|z\|)\zeta_1(\tau^{-2}s) \\ III_t &\leftrightarrow [1 - \zeta_1(t^{-1}\|z\|)]\zeta_1(\tau^{-2}s) \\ IV_t &\leftrightarrow [1 - \zeta_1(t^{-1}\|z\|)] \cdot [1 - \zeta_1(\tau^{-2}s)]. \end{aligned}$$

The four will be analyzed separately. To each corresponds a portion of (9.3). These portions will be denoted $\overline{I}_t$, $\overline{II}_t$ and so on. For instance $\overline{I}_t$ is

$$\int I_\gamma(z)\mathscr{A}_0(s)\ell_t(\bar{s}^{-1}v^{-1}z^{-1})[\zeta_1(t^{-1}\|z\|)(1 - \zeta_1(\tau^{-2}s))]\eta_0(x \times v)\, ds\, dz$$

where still $y = x \times v$; $\overline{I}_t$ is obtained from I_t by replacing L_t by ℓ_t, $\mathscr{A}$ by $\mathscr{A}_0$, u by $\bar{s}^{-1}v^{-1}z^{-1}$, J by 1, and the cutoff function $\eta_1^2(x \times z^{-1})\eta_0(x \times z^{-1} \times u^{-1})$ by $\eta_0(x \times v)$. $\overline{II}_t$ is obtained from II_t by the same modifications and so on. There are two extra terms:

$$\overline{V} = \int I_\gamma(z)\mathscr{A}_0(s)\left[\int_1^\infty \ell_t(\bar{s}^{-1}v^{-1}z^{-1})\frac{dt}{t}\right] ds\, dz$$

and

$$\overline{VI} = \mu^{\infty} * L * I_{\gamma}(v)$$

where μ^{∞} is the convolution "kernel" for $|[\widehat{X}, \widehat{Y}]| - |[\widehat{X}, \widehat{Y}]|_0$ on G. Then

$$\eta_2(x) \cdot \int_0^1 (\overline{I}_t + \overline{II}_t + \overline{III}_t + \overline{IV}_t)\frac{dt}{t} + \eta_2(x)\eta_0(x \times v)[\overline{V} + \overline{VI}]$$

is the kernel for (9.3). The plan is to study I_t, $\overline{I}_t$, $II_t - \overline{II}_t$, $III_t - \overline{III}_t$, IV_t, $\overline{IV}_t$, $\overline{V}$, and $\overline{VI}$ separately. The proof will show that II_t and $\overline{II}_t$ give rise to the main parts of the kernels of $\Lambda_{\gamma}\mathscr{P}_0 B$ and (9.3).

In the analysis we proceed in a formal way, writing integrals where distributions abound, and omitting justification when the order of integration in a multiple integral is changed. Such manipulations could be justified by truncating L as in §7, regularizing $\mathscr{A}_0$ and I_{γ}, and passing to the limit, assuming f to be smooth.

Let the variable s be assigned weight 2, as in §8.

Lemma 9.5. *If $x \times v = \gamma(x \times z^{-1} \times u^{-1}, -s)$ then $v = Q_1(s, u, z)$ and $u = Q_2(s, v, z)$ where Q_1, Q_2 are C^{∞} functions of x, (u or v), z, and s. Their components satisfy*

$$Q_1(s, u, z)_{\alpha} = (z^{-1} u^{-1} \bar{s}^{-1})_{\alpha} + O(s, u, z)^{w(\alpha)+1}$$

and

$$Q_2(s, v, z)_{\alpha} = (\bar{s}^{-1} v^{-1} z^{-1})_{\alpha} + O(s, v, z)^{w(\alpha)+1}.$$

The proof is the same as for Lemma 7.4.

Begin the analysis of I_t. Freeze s and consider the integral with respect to u and z in (9.4), omitting the harmless factor $\eta_2(x)$ as well as $\mathscr{A}_0(s)(1 - \zeta_1(\tau^{-2}s))$:

$$\begin{aligned}\iint & I_{\gamma}(z)\zeta_1(t^{-1}\|x\|)L_t(u) \\ & \cdot \eta_1^2(x \times z^{-1})\eta_0(x \times z^{-1} \times u^{-1}) f(\gamma(x \times z^{-1} \times u^{-1}, -s))\, dz\, du \\ = & \iint I_{\gamma}(z)\zeta_1(t^{-1}\|z\|)a(x, w, z)L_t(Q_4(w, z))dz\, f(\gamma(x \times w, -s))\, dw\end{aligned}$$

where $a \in C^{\infty}$ incorporates a Jacobian factor, and we have substituted $u \to w$, where $x \times w = x \times z^{-1} \times u^{-1}$, $w = Q_3(u, z)$, $u = Q_4(w, z)$ and as always both functions Q depend implicitly on x. By Baker-Campbell-Hausdorff, $Q_4(w, z) = w^{-1}z^{-1} + O(w, z)^{m+1}$. The inner integral

$$\int I_{\gamma}(z)\zeta_1(t^{-1}\|z\|)L_t(Q_4(w, z))a(x, w, z)\, dz \tag{9.7}$$

is completely independent of s; what we have done is to associate $\Lambda_{\gamma}\mathscr{P}_0 B = (\Lambda_{\gamma}\mathscr{P}_0)B$. By the reasoning of §8, (9.7) equals

$$I_t''(w) = \int I_{\gamma}(z)\zeta_1(t^{-1}\|z\|)L_t(w^{-1}z^{-1})a(x, w, 0)\, dz$$

plus an error term $I_t'(w)$, which depends also on x, whose L^∞ norm is majorized by $Ct^{-d-\gamma+2}t^{1-c\omega} = Ct^{-d-\gamma+2+\varepsilon}$ where $\varepsilon > 0$. By the triangle inequality both I_t' and (9.7) itself are supported where $\|w\| \prec t$, so $\|I_t'\|_{L^1(dw)} \prec t^{-\gamma+2+\varepsilon}$.

The contribution of I_t' to K_t is the kernel for the operator

$$\begin{aligned} f \to \eta_2(x) \int I_t'(w)\mathscr{A}_0(s)(1-\zeta_1(\tau^{-2}s))f(\gamma(x\times w, -s))\,ds \\ = \eta_2(x)\int I_t'(w(v,s))\mathscr{A}_0(s)(1-\zeta_1(\tau^{-2}s))f(x\times v)\,ds \end{aligned}$$

where $y = x\times v = \gamma(x\times w, -s)$. As in Lemma 7.4, $v_\alpha = (w\bar{s}^{-1})_\alpha + O(w,\bar{s})^{w(\alpha)+1}$ for each coordinate. Thus the contribution of I_t' to K_t is

$$\eta_2(x)\int I_t'(w(v,s))\mathscr{A}_0(s)(1-\zeta_1(\tau^{-2}s))\,ds. \tag{9.8}$$

Suppose $\rho(x,y) = \|v\|$ is comparable to r. By the triangle inequality, since $1-\zeta_1(\tau^{-2}s)$ is supported where $\|\bar{s}\| \succ \tau >> t$, and $I_t'(w)$ where $\|w\| \prec t$, (9.8) is supported where $\|v\| \succ \tau$. Moreover if $r \succ \tau$ then the integrand in (9.8) can be nonzero only when $\|\bar{s}\| \sim r$, that is when $|s| \sim r^2$. Therefore

$$\begin{aligned} \int_{\|v\|\sim r} |(9.8)|\,dv &\prec \int_{\|v\|\sim r}\int_{|s|\sim r^2} |s|^{-2}|I_t'(w(v,s))|\,ds\,dv \\ &\prec \int_{|s|\sim r^2} r^{-4}\,ds \cdot \int_{\|v\|\sim r} |I_t'(w)|\,dw \\ &\prec r^{-2}\|I_t'\|_1 \\ &\prec r^{-2}t^{-\gamma+2+\varepsilon}. \end{aligned}$$

Integrate over the relevant range $\tau = t^{1-\omega_1} \prec r$ to obtain the following bound for the contribution of $\int_0^1 I_t'\frac{dt}{t}$ to $\sup_x \int_{\rho(x,y)\sim r}\int_0^1 |I_t(x,y)|\frac{dt}{t}dy$:

$$\begin{aligned} Cr^{-2}\int_0^{r^{1/(1-\omega_1)}} t^{-\gamma+2+\varepsilon}\frac{dt}{t} &= Cr^{-2}r^{(-\gamma+2+\varepsilon)/(1-\omega_1)} \\ &\le Cr^{-2}r^{-\gamma+2+\varepsilon} \\ &\prec r^{-\gamma+\varepsilon} \end{aligned}$$

as desired.

(9.7) was split into two parts—turn now to the principal term,

$$I_t''(w) = \int I_\gamma(z)\zeta_1(t^{-1}\|z\|)L_t(w^{-1}z^{-1})\,dz\cdot a(x,w,0).$$

The integral is merely a convolution on G, and after rescaling as in §8 (let $z' = \delta_t z$, $w' = \delta_t w$), Lemma 5.11 applies and demonstrates that $\|I_t''\|_{L^1} \prec t^{-\gamma+2}$. Continue as for I_t', arriving finally at the bound $r^{-2}r^{(-\gamma+2)/(1-\omega_1)} = r^{-\gamma+\varepsilon}$,

where $\varepsilon > 0$ since $\omega_1 > 0$ and $\gamma < 2$. We have shown that

$$\sup_x \int_{\rho(x,y)\sim r} |k(x,y)|\,dy \prec r^{-\gamma+\varepsilon}$$

for the contribution of $\int_0^1 I_t \frac{dt}{t}$. It may easily be deduced from the particular forms of the operators and the estimates obtained that the same goes for $\sup_y \int_{\rho(x,y)\sim r} |k(x,y)|\,dx$.

$\overline{I}_t$ may be handled in the same way, and the argument simplifies since $\overline{I}_t$ merely involves convolutions on G. Proposition 5.6′ must be invoked to control the smoothness of ℓ_t. We omit the details.

Consider the main term, $II_t - \overline{II}_t$. II_t has the form

$$\iint I_\gamma(z)\zeta_1(t^{-1}\|z\|)\mathscr{A}_0(s)\zeta_1(\tau^{-2}s)L_t(Q_2(s,v,z))a(x,s,v,z)\,ds\,dt \tag{9.9}$$

where $a(x,0,v,0) \equiv \eta_0(x\times v)$. Compare this first to

$$\iint I_\gamma(z)\zeta_1(t^{-1}\|z\|)\mathscr{A}_0(s)\zeta_1(\tau^{-2}s)L_t(\bar{s}^{-1}v^{-1}z^{-1})a(x,0,v,0)\,ds\,dz. \tag{9.10}$$

Make the changes of variable $s = \tau^2 s'$, $v = \delta_\tau v'$, $z = \delta_\tau z'$ and argue as in §8 to obtain a pointwise bound $C\tau^{-d-\gamma}\tau^{1-c\omega-c\omega_1}$ for the difference of (9.9) and (9.10). This is $\prec \tau^{-d-\gamma+1/2}$ if ω, ω_1 are both chosen sufficiently small. It is essential here that $\gamma < 2$, because of the limited order of L^q differentiability of ℓ_t guaranteed by Proposition 5.6′. The triangle inequality implies that (9.9)–(9.10) is supported where $\|v\| \prec \tau$. Therefore for $\|v\| \sim r$,

$$\int_0^1 |(9.9) - (9.10)|\frac{dt}{t} \prec \int_{r\prec\tau\prec 1} \tau^{-d-\gamma+1/2}\frac{d\tau}{\tau} \sim r^{-d-\gamma+1/2},$$

which implies the desired L^1 estimate. $II_t - \overline{II}_t$ has been reduced to $(9.10) - \overline{II}_t$, which equals

$$\iint I_\gamma(z)\zeta_1(t^{-1}\|z\|)\mathscr{A}_0(s)\zeta_1(\tau^{-2}s)[(L_t - \ell_t)(\bar{s}^{-1}v^{-1}z^{-1})]\,ds\,dz\cdot\eta_0(x\times v).$$

As we have already seen, it is only necessary to examine small values of t. Then $\varphi(s)\zeta_1(\tau^{-2}s) \equiv \zeta_1(\tau^{-2}s)$, so $\mathscr{A}_0$ may be replaced by $\mathscr{A}$. The change of variables $z = \delta_\tau z'$, $v = \delta_\tau v'$, $s = \tau^2 s'$ re-expresses the last integral as $\eta_0(x\times v)$ times $\tau^{2-\gamma}t^{-d+2}$ times

$$\iint I_\gamma(z')\zeta_1((\tau/t)\|z'\|)\mathscr{A}(s')\zeta_1(s')[(\ell_1 * (\varphi_{t^\omega} - \delta))(\delta_{\tau/t}\bar{s}'^{-1}v'^{-1}z'^{-1})]\,ds'\,dz', \tag{9.11}$$

which again involves only convolutions on G. For the slightly simpler expression

$$\iint I_\gamma(z')\zeta_1(\|z'\|)\mathscr{A}(s')\zeta_1(s')[(\ell_1 * (\varphi_{t^\omega} - \delta))(\bar{s}'^{-1}v'^{-1}z'^{-1})]\,ds'\,dz'\cdot\eta_0(x\times v),$$

Proposition 5.6′ plus Corollary 5.9 demonstrate that it is in L^q with respect to v', so long as $\gamma < 2$, with L^q norm $\prec t^{\varepsilon_0 \omega}$ for some $\varepsilon_0 > 0$. Clearly (9.11) will satisfy the same bound, worsened by a factor $(\tau/t)^{C_0} = t^{-C_0\omega_1}$ for some $C_0 < \infty$. Therefore

$$\|\tau^{2-\gamma} t^{-d+2} \cdot (9.11)\|_{L^q(dv')} \prec \tau^{2-\gamma} t^{-d+2} t^{\varepsilon_0 \omega} t^{-C_0\omega_1} = \tau^{-d-\gamma} \tau^{-C_1\omega_1} \tau^{\varepsilon_1 \omega},$$

where still $C_1 < \infty$, $q > 1$, and $\varepsilon_1 > 0$. So far the only restriction on ω and ω_1 has been that both should be sufficiently small. We now require also that

$$\omega_1 < (\varepsilon_1/2C_1)\omega.$$

Then $\|\tau^{2-\gamma} t^{-d+2} \cdot (9.11)\|_{L^q(dv')} \prec \tau^{-d-\gamma+\varepsilon}$ where $\varepsilon > 0$. Rescaling gives

$$\|(9.10) - \overline{II}_t\|_{L^{q'}(dv)} \prec \tau^{-\gamma}$$

for some $q' > 1$, uniformly in t. By the triangle inequality, $(9.10) - \overline{II}_t$ is supported where $\|v\| \prec \tau$, so for the integral with respect to dt/t we again find the bound $r^{-\gamma}$ for the $L^{q'}(dv)$ norm on the region $\{\|v\| \sim r\}$. Then Hölder's inequality gives (9.10).

IV_t and $\overline{IV}_t$ are easier, for the cutoff functions $(1 - \zeta_1(t^{-1}\|z\|))$ and $(1 - \zeta_1(\tau^{-2}s))$ remove the singularities of I and $\mathscr{A}_0$. The two terms are quite similar, so we discuss only the more complicated one, IV_t. In the integral (9.6) defining K_t, freeze s and perform the z-integration first. As in the treatment of I_t above, obtain

$$IV_t'(w) = \int I_\gamma(z)(1 - \zeta_1(t^{-1}\|z\|))L_t(Q_4(w, z))a(x, w, z)dz,$$

analogous to (9.7) rather than to $I_t'(w)$. Since

$$\|L_t\|_{L^1} \prec t^2 \quad \text{and} \quad |I_\gamma(z) \cdot (1 - \zeta_1(t^{-1}\|z\|))| \prec \|z\|^{-d-\gamma}\chi_{\|z\| \geq t},$$

the triangle inequality gives

$$|IV_t'(w)| \prec t^{-d-\gamma+2}(1 + t^{-1}\|w\|)^{-d-\gamma}. \tag{9.12}$$

Consequently

$$\|IV_t'\|_{L^1} \prec t^{2-\gamma}.$$

Now writing k_t for the contribution of IV_t' to K_t,

$$\begin{aligned} |k_t(x, y)| &= \left|\eta_2(x) \int IV_t'(w(s, v))\mathscr{A}_0(s)(1 - \zeta_1(\tau^{-2}s))\, ds\right| \\ &\prec \int_{|s| \succ \tau^2} |IV_t'(w(s, v))|\, |s|^{-2}\, ds \end{aligned} \tag{9.13}$$

where $y = x \times v = \gamma(x \times w, -s)$. Consider the behavior on $\{v : \|v\| \sim r\}$. First look at the case when $r \leq C_0\tau$ where C_0 is small. By the triangle inequality

$\|w(s,v)\| \sim \|\bar{s}\|$ when $|s| \succ \tau^2$. Hence by (9.12)

$$\begin{aligned} |k_t(x,y)| &\prec \int_{|s|\succ\tau^2} t^2 |s|^{-(d+\gamma)/2} |s|^{-2}\, ds \\ &\prec t^2 \tau^{-d-\gamma-2} \\ &= \tau^{-d-\gamma} t^{2w_1} \\ &= \tau^{-d-\gamma+\varepsilon} \end{aligned}$$

where $\varepsilon > 0$. This suffices for our purpose, since $\int_{r\prec\tau\prec 1} \tau^{-d-\gamma+\varepsilon}\, d\tau/\tau \sim r^{-d-\gamma+\varepsilon}$. When $r \succ \tau$ three cases must be distinguished in the integral: (i) $\|\bar{s}\| \le C_1^{-1} r$, (ii) $\|\bar{s}\| \ge C_1 r$, and (iii) $\|\bar{s}\| \sim r$, where C_1 is a large constant. In case (i) $\|w(s,v)\| \sim r$ by the triangle inequality. By (9.12), the contribution of this region of s to the integral (9.13) is

$$\begin{aligned} &\prec \int_{\tau^2\prec|s|\prec r^2} t^2 r^{-d-\gamma} |s|^{-2}\, ds \\ &\prec (t/\tau)^2 r^{-d-\gamma} \\ &= t^{2\omega_1} r^{-d-\gamma} \le t^{\omega_1} r^{-d-\gamma+\omega_1} \end{aligned}$$

which again suffices since $\int_0^1 t^{\omega_1} dt/t < \infty$. In region (ii) $\|w(s,v)\| \sim \|\bar{s}\|$ so an upper bound is

$$\begin{aligned} \int_{|s|\succ r^2} t^2 \|\bar{s}\|^{-d-\gamma} |s|^{-2}\, ds &\prec t^2 r^{-d-\gamma-2} \le r^{-d-\gamma} t^2 \tau^{-2} \\ &= r^{-d-\gamma} t^{2\omega_1} \le r^{-d-\gamma+\omega_1} t^{\omega_1} \end{aligned}$$

once again. For region (iii) we do not make a pointwise estimate, but instead observe that the contribution to $\int_{\|v\|\sim r} |k_t(x,y)|\, dy$ is majorized by a constant times

$$\int_{\|v\|\sim r} \int_{|s|\sim r^2} |IV_t'(w(s,v))|\, |s|^{-2}\, ds\, dv.$$

Reverse the order of integration. Since $v \to w(s,v)$ is a diffeomorphism, the inner integral is majorized by $C\|IV_t'\|_1 \prec t^{2-\gamma}$, so in all the integral is

$$\begin{aligned} &\prec t^{2-\gamma} \int_{|s|\sim r^2} |s|^{-2}\, ds \\ &\sim r^{-2} t^{2-\gamma} = r^{-\gamma} (t/r)^{2-\gamma} \\ &\prec r^{-\gamma} (t/\tau)^{2-\gamma} = r^{-\gamma} t^{(2-\gamma)\omega_1} \\ &= r^{-\gamma} t^{2\varepsilon} \prec r^{-\gamma+\varepsilon} t^{\varepsilon} \end{aligned}$$

where $2\varepsilon = (2-\gamma)\omega_1 > 0$. Integration with respect to dt/t yields the desired bound $r^{-\gamma+\varepsilon}$.

Next is $III_t - \overline{III}_t$, which may be treated by small modifications of the arguments already given. Let ω_2 be another small constant and write

$$1 - \zeta_1(t^{-1}\|z\|) = (1 - \zeta_1(t^{-1+\omega_2}\|z\|)) + (\zeta_1(t^{-1+\omega_2}\|z\|) - \zeta_1(t^{-1}\|z\|)),$$

thereby decomposing each of III_t and $\overline{III}_t$ into two terms. The difference of the two terms spawned by $(\zeta_1(t^{-(1-\omega_2)}\|z\|) - \zeta_1(t^{-1}\|z\|))$ may clearly be majorized as we did $II_t - \overline{II}_t$, provided that both ω_1 and ω_2 are sufficiently small, relative to ω. The two terms derived from $(1 - \zeta_1(t^{-(1-\omega_2)}\|z\|))$ may be estimated separately; the analysis for I_t applies, with straightforward modifications due to the reversal of the roles of I_γ and B, provided that ω_1 and ω_2 are chosen so that $\omega_1 << \omega_2 << \omega$.

By Lemma 5.10 $\overline{V}$ is majorized by an L^q function of v independent of x, for some $q > 1$, so Hölder's inequality yields the desired bound (9.2). The same goes for $\overline{VI}$. This completes the proof of Lemma 9.4.

10. Smoothing properties of the parametrix

We continue to work in the lifted setting of §4 and continue to denote by $A = \Delta + B$ the operator $\tilde{A}$ of that section.

Proposition 10.1. *Each of*

$$(X \text{ or } Y)(X \text{ or } Y)\mathscr{P}_0, \qquad (X \text{ or } Y)\mathscr{P}_0(X \text{ or } Y), \qquad \mathscr{P}_0(X \text{ or } Y)(X \text{ or } Y)$$

extends to a bounded operator on $L^p(U)$ *for all* $p \in (1, \infty)$.

Proof. We must show that these are in a suitable sense Calderón-Zygmund singular integral operators. Consider $X\mathscr{P}_0X$; the others may be treated in the same way. For $f \in C^\infty$,

$$\mathscr{P}_0Xf(x) = \eta_1(x)\int\left(\int_0^1 L_t(u)\frac{dt}{t}\right)(Xf)(x \times u^{-1})\eta_0(x \times u^{-1})du$$

$$= \lim_{s\to 0}\eta_1(x)\int\left(\int_s^1 L_t(u)\frac{dt}{t}\right)(Xf)(x \times u^{-1})\eta_0(x \times u^{-1})du$$

in C^∞. It suffices to show that after application of X on the left, the resulting operator remains bounded on L^p, uniformly as $s \to 0$. So let $s > 0$ be fixed.

Various error terms may be peeled off. First the operator

$$(X\eta_1)(x)\iint_s^1 L_t(u)\frac{dt}{t}(Xf \cdot \eta_0)(x \times u^{-1})du$$

has a kernel uniformly integrable in each variable separately, uniformly in s, so is bounded. On the other hand for the main term, where X falls on the integral rather than on η_1, one may proceed as in Lemma 7.2 and §8 to write it as

$$\eta_1(x)\int\left(\int_s^1 h_t(u)\frac{dt}{t}\right)(f\eta_0)(x \times u^{-1})du = \int_s^1 T_tf(x)\frac{dt}{t},$$

where h_t is the convolution kernel for the invariant operator $g \to X(L_t * (Xg))$ on G, plus various error terms whose kernels are uniformly integrable in each variable separately.

To deduce that $\int_s^1 T_t\,dt/t$ is bounded on L^2, uniformly in s, apply the Cotlar-Knapp-Stein [KS] almost-orthogonality lemma. Writing $\|\cdot\|$ for the L^2 operator norm, what must be shown is that for $0 < t, \tau \le 1$, $\|T_t^* T_\tau\| + \|T_t T_\tau^*\| \prec \operatorname{Min}(t/\tau, \tau/t)^\varepsilon$ for some $\varepsilon > 0$. The following properties of h_t will be used: $h_t(u) = t^{-d} h^t(\delta_t^{-1} u)$ where h^t is supported on $\{\|\cdot\| \prec 1\}$, $\|h^t\|_{L^q_{\varepsilon_0}} \prec 1$, $\|h^t\|_{C^1} \prec t^{-c\omega}$, and $\int_G h^t(z)dz = 0$. These hold uniformly for $t \in (0,1]$; $\omega > 0$ is the small exponent in the definition (6.3) of $\mathscr{P}_0$. Write $\tilde{h}_t(z) = \overline{h}_t(z^{-1})$.

By the usual analysis the kernel $k(x,y)$ for $T_t T_\tau^*$ may be decomposed as a principal part plus an error. Both are supported where $\rho(x,y) \prec \max(t,\tau)$, by the triangle inequality. Suppose for definiteness that $t \le \tau$. The error kernel is bounded pointwise by $C\tau^{-d}(t/\tau)^{1-c\omega}$ since $\int h^t = 0$ and $\|\tilde{h}^\tau\|_{C^1} \prec \tau^{-c\omega}$. The principal part is of the form $a(x,y)\cdot[\tilde{h}_\tau * h_t(u)]$, where $y = x \times u^{-1}$ and $a \in C_0^\infty(U \times U)$. Since $h^t \in L^1$, $\tilde{h}^\tau \in L^q_{\varepsilon_2}$ for some $\varepsilon_2 > 0$ and $\int h^t = 0$, it may be deduced that $\|\tilde{h}_\tau * h_t\|_{L^1} \prec (t/\tau)^\varepsilon$ for some $\varepsilon > 0$. Therefore $\|T_t T_\tau^*\| \prec (t/\tau)^\varepsilon$ for $t \le \tau$. The case $t > \tau$, as well as $T_t^* T_\tau$, may be treated in the same way. Therefore T is bounded on L^2, uniformly for $s \in (0,1]$.

The L^p boundedness is obtained from the Calderón-Zygmund theory; it suffices to show that both T and its transpose are of weak type (1,1) with bounds independent of s. Note that U is a space of homogeneous type in the sense of Coifman and Weiss [CW], for if $B(x,r)$ is defined to be $\{y : \rho(x,y) \le r\}$ then $|B(x,r)| \sim r^d$ for $r \in (0,1]$ and $x \in U$, and ρ satisfies a quasi-triangle inequality. Therefore T is of weak type $(1,1)$ (uniformly in s) if it can be shown that for any $x_0 \in U$, any $r \in (0,1]$ and any $b \in L^1(U)$ supported in $B(x_0,r)$ and satisfying $\int b = 0$,

$$\|Tb\|_{L^1(U \setminus B(x_0,2r))} \prec \|b\|_{L^1}.$$

We will show that there exists $\varepsilon > 0$ such that $\|T_t b\|_1 \prec (r/t)^\varepsilon \|b\|_1$ for all r, t. This suffices, since on $U \setminus B(x_0, 2r)$, $Tb \equiv \int_{c_0 r}^1 T_t b\,dt/t$ for a small constant c_0; $T_t b$ is supported in $B(x_0, 2r)$ for $t \le c_0 r$. Fix x_0, r and b. For $t \ge c_0 r$ split $h^t = h_1^t + h_2^t$ where $\|h_2^t\|_{C^1} \prec (r/t)^{-\eta}$ and $\eta > 0$ is a small constant to be chosen below, and $\|h_1^t\|_{L^1} \prec (r/t)^\varepsilon$, where $\varepsilon = \varepsilon(\eta)$ is some positive constant, dependent on η. This is possible since $h^t \in L^q_{\varepsilon_0}, \varepsilon_0 > 0$ and it may be done so that still $\|h_2^t\|_{C^1} \prec t^{-c\omega}$. This splits each T_t as $T_t^1 + T_t^2$. The kernel for T_t^1 is uniformly integrable in each variable separately, with bound $\prec (r/t)^\varepsilon$, so $\|T_t^1 b\|_1 \prec (r/t)^\varepsilon \|b\|_1$. By the triangle inequality, $T_t b$ is supported where $\rho(x, x_0) \le t$, and since b has mean value zero it follows, once more by our

usual arguments, that

$$\begin{aligned}\|T_t b\|_1 &\prec \|h_1^t\|_1 \|b\|_1 + (r/t)\|h_2^t\|_{C^1} \|b\|_1 \\ &\prec (r/t)^{\varepsilon} \|b\|_1 + (r/t)^{1-c\omega-c\eta} \|b\|_1 \\ &\prec (r/t)^{\varepsilon} \|b\|_1\end{aligned}$$

if ω and η are small enough. It follows by the same reasoning that the transpose of T is also of weak type $(1,1)$, uniformly in s.

Lemma 10.2. *$A \circ \mathscr{P}_0$ and $\mathscr{P}_0 \circ A$ are bounded on L^p for all $p \in (1,\infty)$.*

Proof. $\Delta \circ \mathscr{P}_0$ and $\mathscr{P}_0 \circ \Delta$ have already been treated. The two operators $B\mathscr{P}_0$ and $\mathscr{P}_0 B$ are quite similar to one another and may be treated in the same way, so we examine only the latter. Let $\mathscr{K} = \mu * \ell$ on G, where $|[\hat{X}, \hat{Y}]|$ is given by convolution with μ. Then $\mathscr{K}$ is a distribution homogeneous of the critical degree $-d$, and it belongs to some L^q_α away from 0, with $\alpha > 0$, by Proposition 5.6$'$. Therefore the proof of Proposition 10.1 applies equally to the operator

$$f \mapsto \eta_1(x) \int \mathscr{K}(u) f(x \times u^{-1}) \eta_0(x \times u^{-1})\, du.$$

But in the course of the proof of Proposition 7.1 we showed that $\mathscr{P}_0 \circ B$ differs from this by an operator given by integration against a distribution-kernel which is uniformly integrable in each variable separately.

Recall from §6 the definitions $\mathscr{P} = (I + \mathscr{E}_0 + \mathscr{E}_0^2 + \cdots + \mathscr{E}_0^M)\mathscr{P}_0$ and $\mathscr{E} = -\mathscr{E}_0^{M+1}$, so that $\mathscr{P}A = I + \mathscr{E}$, where M is a large number to be chosen below.

Corollary 10.3. *The conclusion of Proposition* 10.1 *remains valid if $\mathscr{P}_0$ is replaced by $\mathscr{P}$.*

Proof. By expanding $\mathscr{E} = -(I - \mathscr{P}_0 \circ A)^{M+1}$ as a sum of coefficients times powers of $\mathscr{P}_0 \circ A$, we reduce matters to analyzing $(\mathscr{P}_0 \circ A)^i \circ \mathscr{P}_0$ for $i > 0$. By the associative law it follows at once from the last lemma and proposition that $(X \text{ or } Y)(X \text{ or } Y)(\mathscr{P}_0 A)^i \mathscr{P}_0$ is bounded, and the same if we compose instead with $(X \text{ or } Y)(X \text{ or } Y)$ on the right. Then for the case when one factor of $(X \text{ or } Y)$ falls on the left and one on the right it suffices to apply the next lemma.

Lemma 10.4. *Let T be a linear operator mapping smooth functions to smooth functions. Let $p \in (1,\infty)$. Suppose that all nine of the operators T, $(X \text{ or } Y)(X \text{ or } Y)T$, $T(X \text{ or } Y)(X \text{ or } Y)$ are bounded in the L^p norm. Then the operators $(X \text{ or } Y)T(X \text{ or } Y)$ are also bounded in L^p norm.*

This may be proved by a complex interpolation using the analytic family of operators Λ_z. See [Fo] for a closely related result. We omit the routine details.

Lemma 10.5. *If M is sufficiently large then $(X \text{ or } Y)(X \text{ or } Y)\mathscr{E}$ and $(X \text{ or } Y)\mathscr{E}$ are bounded on L^p for all $p \in (1, \infty)$.*

Proof. We claim that there exists $c < \infty$ such that for all $0 \le s \in \mathbf{R}$, $\mathscr{E}_0$ maps L^p_s to L^p_{s-c} boundedly. The definition of $\mathscr{P}_0$ and the fact that $L \in L^1$, coupled with the diffeomorphism-invariance of L^p_s, imply that $\mathscr{P}_0$ is bounded on all L^p_s (actually it is smoothing). Similarly A maps L^p_s to L^p_{s-c}.

By Proposition 7.1, $\mathscr{E}_0$ maps L^p to L^p_ε for some $\varepsilon > 0$. Interpolating with the boundedness of $\mathscr{E}_0$ from L^p_s to L^p_{s-c} and letting s tend to $+\infty$ we find that $\mathscr{E}_0$ is bounded from L^p_α to $L^p_{\alpha+\delta}$ for all $\alpha \ge 0$ and $\delta < \varepsilon$. The lemma follows.

This completes the proof of Proposition 4.4, and therefore of Theorems A and A$'$ as well. There remains only the application to the Szegö kernel.

11. Balls and the metric. Local L^2 estimates,

Nagel, Stein, and Wainger [NSW] have shown how the manifold M may be endowed with the structure of a space of homogeneous type, in a manner which is natural with respect to $\overline{\partial}_b$. In a small coordinate patch U of the type in which we have been working, for each commutator $W = [(X \text{ or } Y), [(X \text{ or } Y), \dots] \cdots]$ with some number k of factors, consider the function

$$\lambda_W(x) = |\det\{X(x), Y(x), W(x)\}|,$$

the absolute value of the determinant of three vectors in $\mathbf{R}^3$. Set

$$\lambda(x, r) = \sum_{W : k \le m} r^{k+2} \lambda_W(x),$$

a polynomial in r. Note that $k \ge 2$, so the lowest power of r which might possibly occur is r^4. Therefore we have for all x and all $0 < r \le s \prec 1$,

$$\lambda(x, s) \succ (s/r)^4 \lambda(x, r)$$

uniformly in x, r, s. Moreover $\lambda(x, r) \succ r^{m+2}$ by the finite type hypothesis.

For each x and each $0 < r \prec 1$ choose V, a commutator of X, Y with some number $i = i(x, r)$ of factors of $(X \text{ or } Y)$, so that

$$r^{i+2} \lambda_V(x) \ge \tfrac{1}{2} \max_{W : k \le m} r^{k+2} \lambda_W(x). \tag{11.1}$$

Define "balls"

$$B(x, r) = \{\exp(c_0 r u_1 X + c_0 r u_2 Y + c_0 r^i u_3 V)x : u \in \mathbf{R}^3 \text{ and } |u| \le 1\}$$

where $c_0 > 0$ is a small fixed constant. Set also

$$\rho(x, y) = \inf\{r : y \in B(x, r)\}.$$

Let $\widehat{B} = \{|u| \leq 1\} \subset \mathbf{R}^3$ and for each x_0, r define a coordinate map $\phi_{x_0,r} = \phi : \widehat{B} \leftrightarrow B(x_0, r)$ by

$$\phi(u) = \exp(c_0 r u_1 X + c_0 r u_2 Y + c_0 r^i u_3 V)x.$$

[NSW] show that if c_0 is chosen sufficiently small, depending on m and on U, then for all sufficiently small $r > 0$ and all x, y, z within Euclidean distance c_0 of $0 \in U$,

Theorem 11.1 [NSW].

(1) $|B(x, r)| \sim \lambda(x, r)$.
(2) $\rho(x, y) \prec \rho(x, z) + \rho(z, y)$.
(3) $\rho(x, y) \sim \rho(y, x)$.
(4) $\phi_{x,r} : \widehat{B} \leftrightarrow B(x, r)$ *is a bijective diffeomorphism.*
(5) $|det(\frac{\partial \phi_{x,r}}{\partial u})| \sim \lambda(x, r)$ *on* $\widehat{B}$, *uniformly in* x, r.
(6) $|\frac{\partial^\alpha}{\partial u^\alpha} det(\frac{\partial \phi_{x,r}}{\partial u})| \prec \lambda(x, r)$ *on* $\widehat{B}$, *for all* α, *uniformly in* x, r.

Moreover if for each x, r *another* V' *satisfying* (11.1) *is chosen in place of* V, *then the resulting balls* B' *and metric* ρ' *are equivalent in the sense that* $\rho \sim \rho'$.

The diffeomorphism $\phi_{x_0,r}$ may be used to pull back the vector fields X, Y from $B(x_0, r)$ to vector fields $\widehat{X}, \widehat{Y}$ on $\widehat{B}$; of course these depend on x_0, r. V pulls back to $\widehat{V}$, an appropriate commutator of $\widehat{X}, \widehat{Y}$. From the results of [NSW] there follows also

Lemma 11.2. $\widehat{X}, \widehat{Y}, \widehat{V}$ *are linearly independent on* $\widehat{B}$, *and their coefficients are* C^∞ *functions on* $\widehat{B}$. *Both conclusions hold uniformly in* $x_0 \in U$ *and* $r > 0$.

We always assume implicitly that r is rather small. By the coefficients of $\widehat{X}$, for instance, we mean the functions a_j satisfying $\widehat{X}(\xi) = \sum_{j=1}^3 a_j(\xi)\frac{\partial}{\partial \xi_j}$ on $\widehat{B}$.

For any (almost) function g on $B(x_0, r)$, define

$$\hat{g}(\xi) = g(\phi(\xi))$$

on $\widehat{B}$. Then the equation $\overline{L}u = f$ on $B(x_0, r)$ pulls back to $\widehat{\overline{L}}\hat{u} = r\hat{f}$ on $\widehat{B}$, where of course $\widehat{\overline{L}} = \widehat{X} + i\widehat{Y}$. The equation $\overline{L}^* v = (X - iY)v + cv = g$ pulls back to $\widehat{\overline{L}}^* \hat{v} = r\hat{g}$, where $\widehat{\overline{L}}^* = \widehat{X} - i\widehat{Y} + \hat{c}$ but where we are forced to contravene convention by setting $\hat{c}(\xi) = rc(\phi(\xi))$ so that the equation will hold. Certainly the functions $\hat{c}$ are C^∞ on $\widehat{B}$, uniformly in x_0, r.

Our present goal is to prove Theorem C, enunciated in §12, which says that the distribution-kernel $K(x, y)$ for the Szegö projection S satisfies certain natural pointwise bounds expressed in terms of the quantities ρ and λ. The idea of the proof is to estimate $K(x_0, y)$, when $\frac{r}{2} \leq \rho(x_0, y) \leq r$, by pulling matters back to $\widehat{B}$ via $\phi_{x_0,r}$, and re-running the reasoning behind Theorems A$'$ and B on $\widehat{B}$, thus obtaining bounds which are suitably scale- and basepoint-invariant so that the desired bounds on K will follow. The principal obstacle to

this line of attack is that the fundamental estimate, $\|u\|_2 \prec \|f\|_2$ when $\overline{\partial}_b u = f$ and $u \perp \mathscr{H}_b$, does not scale correctly. The estimate which should be used as the starting point for the analysis on $\widehat{B}$ is $\|r^{-1}\hat{u}\|_{L^2(\widehat{B})} \prec \lambda(x_0, r)^{-1/2}\|f\|_2$, in other words

$$\|\hat{u}\|_{L^2(\widehat{B})} \prec r\|f\|_2 .$$

From the closed range inequality it follows only that this holds without the crucial factor of r on the right-hand side. The purpose of this section is to obtain the extra factor of r.

Lemma 11.3. *Let $x_0 \in U$, let $r > 0$ be small and define*

$$K(x, y) = [\rho(x, y)/\lambda(x, \rho(x, y))].$$

Then for $g \in L^2(U)$,

$$\left\| \int K(x, y) g(y) dy \right\|_{L^2(B(x_0, r))} \prec r\|g\|_{L^2(U)},$$

uniformly in x_0, r.

Proof. The L^2 operator norm of some integral operator $g \to \int L(x, y) g(y) dy$ is majorized by

$$\left(\sup_x \int |L(x, y)| dy\right)^{1/2} \left(\sup_y \int |L(x, y)| dx\right)^{1/2}. \tag{11.2}$$

Set $K' = K(x, y)\chi_{B(x_0, r)}(x)$ where χ denotes the characteristic function. We must estimate the operator norm of the operator whose integral kernel is K'. Decompose $K' = \sum_{j=-\infty}^{\log(Cr^{-1})} K_j$ where $K_j(x, y) = K'(x, y)\chi_{2^{j-1} \le \rho(x, y) \le 2^j r}$. For $j > 0$, K_j is supported on $\{\rho(x, x_0) \le r \text{ and } \rho(y, x_0) \le C2^j r\}$. For $x \in B(x_0, r)$,

$$\begin{aligned} \int |K_j(x, y)| dy &\le \|K_j\|_\infty |B(x_0, C2^j r)| \\ &\prec 2^j r \lambda(x_0, 2^j r)^{-1} \lambda(x_0, C2^j r) \\ &\sim 2^j r. \end{aligned}$$

On the other hand for $y \in B(x_0, C2^j r)$,

$$\begin{aligned} \int K_j(x, y) dx &\prec \|K_j\|_\infty |B(x_0, r)| \\ &\prec 2^j r \left(\lambda(x_0, r)/\lambda(x_0, 2^j r)\right) \\ &\prec 2^j r 2^{-4j}, \end{aligned}$$

so for K_j (11.2) is $\prec (2^j r)^{1/2}(2^{-3j} r)^{1/2} = 2^{-j} r$. When $j \le 0$, K_j is supported where $\rho(x, x_0) \prec r$, $\rho(y, x_0) \prec r$, and $\rho(x, y) \prec 2^j r$. The same holds true

when the roles of x and y are reversed, so (11.2) is $\prec 2^j r$. Summing over j concludes the proof.

The classical analogue of this lemma is that in $\mathbf{R}^n$, fractional integration of order 1 maps L^2 to L^p, $p^{-1} = 2^{-1} - n^{-1}$. However no inequality involving a higher order of integrability is appropriate in the present context, for no such inequality which implies the conclusion of the lemma can be valid. In $\mathbf{R}^n$ the analogue of the lemma follows by Hölder's inequality from the L^p inequality, because of the interplay of the formulas $p^{-1} = 2^{-1} - n^{-1}$ and $|B(x,r)| \sim r^n$; p must depend on the dimension n. In the present context one should think heuristically of $\log_r \lambda(x,r)$, the logarithm to the base r, as a variable dimension depending on both x and r. Thus there is no fixed p which works.

Lemma 11.4. *Suppose $f \in C^1(U)$. Then for any $r > 0$ and $x_0 \in U$,*

$$\|f\|_{L^2(B(x_0,r))} \prec r\left(\|Xf\|_{L^2(U)} + \|Yf\|_{L^2(U)} + \|f\|_{L^2(U)}\right),$$

uniformly in x_0, r, and f.

This lemma is not quite precisely phrased—the actual hypothesis is that $B(x_0,r)$ lies inside a compact subset of U, and the bound in the inequality is permitted to depend on this compact subset. The same goes in the next two lemmas.

Proof. Let $\Delta' = -X^2 - Y^2$. By Theorem 10 of [RS] plus Theorem 5 of [NSW] there exist kernels $k_i(x,y)$, $i = 0, 1$, such that

$$f(x) = \int k_1(x,y)(\Delta' f)(y)dy + \int k_0(x,y)f(y)dy$$

where

$$|k_0(x,y)| \prec \rho(x,y)\lambda(x,\rho(x,y))^{-1},$$
$$|k_1(x,y)| \prec \rho(x,y)^2\lambda(x,\rho(x,y))^{-1},$$

and for any differential operator $D = (X \text{ or } Y)\circ(X \text{ or } Y)\circ\cdots$ with any number n of factors, acting in the y variable,

$$|Dk_1(x,y)| \prec \rho(x,y)^{2-n}\lambda(x,\rho(x,y))^{-1}.$$

Integration by parts is therefore valid and yields

$$f(x) = \int k_2(x,y)(Xf)(y)dy + \int k_3(x,y)(Yf)(y)dy + \int k_0(x,y)f(y)dy$$

where all three kernels satisfy

$$|k_i(x,y)| \prec \rho(x,y)\lambda(x,\rho(x,y))^{-1}.$$

Therefore it suffices to apply Lemma 11.3.

The hypothesis $f \in C^1$ may be weakened by an approximation argument; it suffices to know that $f \in L^2$ and that $Xf, Yf \in L^2$ (in the sense of distributions) as well.

Corollary 11.5. *Suppose* $f \in \mathrm{Range}(\overline{\partial}_b)$ *and that* $u \in L^2$ *is the unique solution of* $\overline{\partial}_b u = f$ *satisfying* $u \perp \mathscr{H}_b$. *Then there exists* $C < \infty$ *such that for all* $x_0 \in U$ *and small* $r > 0$,

$$\|u\|_{L^2(B(x_0,r))} \leq Cr\|f\|_{L^2(M)}.$$

This follows from the last lemma plus Theorem A$'$. But what we shall actually require is a variant:

Corollary 11.5*. *Suppose that* $g \in L^2(M)$ *belongs to* $\mathrm{Range}(\overline{\partial}_b^*)$. *Let* v *(an* L^2 *section of* $B^{0,1}(M)$*) be the unique solution of* $\overline{\partial}_b^* v = g$ *satisfying* $v \perp \mathrm{kernel}(\overline{\partial}_b^*)$. *Then for all* $x_0 \in U$ *and* $r > 0$,

$$\|v\|_{L^2(B(x_0,r))} \prec r\|g\|_{L^2(M)}.$$

This is proved by reversing the roles of $\overline{\partial}_b$ and $\overline{\partial}_b^*$ throughout the entire theory. Since $\overline{\partial}_b$ has closed range, $\overline{\partial}_b^*$ automatically does also. There are two slight differences between the two: one maps functions to sections of $B^{0,1}$ while the other does the reverse, and $\overline{\partial}_b$ annihilates constants while $\overline{\partial}_b^*$ maps them to constant multiples of $c(x) \in C^\infty$. Neither difference has any effect in the proofs.

12. Pointwise bounds for the Szegö kernel

Let $S : L^2(M) \to \mathscr{H}_b$ denote the Szegö projection. We continue to assume that M is pseudoconvex, of finite type m, and that $\overline{\partial}_b$ has closed range.

Theorem C. *There exists* $K \in C^\infty(M \times M \backslash diagonal)$ *such that for any* $f, g \in L^2(M)$ *with disjoint supports,*

$$\langle g, Sf\rangle = \iint g(x)f(y)K(x,y)\,dx\,dy.$$

K *satisfies*

$$|K(x,y)| \leq C_0\lambda(x,\rho(x,y))^{-1}$$

for all $x \neq y \in M$. *More generally*

$$|D_x D_y K(x,y)| \leq C_{n_1,n_2}\rho(x,y)^{-n_1-n_2}\lambda(x,\rho(x,y))^{-1}$$

for any $D_x = (X \text{ or } Y) \circ (X \text{ or } Y) \circ \cdots$ *with* n_1 *factors acting in the* x *variable and* $D_y = (X \text{ or } Y) \circ (X \text{ or } Y) \circ \cdots$ *with* n_2 *factors, acting in the* y *variable, for any* $n_1, n_2 \geq 0$. C_0 *and* C_{n_1,n_2} *depend only on* M.

This was already known in special cases [NSW, M, NRSW1]. Recently Nagel, Rosay, Stein, and Wainger have obtained the same result, in the case where M is the boundary of a (pseudoconvex, finite type) domain in $\mathbf{C}^2$ [NRSW2]. Even more recently such a result has been obtained in certain higher-dimensional cases by Machedon [M2]. Note that although the arguments of §2 yield an expression for S as a sum of three principal terms (one each for P^0, P^-, P^+)

plus smoothing error terms, some of these terms appear not to be C^∞ away from the diagonal, since the kernel for $\mathscr{E}_0$ is not.

Let $\widehat{B}, \widehat{X}, \widehat{Y}$ be as in the last section. The proof of Theorem A$'$ establishes the following:

Proposition 12.1. *Let $\varepsilon_0, \varepsilon_1 > 0$. Let $\mathscr{B} \subset \widehat{B}$ be a Euclidean ball of radius ε_0, whose distance to the boundary of $\widehat{B}$ is at least ε_1. For any $\varepsilon_0, \varepsilon_1$ and any $\hat{u}, \hat{v}, \hat{g} \in L^2(\widehat{B})$ satisfying $\widehat{\overline{L}}\hat{u} = \hat{g}$ and $\widehat{\overline{L}}^* \hat{v} = \hat{u}$, for any n and any $\widehat{D} = (\widehat{X}$ or $\widehat{Y}) \circ (\widehat{X}$ or $\widehat{Y}) \circ \cdots$ with n factors,*

$$\|\widehat{D}\hat{u}\|_{L^\infty(\mathscr{B})} \le C_n \left(\|\hat{g}\|_{C^k(\widehat{B})} + \|\hat{u}\|_{L^2(\widehat{B})} + \|\hat{v}\|_{L^2(\widehat{B})} \right).$$

C depends only on $n, \varepsilon_0, \varepsilon_1$, on upper bounds for finitely many derivatives of the coefficients of $\widehat{X}, \widehat{Y}$, on the degree of the commutator $\widehat{V}$, and on a lower bound for $|\det(\widehat{X}, \widehat{Y}, \widehat{V})|$ on $\widehat{B}$. k depends on n and on the type m of M.

We now show how Theorem C is a formal consequence of Proposition 12.1, Corollary 11.5* and the machinery of [NSW].

Lemma 12.2. *The distribution-kernel K for $I - S$ is C^∞ away from the diagonal.*

The purpose of this preliminary lemma is just to permit us to discuss K as a function without qualms.

Proof. Let $f \in C^1$ be supported in some compact set A. Outside a neighborhood $\widetilde{A}$ of A, $u = (I - S)f$ satisfies $\overline{\partial}_b u = \overline{\partial}_b f \equiv 0$, and $u = \overline{\partial}_b^* v$ where $v \in L^2$ and $\|u\|_2, \|v\|_2 \prec \|f\|_2$. By Proposition 12.1, $u \in C^\infty$ on $M \backslash \widetilde{A}$, with bounds which depend only on $\|u\|_2$ and $\|v\|_2$, which in turn are majorized by a constant multiple of $\|f\|_2$. Since $I - S$ is bounded on L^2 we may pass to the limit to deduce that $I - S$ is bounded from $L^2(A)$ to $C^\infty(M \backslash \widetilde{A})$. Therefore for any $x_0 \in M \backslash \widetilde{A}$, any partial derivative with respect to x of $K(x, y)$ belongs to $L^2(A, dy)$. But $I - S$ is selfadjoint, so $K(x, y) = K(y, x)$ times a nonvanishing C^∞ function of x, y, so the roles of the x and y variables may be reversed. By a partition of unity argument and the Sobolev embedding lemma it follows that $K \in C^\infty$ on $M \times M$ minus any neighborhood of the diagonal.

The proof of Theorem C is just a more quantitative version of this argument. Consider two distinct points x_0, y_0 in some coordinate patch on M, at which we wish to estimate K and its derivatives. By the last lemma it suffices to consider points which are very close together. Let $r = \rho(x_0, y_0) > 0$. Let $\varepsilon > 0$ be a small constant. We study K on a small neighborhood of $(x_0, y_0) \in M \times M$. Let C_1 be a large constant and $\phi = \phi_{x_0, C_1 r}$. If C_1 is large enough then $\xi_0 = \phi^{-1} x_0$ and $\eta_0 = \phi^{-1} y_0$ are separated from one another and from the boundary of $\widehat{B}$ by a fixed distance, uniformly in x_0, y_0. Indeed it may

easily be shown that if $\hat{\rho}$ is constructed from $\widehat{X}, \widehat{Y}$ as ρ was from X, Y, then $\hat{\rho}(\xi, \eta) \sim r^{-1}\rho(\phi\xi, \phi\eta)$. Moreover $|\xi - \eta| \prec \hat{\rho}(\xi, \eta) \prec |\xi - \eta|^{m^{-1}}$, by the results of [NSW].

Let $B_1 = \{\xi \in \widehat{B} : |\xi - \xi_0| < \varepsilon\}$ and $B_2 = \{\eta \in \widehat{B} : |\eta - \eta_0| < \varepsilon\}$ where $\varepsilon > 0$ is a small constant, independent of x_0, y_0. Let $f \in L^2(\phi(B_2))$ be arbitrary. Let $u = (I - S)f$. Then on $\phi(B_1)$, $u(x) = \int K(x, y)f(y)dy$. Let $\hat{u}, \hat{f}$ be the pullbacks to $\widehat{B}$.

$$\begin{aligned}\hat{u}(\xi) &= u(\phi\xi)\\ &= \int K(\phi\xi, y)f(y)dy\\ &= \int K(\phi\xi, \phi\eta)\hat{f}(\eta)J(\eta)d\eta\end{aligned}$$

where $J = |\frac{\partial y}{\partial \eta}|$ is C^∞. By Theorem 11.1(5), $J \sim \lambda(x_0, r)$ on $\widehat{B}$, and each partial derivative $\frac{\partial^\alpha J}{\partial \eta^\alpha}$ is also $\prec \lambda(x_0, r)$, uniformly in x_0, r. For $\xi \in B_1$,

$$\hat{u}(\xi) = \int \widehat{K}(\xi, \eta)\hat{f}(\eta)d\eta,$$

where $\widehat{K}(\xi, \eta) = K(\phi\xi, \phi\eta)J(\eta)$.

Since $u = (I - S)f$ is orthogonal to $\mathscr{H}_b = \text{kernel}(\overline{\partial}_b)$, u belongs to Range $(\overline{\partial}_b^*)$, so there exists an L^2 section v of $B^{0,1}$ such that $v \perp \text{kernel}(\overline{\partial}_b^*)$, $\overline{\partial}_b^* v = u$ and by Corollary 11.5*, $\|r^{-1}v\|_{L^2(\phi(\widehat{B}))} \prec \|u\|_2$, which is $\prec \|f\|_2$. Therefore on $\widehat{B}$ we have $\widehat{\overline{L}}\hat{u} = \widehat{\overline{L}}\hat{f}$, $\widehat{\overline{L}}^* r^{-1}\hat{v} = \hat{u}$, and $\|\hat{u}\|_{L^2(\widehat{B})} + \|r^{-1}\hat{v}\|_{L^2(\widehat{B})} \prec \|\hat{f}\|_{L^2(\widehat{B})}$. Moreover on $2B_1$, the double of B_1, $\widehat{\overline{L}}\hat{f} \equiv \widehat{\overline{L}}0 \equiv 0$. We apply Proposition 12.1 on $2B_1$, with $\hat{g} = \widehat{\overline{L}}\hat{f} = 0$, to conclude that $\|\hat{u}\|_{C^n(B_1)} \prec \|\hat{f}\|_2$ for all n. $\hat{f} \in L^2(B_2)$ is completely arbitrary. Therefore for each $\widehat{D} = (\widehat{X} \text{ or } \widehat{Y}) \circ (\widehat{X} \text{ or } \widehat{Y}) \circ \cdots$ acting in the ξ variable,

$$\sup_{\xi \in B_1} \|\widehat{D}\widehat{K}(\xi, \cdot)\|_{L^2(B_2)}$$

is majorized by a constant which depends on the degree of $\widehat{D}$ but not on x_0, y_0. Again $\widehat{K}(\xi, \eta) = \widehat{K}(\eta, \xi)$ times a nonvanishing C^∞ factor, which is bounded above and below on $\widehat{B}$ uniformly in x_0, y_0 because $J \sim \lambda(x_0, r)$ uniformly. It follows from the Sobolev embedding lemma that $\widehat{K} \in C^\infty(B_1 \times B_2)$, uniformly in x_0, y_0. Since $\widehat{X}, \widehat{Y}$ have uniformly C^∞ coefficients, $\|\widehat{D}_\xi \widehat{D}_\eta \widehat{K}\|_{L^\infty(B_1 \times B_2)}$ is bounded uniformly for each $\widehat{D}_\xi, \widehat{D}_\eta = (\widehat{X} \text{ or } \widehat{Y}) \circ (\widehat{X} \text{ or } \widehat{Y}) \circ \cdots$ acting in the ξ, η variables respectively, with a bound depending only on the total number of factors of $(\widehat{X} \text{ or } \widehat{Y})$. If $\widehat{K}(\xi, \eta)$ is replaced by $K(\phi\xi, \phi\eta)$, the same result follows but with a factor of $\lambda(x_0, r)^{-1}$, since $\|J^{-1}\|_{C^n} \prec \lambda(x_0, r)^{-1}$ for all n. The general relation $\widehat{D}\hat{g}(\xi) \equiv r^n Dg(\phi\xi)$, where n is the number of factors of $(\widehat{X} \text{ or } \widehat{Y})$ in $\widehat{D}$, now implies the desired conclusion.

The idea of rescaling was pointed out to the author in connection with this problem some time ago by A. Nagel.

Since U has the structure of a space of homogeneous type, any operator bounded on L^2 whose distribution-kernel satisfies the estimates of Theorem C is automatically bounded on L^p for all $p \in (1, \infty)$. So Theorem B is proved as well.

Added in proof. Closely related results for domains in $\mathbf{C}^2$ have also been obtained by McNeal [Mc].

References

[BS] H. Boas and M. C. Shaw. *Sobolev estimates for the Lewy operator on weakly pseudoconvex boundaries*, Math. Ann. **274** (1986), 221–231.

[BMS] L. Boutet de Monvel and J. Sjöstrand, *Sur la singularité des noyaux de Bergmann et Szegö*, Soc. Mat. France Asterisque **34–35** (1976), 123–164.

[C1] M. Christ, *Hilbert transforms along curves*, III: *Rotational curvature*, preprint, unpublished, December 1984.

[C2] ——, *On the regularity of inverses of singular integral operators*, Duke Math. J. (to appear).

[C3] ——, *Pointwise estimates for the relative fundamental solution of* $\overline{\partial}_b$, Proc. Amer. Math. Soc. (to appear).

[C4] ——, *On the* $\overline{\partial}_b$ *equation and Szegö projection on CR manifolds*, Proc. 1987 El Escorial Conf. on Harmonic Analysis and Partial Differential Equations, Lecture Notes in Math., Springer-Verlag, Berlin (to appear).

[C5] ——, *Inversion in some algebras of singular integral operators*, MSRI preprint, April 1988.

[CG] M. Christ and D. Geller, *Singular integral characterizations of Hardy spaces on homogeneous groups*, Duke Math. J. **51** (1984), 547–598.

[CNSW] M. Christ, A. Nagel, E. M. Stein, and S. Wainger (in preparation).

[CW] R. R. Coifman and G. Weiss, *Analyse harmonique sur certains espaces homogenes*, Lecture Notes in Math., no. 242, Springer-Verlag, Berlin, 1971.

[F] C. Fefferman, *The Bergman kernel and biholomorphic mappings of pseudoconvex domains*, Invent. Math. **26** (1974), 1–65.

[FK] C. Fefferman and J. J. Kohn, *Hölder estimates on domains in two complex dimensions and on three-dimensional CR manifolds*, preprint.

[FS] C. Fefferman and A. Sánchez-Calle, *Fundamental solutions for second order subelliptic operators*, Ann. of Math. **124** (1986), 247–272.

[Fo] G. B. Folland, *Subelliptic estimates and function spaces on nilpotent Lie groups*, Ark. Math. **13** (1975), 161–207.

[FSt1] G. B. Folland and E. M. Stein, *Estimates for the* $\overline{\partial}_b$ *complex and analysis on the Heisenberg group*, Comm. Pure Appl. Math. **27** (1974), 429–522.

[FSt2] ——, *Hardy spaces on homogeneous groups*, Princeton Univ. Press, Princeton, N.J., 1982.

[GS] P. C. Greiner and E. M. Stein, *Estimates for the* $\overline{\partial}$*-Neumann problem*, Princeton Univ. Press, Princeton, N.J., 1977.

[H1] L. Hörmander, *Hypoelliptic second order differential equations*, Acta Math. **119** (1967), 147–171.

[H2] ——, *Introduction to complex analysis in several variables*, North-Holland, Amsterdam, 1973.

[H3] ——, *Differential equations without solutions*, Math. Ann. **140** (1960), 169–173.

[KeSt] N. Kerzman and E. M. Stein, *The Szegö kernel in terms of Cauchy-Fantappie kernels*, Duke Math. J. **45** (1978), 197–224.

[KS] A. W. Knapp and E. M. Stein, *Intertwining operators for semisimple groups*, Ann. of Math. **93** (1971), 489–578.

[K1] J. J. Kohn, *The range of the tangential Cauchy-Riemann operator*, Duke Math. J. **53** (1986), 525–545.

[K2] ——, *Estimates for $\overline{\partial}_b$ on pseudoconvex CR manifolds*, Proc. Symp. Pure Math. **43** (1985), 207–217.

[K3] ——, *Microlocal analysis of the $\overline{\partial}$-Neumann problem, $\overline{\partial}_b$, Δ_b and related operators* (in preparation).

[K4] ——, *Boundaries of complex manifolds*, Proc. Conf. Complex Analysis, Minneapolis, 1964, Springer-Verlag, Berlin, 1965.

[L] H. Lewy, *An example of a smooth linear partial differential equation without solution*, Ann. of Math. **66** (1957), 155–158.

[M1] M. Machedon, *Estimates for the parametrix of the Kohn Laplacian on certain domains*, Invent. Math. **91** (1988), 339–364.

[M2] ——, *Szegö kernels on pseudoconvex domains with one degenerate eigenvalue*, preprint.

[Mc] J. McNeal, *Boundary behavior of the Bergman kernel*, preprint.

[N] A. Nagel, *Vector fields and nonisotropic metrics*, Beijing Lectures in Harmonic Analysis (E. M. Stein, ed.), Princeton Univ. Press, Princeton, N.J., 1986.

[NRSW1] A. Nagel, J.-P. Rosay, E. M. Stein, and S. Wainger, *Estimates for the Bergman and Szegö kernels in certain weakly pseudoconvex domains*, Bull. Amer. Math. Soc. **18** (1988), 55–59.

[NRSW2] ——, *Estimates for the Bergman and Szegö kernels in* $\mathbf{C}^2$, MSRI preprint, February 1988.

[NSW] A. Nagel, E. M. Stein, and S. Wainger, *Balls and metrics defined by vector fields*, I: *Basic properties*, Acta Math. **155** (1985), 103–147.

[RS] L. P. Rothschild and E. M. Stein, *Hypoelliptic differential operators and nilpotent groups*, Acta Math. **137** (1976), 247–320.

[S] A. Sánchez-Calle, *Fundamental solutions and geometry of the sum of squares of vector fields*, Invent. Math. **78** (1984), 143–160.

[St] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Univ. Press, Princeton, N.J., 1970.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CALIFORNIA 90024