ON SOME SUBELLIPTIC REAL ANALYSIS

PO-LAM YUNG

In this and the next talk we shall analyze some real partial differential operators that are subelliptic in nature. This will pave our way towards a discussion of some subelliptic systems of equations that naturally arise in several complex variables in the last talk.

We shall begin by discussing some background material, and work ourselves towards some Sobolev type $L^p - L^q$ estimates in this context.

1. Hormander's Theorem

Let $X_1, \dots, X_n$ be some smooth real vector fields on $\mathbb{R}^N$. A commutator of these vector fields of length r is a vector field of the form

$$[X_{i_1}, [X_{i_2}, [\dots, [X_{i_{r-1}}, X_{i_r}]]]].$$

Definition 1 (Hormander [3])**.** *The vector fields $X_1, \dots, X_n$ are said to satisfy* Hormander's finite type condition *at a point ξ if they and their commutators of length $\leq r$ span the tangent space of $\mathbb{R}^N$ at ξ for some positive integer r. The smallest r for which this holds is called the* type *of $X_1, \dots, X_n$ at ξ.*

For simplicity, in such a situation we shall often just say that $X_1, \dots, X_n$ are of finite type r. Note that in general we may have fewer vector fields than the Euclidean dimension of the underlying space. In other words, n may be smaller than N.

One trivial example is the case when we have N vector fields $\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_N}$ on $\mathbb{R}^N$. These vector fields are of finite type 1. The analysis associated with these vector fields are well-known. What we do below can be thought of as a generalization of this analysis.

Two more interesting examples are:

Example 1. *On $\mathbb{R}^2$, where the coordinates are given by (x,t), let*

$$X_1 = \frac{\partial}{\partial x}, \quad X_2 = x^r \frac{\partial}{\partial t}, \qquad r \geq 1.$$

Then $[X_1, [X_1, [\dots, [X_1, X_2]]]] = r! \frac{\partial}{\partial t}$ where the bracket has length $r+1$. Hence X_1, X_2 are of finite type $r+1$ at 0. (For simplicity, we shall focus below on the case where $r = 1$.)

Example 2. *On $\mathbb{R}^3$, where the coordinates are given by (x,y,t), let*

$$X_1 = \frac{\partial}{\partial x}, \quad X_2 = \frac{\partial}{\partial y} + x\frac{\partial}{\partial t}.$$

Then $[X_1, X_2] = \frac{\partial}{\partial t}$, so X_1, X_2 are of finite type 2 at 0.

Date: October 28, 2009.

These will serve as our main motivating examples. As we shall see shortly, in Example 2, it is possible to make $\mathbb{R}^3$ into a Lie group H, so that X_1, X_2 become left-invariant vector fields on H. This makes its study easier than Example 1.

In studying vector fields that satisfy this finite type condition, Hormander introduced the following 'sum of squares' operator

$$L = \sum_{j=1}^{n} X_j^2. \tag{1}$$

When $X_1, \dots, X_N$ are just $\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_N}$ on $\mathbb{R}^N$, this L is just the ordinary Laplace operator; it is second order elliptic. The remarkable fact about this L is that while in general L fails to be elliptic, it is nonetheless *hypoellpitic*:

Definition 2. *A partial differential operator L is said to be hypoelliptic on an open set Ω if for any open subset $U \subseteq \Omega$ and for any distribution u on Ω that satisfies $Lu \in C^\infty(U)$, we have $u \in C^\infty(U)$.*

Theorem 1 (Hormander)**.** *Suppose $X_1, \dots, X_n$ are smooth real vector fields of finite type r at every point on an open set $\Omega \subseteq \mathbb{R}^N$, and L be the sum of squares operator defined by (1). Then there exists $\varepsilon > 0$ such that for all $u \in C_c^\infty(\Omega)$ and $s, m \in \mathbb{R}$, we have*

$$\|u\|_{W^{s+\varepsilon,2}} \lesssim \|Lu\|_{W^{s,2}} + \|u\|_{W^{-m,2}}. \tag{2}$$

In fact $\varepsilon = 2^{1-r}$ will do.

It then follows that such L is hypoelliptic on Ω. This theorem is very remarkable because in general such operators L are not elliptic. For example, in Example 1 above, if $r = 1$ and $L = X_1^2 + X_2^2 = \frac{\partial^2}{\partial x^2} + x^2 \frac{\partial^2}{\partial t^2}$, then a simple dilation invariance argument shows that we cannot have

$$\|u\|_{W^{2,2}} \lesssim \|Lu\|_{L^2} + \|u\|_{L^2}$$

for all smooth functions u supported in a neighborhood of 0; in other words we cannot gain as many as two derivatives in L^2 knowing only $Lu \in L^2$. In fact if $\|u\|_{W^{\varepsilon,2}} \lesssim \|Lu\|_{L^2} + \|u\|_{L^2}$ for all smooth u supported in a neighborhood of 0, then $\varepsilon \leq 1$. Hence L cannot be elliptic.

Opeartors satisfying (2) are said to be *subelliptic* on Ω. They are in general not as well behaved as elliptic operators; nevertheless they are hypoelliptic. Similar phenomena were in fact first observed in the study of several complex variables. The analysis there, however, is more complicated, because the equations involved there are systems of equations (rather than a single scalar equation as we had above). We shall only turn to several complex variables in the next talk.

The finite type condition is really the correct condition to impose here, because without this, Hormander's theorem, as well as all the results that we shall describe in what follows, will fail to hold by simple considerations of dilation invariance.

A word about the proof of Hormander's theorem. It is based on the technique of *commutator estimates*. For instance, in the Example 1 we had above, it is easy to see that

$$\|X_1 u\|_{L^2}^2 + \|X_2 u\|_{L^2}^2 \leq \|Lu\|_{L^2}^2 + \|u\|_{L^2}^2 \tag{3}$$

for all $u \in C_c^\infty(\Omega)$ where Ω is a neighborhood of 0 in $\mathbb{R}^2$; one just has to observe that upon integrating by parts,

$$\|X_1 u\|_{L^2}^2 + \|X_2 u\|_{L^2}^2 = (u, Lu)_{L^2} \leq \|u\|_{L^2} \|Lu\|_{L^2}.$$

Thus it remains to bound

$$\left\| \frac{\partial}{\partial t} u \right\|^2_{W^{-\frac{1}{2},2}} = \left([X_1, X_2]u, \Delta^{-\frac{1}{2}} \frac{\partial u}{\partial t} \right)_{L^2}.$$

But integrating by parts, we can write this as

$$-\left(X_2 u, X_1 \Delta^{-\frac{1}{2}} \frac{\partial u}{\partial t} \right)_{L^2} + \left(X_1 u, X_2 \Delta^{-\frac{1}{2}} \frac{\partial u}{\partial t} \right)_{L^2}.$$

Commuting X_1 and X_2 past $\Delta^{-\frac{1}{2}} \frac{\partial}{\partial t}$, and using that $\Delta^{-\frac{1}{2}} \frac{\partial}{\partial t}$ preserves L^2 with (3), one bounds this by

$$\begin{aligned} &\|X_2 u\|_{L^2} \left(\|X_1 u\|_{L^2} + C\|u\|_{L^2} \right) + \|X_1 u\|_{L^2} \left(\|X_2 u\|_{L^2} + C\|u\|_{L^2} \right) \\ \lesssim &(\|Lu\|_{L^2} + \|u\|_{L^2})^2 \end{aligned}$$

This proves

$$\|u\|_{W^{\frac{1}{2},2}} \lesssim \|Lu\|_{L^2} + \|u\|_{L^2}$$

for all smooth u supported near 0. The proof of the general theorem follows a similar line and estimates

$$\|u\|_{W^{\varepsilon,2}} \lesssim \sum_{k=1}^{r} \sum_{i_1,\dots,i_k} \left\| [X_{i_1}, [X_{i_2}, [\dots, [X_{i_{k-1}}, X_{i_k}]]]]u \right\|_{W^{\varepsilon-1,2}} + \|u\|_{L^2}$$

by commutator estimates.

2. Sharp L^p estimates

Hormander's theory was based on commutator estimates and thus L^2 in nature. The next breakthrough in the theory came when Rothschild and Stein [5] proved the following L^p estimate for the sum of squares operator.

Theorem 2 (Rothschild-Stein). *Let $L = \sum_{j=1}^n X_j^2$ be as in Theorem 1. Then for all all $\Omega' \Subset \Omega$, there is a constant $C > 0$ such that for all functions $u \in C^\infty(\Omega)$, we have*

$$\|X_j X_k u\|_{L^p(\Omega')} \leq C \left(\|Lu\|_{L^p(\Omega)} + \|u\|_{L^p(\Omega)} \right)$$

for all $1 < p < \infty$.

By introducing an analytic family of operators and interpolation, one can also prove the following estimate:

$$\|au\|_{W^{\frac{2}{r},p}} \lesssim \|Lu\|_{L^p(\Omega)} + \|u\|_{L^p(\Omega)}, \quad 1 < p < \infty$$

for all smooth cut-off $a \in C_c^\infty(\Omega)$ (where the implicit constant depends on a). This improves Hormander's theorem even when $p = 2$. We shall not pursue this further.

2.1. The case of homogeneous groups.

The proof of Theorem 2 comes in two steps. The first step, roughly speaking, consists of studying the special case when the underlying space is a *Lie group*, and when the vector fields $X_1, \dots, X_n$ are *left-invariant*. The advantage of considering such a situation is that one can define *convolutions* in this case:

$$(f * g)(\xi) = \int f(\xi \cdot \eta^{-1}) g(\eta) d\eta,$$

and left-invariant vector fields are very compatible with convolutions:

$$X_j(f * g) = f * (X_j g).$$

In particular

$$L(f * g) = f * (Lg).$$

Hence if we can find a distribution K such that

$$LK = \delta_0,$$

then $u = f * K$ solves $Lu = f$. To find such a distribution K we need a little more: we need that the Lie group to carry an *automorphic dilation*, making it a *homogeneous group*. Thus it makes sense to talk about homogeneous functions or distributions on the group, and also the degree of a differential operator. It also allows us to talk about the *homogeneous dimension* Q of the group. What will happen is that L will be a homogeneous differential operator of degree 2, and we will be able to find a homogeneous distribution K of degree $-(Q-2)$ that solves $LK = \delta_0$ if $X_1, \dots, X_n$ form a *basis* of left-invariant vector fields of degree 1. (c.f. the case of the ordinary Laplacian Δ on $\mathbb{R}^N$.) We can then obtain an estimate of u in terms of Lu.

An example is given by our Example 2 above. If we use coordinates (x, y, t) on $\mathbb{R}^3$ and make it a group H by imposing the group law

$$(x, y, t) \cdot (\alpha, \beta, \gamma) = (x + \alpha, y + \beta, t + \gamma + x\beta),$$

then our vector fields $X_1 = \frac{\partial}{\partial x}$ and $X_2 = \frac{\partial}{\partial y} + x\frac{\partial}{\partial t}$ become left-invariant. We can also define a dilation on H by setting

$$\delta_\lambda(x, y, t) = (\lambda x, \lambda y, \lambda^2 t), \quad \lambda > 0.$$

δ_λ is a group automorphism for each $\lambda > 0$. A function f on H is said to be homogeneous of degree k if

$$f(\delta_\lambda(x, y, t)) = \lambda^k f(x, y, t)$$

for all (x, y, t) and λ. For example, the 'norm function'

$$|(x, y, t)| := |x| + |y| + |t|^{1/2}$$

is homogeneous of degree 1, and $|(x, y, t)|^k$ is homogeneous of degree k. If P is a differential operator on H, then it is said to be homogeneous of degree k if P lowers the degree of every homogeneous function by k. For instance, $X_1 = \frac{\partial}{\partial x}$ and $X_2 = \frac{\partial}{\partial y} + x\frac{\partial}{\partial t}$ are of degree 1, and $L = X_1^2 + X_2^2$ is of degree 2. In fact X_1, X_2 is a basis of left-invariant vector fields of degree 1. The homogeneous dimension Q of this group is said to be 4, because if we pull back the Haar measure $dxdydt$ of the group by the automorphic dilation δ_λ, we get $\lambda^4 dxdydt$. In particular, a distribution of degree -4 on the group would just barely fail to be in integrable with respect to the Haar measure (just like $|x|^{-n}$ barely fails to be integrable in $\mathbb{R}^n$).

In this example it is particularly easy to invert the operator L and obtain estimates as in the Theorem; in fact there is a homogeneous distribution K of degree $-(4-2)$ such that $u = f * K$ solves $Lu = f$. Hence for $u \in C_c^\infty(H)$,

$$X_j X_k u = X_j X_k (Lu * K),$$

and letting the derivatives fall on K, we see that $X_j X_k u$ can be obtained from Lu by convolving against a distribution of the critical degree -4. It turns out that there is a variant of the theory of singular integrals on H, and applying that theory one can conclude that

$$\|X_j X_k u\|_{L^p} \lesssim \|Lu\|_{L^p}$$

for all $1 < p < \infty$.

The group in this example is usually called the first *Heisenberg group*, and it arises naturally from considerations in both several complex variables and quantum mechanics.

In general, if $X_1, \dots, X_n$ is a basis of left-invariant vector fields of degree 1 on a homogeneous group of homogeneous dimension Q, and $L = \sum_{j=1}^n X_j^2$, then a solution operator to L is given by convolving against a homogeneous distribution of degree $-(Q-2)$, and we can conclude the proof of the theorem in this situation.

2.2. **Lifting of vector fields.** The second step to proving the theorem involves a reduction to the special case considered in the first step. This is done by a technique called *lifting* of the vector fields involved. It is the most transparent to illustrate this with our Example 1 above. Suppose again we are on $\mathbb{R}^2$ with coordinates (x,t) and $X_1 = \frac{\partial}{\partial x}$, $X_2 = x\frac{\partial}{\partial t}$, $L = X_1^2 + X_2^2$. It is impossible to make X_1 and X_2 the left-invariant vector fields of any homogeneous group; in fact the left-invariant vector field of any Lie group cannot vanish on a lower dimensional submanifold, but X_2 does. It is, nonetheless, possible to lift these vector fields to the homogeneous group $H \simeq \mathbb{R}^3$ we considered in the example in the first step and make such left-invariant vector fields on H: Consider the projection

$$\pi\colon H \to \mathbb{R}^2, \quad \pi(x,y,t) = (x,t),$$

and let

$$\tilde{X}_1 = \frac{\partial}{\partial x}, \quad \tilde{X}_2 = \frac{\partial}{\partial y} + x\frac{\partial}{\partial t}.$$

Then

$$d\pi(\tilde{X}_1) = X_1, \quad d\pi(\tilde{X}_2) = X_2,$$

and $\tilde{X}_1, \tilde{X}_2$ is called a lift of X_1, X_2. The advantage here is that $\tilde{X}_1, \tilde{X}_2$ forms a basis of left-invariant vector fields of degree 1 on H, and we know how to handle such. Any function u on $\mathbb{R}^2$ can be lifted to a function $\tilde{u}$ on H by pulling back via π: we define $\tilde{u} := u \circ \pi$, i.e.

$$\tilde{u}(x,y,t) = u(x,t).$$

Since $d\pi(\tilde{X}_k) = X_k$, we have

$$\tilde{X}_k \tilde{u} = \widetilde{X_k u}.$$

Define now

$$\tilde{L} := \tilde{X}_1^2 + \tilde{X}_2^2.$$

Then if $Lu = f$ on $\mathbb{R}^2$, we have $\tilde{L}\tilde{u} = \tilde{f}$, so $\tilde{u}$ can be recovered from $\tilde{f}$ by a convolution, and this enables one to solve $Lu = f$. This in turn allows one to obtain estimates to the solution of $Lu = f$.

In general, given real vector fields $X_1, \dots, X_n$ satisfying Hormander's condition, it is not always possible to lift the vector fields and make them the left-invariant vector fields of a homogeneous group. What can be done, however, is that we can always *approximate* the lifted vector fields by left-invariant vector fields of a homogeneous group. This incurs additional (lower degree) error terms, and the formula we get will not be exact. Nevertheless one gets a *parametrix* for the sum of squares operator L. It is the size of the kernel of this parametrix to which we now turn.

3. Kernel estimates and subelliptic geometry

First we observe that in the group case, the size of the parametrix to the sum of squares operator L is pretty much dictated by homogeneity. To understand the size of the parametrix in the general case, however, we need a better understanding of the underlying subelliptic geometry.

Let $\Omega \subseteq \mathbb{R}^N$ be an open set. Given a smooth real vector field X on Ω, let $\exp(X)\xi$ be the time-1-flow along the integral curve of X starting at a point ξ (if the flow does not flow out of Ω in time 1). For instance, if $X = \sum_{j=1} a_j \frac{\partial}{\partial x_j}$ has constant coefficients, then $\exp(X)\xi$ is just the point $\xi + a$ where $a = (a_1, \dots, a_N)$ (as long as $\xi + a \in \Omega$). If the underlying space is a Lie group and if X is a left-invariant vector field, then $\exp(X)\xi$ is just the product $\xi \cdot \exp(X)$ taken using the group multiplication.

Suppose as usual that $X_1, \dots, X_n$ are smooth real vector fields of finite type r at every point of Ω. For each $1 \leq j \leq r$, let $\{X_{jk}\}_{k=1}^{\tilde{n_j}}$ be an enumeration of the commutators of $X_1, \dots, X_n$ of length j. For any point ξ and any small $r > 0$, let $B(\xi, r)$ be the set of all points of the form

$$\exp\left(\sum_{j=1}^{r}\sum_{k=1}^{\tilde{n_j}} a_{jk} X_{jk}\right)\xi, \qquad |a_{jk}| \leq r^j \text{ for all } j, k.$$

In other words, roughly speaking one is allowed to go further in some directions (namely those represented by commutators of shorter lengths) than in others (those only represented by long commutators). Thus $B(\xi, r)$ is usually thought of as the *non-isotropic ball* centered at ξ and of radius r. This in turns allows us to define a 'metric' ρ on Ω: we set $\rho(\xi, \eta)$ to be the infimum of all $r > 0$ such that $\eta \in B(\xi, r)$. This metric may not be finite for all $\xi, \eta \in \Omega$, but if η is in a sufficiently small neighborhood of ξ then $\rho(\xi, \eta)$ is finite.

What is remarkable here is that Nagel, Stein and Wainger [4] obtained a bound on the parametrix of $L = \sum_{j=1}^n X_j^2$ in terms of the volumes of the balls, as well as the metric, defined by the vector fields $X_1, \dots, X_n$:

Theorem 3 (Nagel-Stein-Wainger)**.** *If $K(\xi, \eta)$ is the parametrix of L as above, then near the diagonal we have*

$$|K(\xi,\eta)| \lesssim \frac{\rho(\xi,\eta)^2}{|B(\xi,\rho(\xi,\eta))|}, \quad \text{when } N \geq 3;$$

$$|X_j K(\xi,\eta)| \lesssim \frac{\rho(\xi,\eta)}{|B(\xi,\rho(\xi,\eta))|}, \quad \text{when } N \geq 2; \text{ and}$$

$$|X_j X_k K(\xi,\eta)| \lesssim \frac{1}{|B(\xi,\rho(\xi,\eta))|}, \quad \text{when } N \geq 2$$

where in the last two estimates the derivatives X_j and X_k can fall on either ξ or η. Here the volumes of the balls are computed using the Euclidean volume measure.

It is thus important to compute the volumes of these balls defined by the vector fields. In fact in the same paper, Nagel, Stein and Wainger proved a formula for the volumes of these balls, by considering the Jacobian determinant of the exponential map. We refrain, however, from giving the full formula because it is rather complicated. We shall just observe that for our simple Example 1 where we have $X_1 = \frac{\partial}{\partial x}$ and $X_2 = x\frac{\partial}{\partial t}$ on $\mathbb{R}^2$, it is easy to compute the exponential map explicitly, and if one carries out the computation, one sees that

$$|B((x,t),r)| \simeq \begin{cases} |x|r^2 & \text{if } r < |x| \\ r^3 & \text{if } r \geq |x| \end{cases},$$

which is exactly what the formula of Nagel-Stein-Wainger would give us in this special case. It is thus clear that these balls satisfy the *volume doubling property*:

$$|B(\xi, 2r)| \leq C|B(\xi, r)|$$

for all sufficiently small r, which also holds in the general situation by the results of Nagel-Stein-Wainger.

We remark that while the above estimates of the kernel does not look symmetric in ξ and η, it is more or less so: this is because

$$|B(\xi, \rho(\xi, \eta))| \simeq |B(\eta, \rho(\xi, \eta))|$$

by the doubling property.

We also remark that the following notion of *Carnot-Caratheodory distance* also often arises in a discussion of subelliptic geometry. The starting point is the theorem of Caratheodory [1] and Chow [2]:

Theorem 4 (Caratheodory, Chow)**.** *Suppose $X_1, \dots, X_n$ are smooth real vector fields on a connected open set Ω satisfying Hormander's condition at every point. Then for any two points $p, q \in \Omega$, there is a piecewise smooth curve joining p to q such that at every point where the curve is smooth, the tangent vector to the curve is a linear combination of $X_1, \dots, X_n$.*

This allows us to define a metric on Ω.

Definition 3. *Let X_1, $\dots$, X_n be as in the previous theorem. The* Carnot-Caratheodory *distance between two points p and q, denoted $d(p,q)$, is the infimum of all $r > 0$ such that the following holds: there exists a piecewise smooth curve $\phi\colon [0,1] \to \Omega$ joining p to q such that whenever ϕ is smooth,*

$$\phi'(t) = \sum_{j=1}^{n} a_j(t) X_j(\phi(t)) \quad \textit{with } |a_j(t)| \leq r \textit{ for all } j.$$

It is a consequence of the previous theorem that this distance is finite between any two points on Ω (again we are assuming Ω to be connected). This notion of distance is also of interest in control theory.

The problem with such a definition is that it is very difficult to compute. It is thus quite remarkable that Nagel, Stein and Wainger proved in their paper that this Carnot-Caratheodory metric d is locally equivalent to our more explicit metric ρ that we described before.

4. Sobolev inequality for functions

We are now ready to discuss a subellptic Sobolev inequality for functions. The question is the following. Suppose $X_1, \dots, X_n$ are smooth real vector fields on $\mathbb{R}^N$, and that they are finite type at a point, say 0. If u is a nice compactly supported smooth function and we control $u, X_1u, \dots, X_nu$ in L^p, can we say that $u \in L^q$ locally for some $q > p$?

If we are in the Euclidean situation where the vector fields are $\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_N}$ on $\mathbb{R}^N$, then the answer is given by the classical Sobolev inequality: For $u \in C_c^\infty(\mathbb{R}^N)$, we have

$$\|u\|_{L^{p^*}} \lesssim \|\nabla u\|_{L^p}, \quad \frac{1}{p^*} = \frac{1}{p} - \frac{1}{N},$$

where $1 \leq p < N$. The gain in exponent here depends on the dimension of the underlying space. The higher the dimension, the less one gains. What happens, in general, is that we shall need a local notion of non-isotropic dimension attached to these vector fields that satisfy the finite type condition, and only formulate a local Sobolev inequality in terms of that. In fact in the subelliptic case, the non-isotropic dimension will be bigger than the Euclidean dimension N of the underlying space, and thus we gain less than the Euclidean situation (as expected).

Definition 4. *Suppose $X_1, \dots, X_n$ are smooth real vector fields that are of finite type r at 0. Take commutators of $X_1, \dots, X_n$ of length $\leq j$ and restrict them to 0; call the subspace of the tangent space at 0 that they span $V_j(0)$. Clearly $V_{j-1}(0) \subseteq V_j(0)$ for all j; we let*

$$n_1 = dim V_1(0) \quad \text{and} \quad n_j = dim V_j(0) - dim V_{j-1}(0) \quad \text{for } j \geq 2.$$

We then define the non-isotropic dimension Q at 0 to be

$$Q = \sum_{j=1}^{r} j n_j.$$

For example, in Example 1 where we had $\frac{\partial}{\partial x}$, $x\frac{\partial}{\partial t}$ on $\mathbb{R}^2$, at 0 we have $n_1 = 1$ and $n_2 = 1$, so $Q = 1 \times 1 + 2 \times 1 = 3$; in Example 2 where we had $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y} + x\frac{\partial}{\partial t}$ on $\mathbb{R}^3$, at 0 we have $n_1 = 2$ and $n_2 = 1$, so $Q = 1 \times 2 + 2 \times 1 = 4$.

Theorem 5. *Let $X_1, \dots, X_n$ and Q be as in the previous definition. Then there exists a neighborhood Ω of 0 and $C > 0$ such that if $u \in C_c^\infty(\Omega)$ and $1 \leq p < Q$, then*

$$\|u\|_{L^{p^*}(\Omega)} \leq C\left(\|\nabla_b u\|_{L^p(\Omega)} + \|u\|_{L^p(\Omega)}\right), \quad \frac{1}{p^*} = \frac{1}{p} - \frac{1}{Q}$$

where $\nabla_b u = (X_1 u, \dots, X_n u)$ is the subelliptic gradient *of u.*

For example, in Example 1, this says we have a Sobolev inequality where 3 plays the role of the usual dimension; in fact by dilation invariance, we can scale the lower order term $\|u\|_{L^p}$ away, and get

$$\|u\|_{L^{p^*}} \lesssim \left\|\frac{\partial u}{\partial x}\right\|_{L^p} + \left\|x\frac{\partial u}{\partial t}\right\|_{L^p}, \quad \frac{1}{p^*} = \frac{1}{p} - \frac{1}{3}$$

for all $u \in C_c^\infty(\mathbb{R}^2)$, where $1 \leq p < 3$.

A few remarks are in order. First, it was already proven, in the paper of Nagel-Stein-Wainger, that if $u, \nabla_b u \in L^p_{\text{loc}}$, then $u \in W^{\frac{1}{r},p}_{\text{loc}}$ where r is the type of $X_1, \dots, X_n$. This already gives, by the classical Sobolev embedding, that locally $u \in L^q_{\text{loc}}$ for some $q > p$. But this is not as sharp as the previous theorem, because $u \in W^{\frac{1}{r},p}_{\text{loc}}$ is a homogeneous estimate that does not distinguish the good directions from the bad ones. In other words, we are not using the fact that we gain more derivatives in the good directions than in the bad ones, and thus we lose.

Second, the above result is known when $X_1, \dots, X_n$ is a basis of left-invariant vector fields of degree 1 on a homogeneous group. In fact more was known: Caponga, Danielli and Garafalo proved a similar inequality, but with a possibly different Q, that depends on the doubling condition on the volumes of the balls associated to the vector fields. This is, however, not as sharp as what we have above; in fact our Q is the smallest possible Q for which the theorem could hold, by some approximate dilation invariance. Thus the p^* we have is optimal.

As a simple application of this Sobolev inequality, let's observe the following $L^p - L^q$ estimate for the solution of the sum of squares operator $L = \sum_{j=1}^n X_j^2$: in fact under the assumptions of Theorem 2, if $1 < p < Q/2$, we then have

$$\|u\|_{L^{\frac{Qp}{Q-2p}}(\Omega')} \lesssim \|Lu\|_{L^p(\Omega)} + \|u\|_{L^p(\Omega)}.$$

There are two slightly different, but closely related, proofs to the Theorem 5, when $p > 1$. The first one is more complicated, and extends to the case $p = 1$. The second one, while elementary in nature, does not appear to extend to $p = 1$. We shall describe both of those.

The first proof relies on the parametrix we constructed for the sum of squares $L = \sum_{j=1}^n X_j^2$. Take $a, b, c \in C_c^\infty$ supported in a neighborhood of 0, such that $b = 1$ on the support of a and $c = 1$ on the support of b. Then from the parametrix of L, one obtains the following representation formula:

$$au = \sum_{j=1}^n T_j(bX_j u) + T_0(cu)$$

for some integral operators $T_0, \dots, T_n$. In particular, if we take $c \equiv 1$ near 0 and Ω to be an open set contained in the set where $c \equiv 1$, then for all $u \in C_c^\infty(\Omega)$,

$$u = \sum_{j=1}^n T_j(X_j u) + T_0 u.$$

In the Euclidean situation, all that we have done here would be just that if K is the fundamental solution to Δ so that $u = K * \Delta u$, then

$$u = K * \sum_{j=1}^N \frac{\partial^2 u}{\partial x_j^2} = \sum_{j=1}^N \frac{\partial K}{\partial x_j} * \frac{\partial u}{\partial x_j}.$$

This allows us to reproduce u from its derivatives; in fact in the Euclidean case we can also obtain an estimate of these kernels $\frac{\partial K}{\partial x_j}$, and a similar estimate is also available in the subelliptic case. It turns out that the kernels of T_j, which we also denote by $T_j(\xi, \eta)$, satisfy locally

$$|T_j(\xi, \eta)| \lesssim \frac{\rho(\xi, \eta)}{|B(\xi, \rho(\xi, \eta))|}$$

(and are smooth away from the diagonal). The claim is that these kernels are in weak-$L^{\frac{Q}{Q-1}}(d\eta)$ uniformly in ξ, and in weak-$L^{\frac{Q}{Q-1}}(d\xi)$ uniformly in η. It will then follow that the operators T_j maps L^p to weak-L^{p^*} for all $1 \leq p < Q$, and by Marcinkiewicz interpolation we conclude the theorem in the case $p > 1$. The case $p = 1$ will then follow from Maz'ya's truncation argument, which we discussed last time.

So it remains to establish the weak-type estimates for the kernel which is

$$\frac{\rho(\xi, \eta)}{|B(\xi, \rho(\xi, \eta))|}$$

near the diagonal. At this stage it is best to restrict ourselves to the simple Example 1 where where we had $\frac{\partial}{\partial x}$, $x\frac{\partial}{\partial t}$ on $\mathbb{R}^2$. Remember that

$$|B((x,t),r)| \simeq \begin{cases} |x|r^2 & \text{if } r < |x| \\ r^3 & \text{if } r \geq |x| \end{cases}.$$

Hence if we fix $\xi = (x,t)$ and let $r = r(\eta) = \rho(\xi,\eta)$, then

$$\frac{\rho(\xi,\eta)}{|B(\xi,\rho(\xi,\eta))|} \simeq \begin{cases} (|x|r)^{-1} & \text{if } r < |x| \\ r^{-2} & \text{if } r \geq |x| \end{cases}.$$

Now look at the set of η for which the left side above is greater than α. If $\alpha \leq |x|^{-2}$, then this set is just basically where $r^{-2} > \alpha$, i.e. the non-isotropic ball $B(\xi, \alpha^{-1/2})$. Since $\alpha^{-1/2} \geq |x|$, this ball has area $\simeq (\alpha^{-1/2})^3 = \alpha^{-3/2}$, which is what we need (because now $Q = 3$ and we wanted to show that the kernel is uniformly in weak-$L^{\frac{Q}{Q-1}}$). If now $\alpha > |x|^{-2}$, then the desired set is just the set of η where $(|x|r)^{-1} > \alpha$, i.e. the non-isotropic ball $B(\xi, (|x|\alpha)^{-1})$, and since $(|x|\alpha)^{-1} < |x|$, the area of this ball is $\simeq |x|(|x|\alpha)^{-2} \lesssim \alpha^{-3/2}$ as desired. By symmetry we may reverse the role of ξ and η. A similar analysis can be carried out in the general case, and this establishes the theorem.

The second proof is more elementary in nature and consists of a potential estimate for the lifted function. Again it is the most transparent to look at the case $X_1 = \frac{\partial}{\partial x}$, $X_2 = x\frac{\partial}{\partial t}$ on $\mathbb{R}^2$. Recall the lifting of these vector fields to the group $H \simeq \mathbb{R}^3$ via the map

$$\pi\colon H \to \mathbb{R}^2, \quad \pi(x,y,t) = (x,t).$$

We had a basis of left-invariant vector fields of degree 1 on H, namely

$$\tilde{X}_1 = \frac{\partial}{\partial x}, \quad \tilde{X}_2 = \frac{\partial}{\partial y} + x\frac{\partial}{\partial t}.$$

Now given a function $u \in C_c^\infty(\mathbb{R}^2)$. Lift it to H by setting

$$\tilde{u}(x,y,t) = u(x,t).$$

It was known that for functions on H, we have the following pointwise estimate:

$$|\tilde{u}(x,y,t)| \lesssim \int_H \left(|\tilde{X}_1\tilde{u}| + |\tilde{X}_2\tilde{u}|\right)((x,y,t)\cdot(\alpha,\beta,\gamma))\frac{d\alpha d\beta d\gamma}{|(\alpha,\beta,\gamma)|^{\tilde{Q}-1}}, \tag{4}$$

where again

$$|(\alpha,\beta,\gamma)| = |\alpha| + |\beta| + |\gamma|^{\frac{1}{2}}$$

is the norm function, and $\tilde{Q} = 4$. This is just the analog of the Euclidean potential estimate

$$|u(x)| \lesssim \int_{\mathbb{R}^N} |\nabla u|(x+y)\frac{dy}{|y|^{N-1}},$$

which one could prove by applying the fundamental theorem of calculus along straight lines radiating from x, and averaging over all possible directions. A similar construction can be carried out in the group H; a notable feature here is that we are controlling the function by derivatives not in all directions but only in the directions $\tilde{X}_1$ and $\tilde{X}_2$. In other words, we do not have $[\tilde{X}_1, \tilde{X}_2]$ on the right hand side, which is needed to span the tangent space. This is because we can write $[\tilde{X}_1, \tilde{X}_2]$ as $\tilde{X}_1\tilde{X}_2 - \tilde{X}_2\tilde{X}_1$ and integrate by parts once as in the commutator estimate of Hormander, thereby controlling that by $\tilde{X}_1$ and $\tilde{X}_2$ only. This said about formula (4), we shall assume its validity and proceed as follows.

First, observe that (4) reduces to

$$|u(x,t)| \lesssim \int_H (|X_1 u| + |X_2 u|)\,(x+\alpha, t+\gamma+x\beta)\frac{d\alpha d\beta d\gamma}{|(\alpha,\beta,\gamma)|^3}$$

because $(x,y,t)\cdot(\alpha,\beta,\gamma) = (x+\alpha, y+\beta, t+\gamma+x\beta)$ and $(\tilde{X}_j\tilde{u})(x,y,t) = (X_j u)(x,t)$. Writing $\nabla_b u = (X_1 u, X_2 u)$, we get

$$|u(x,t)| \lesssim \int_H |\nabla_b u|(x+\alpha, t+\gamma+x\beta)\frac{d\alpha d\beta d\gamma}{|(\alpha,\beta,\gamma)|^3}.$$

Let now $1<p<Q$ and $\frac{1}{p^*} = \frac{1}{p} - \frac{1}{Q}$. We shall take L^{p^*} norm of both sides in the t and x variables successively. First, taking $L^{p^*}(dt)$ norm, we get

$$\|u\|_{L^{p^*}(dt)}(x) \lesssim \int_{\mathbb{R}^2} \left\| \int_{\mathbb{R}} |\nabla_b u|(x+\alpha, t+\gamma+x\beta)\frac{d\gamma}{|(\alpha,\beta,\gamma)|^3}\right\|_{L^{p^*}(dt)} d\alpha d\beta.$$

In the integral on the right-hand side, we are fixing x, α and β and considering the norm of the integral in γ. If we let $F(t) = |\nabla_b u|(x+\alpha, t+x\beta)$, then this norm is just

$$\left\| \int_{\mathbb{R}} \frac{F(t+\gamma)}{|(\alpha,\beta,\gamma)|^3} d\gamma \right\|_{L^{p^*}(dt)}.$$

But the kernel here, $|(\alpha,\beta,\gamma)|^{-3}$, is not only bounded near $\gamma=0$ (when α,β are not both 0), but also integrable as $\gamma\to\infty$. Hence for almost every (α,β), we can bound this by

$$\left\| \int_{\mathbb{R}} \frac{F(t+\gamma)}{|(\alpha,\beta,\gamma)|^3} d\gamma \right\|_{L^{p^*}(dt)} \lesssim \|F\|_{L^p(dt)} \||(\alpha,\beta,\gamma)|^{-3}\|_{L^r(d\gamma)},$$

where

$$\frac{1}{r} = 1 + \frac{1}{p^*} - \frac{1}{p} = 1 - \frac{1}{Q} = \frac{2}{3}.$$

But then

$$\||(\alpha,\beta,\gamma)|^{-3}\|_{L^r(d\gamma)} \simeq (|\alpha|+|\beta|)^{-3+2\times\frac{2}{3}} = (|\alpha|+|\beta|)^{-\frac{5}{3}}.$$

The remarkable thing in this procedure is that $\|F\|_{L^p(dt)}$ no longer depends on β; in fact

$$\|F\|_{L^p(dt)} = \|\nabla_b u\|_{L^p(dt)}(x+\alpha).$$

Putting these together,

$$\|u\|_{L^{p^*}(dt)}(x) \lesssim \int_{\mathbb{R}^2} \|\nabla_b u\|_{L^p(dt)}(x+\alpha)\frac{d\alpha d\beta}{(|\alpha|+|\beta|)^{\frac{5}{3}}}.$$

Hence we can integrate away the variable β we added in the lifting process, and obtain

$$\|u\|_{L^{p^*}(dt)}(x) \lesssim \int_{\mathbb{R}} \|\nabla_b u\|_{L^p(dt)}(x+\alpha)\frac{d\alpha}{|\alpha|^{\frac{2}{3}}}.$$

If one now invokes fractional integration on $\mathbb{R}$, one gets the desired estimate.

What we have done above is really like a product theory of fractional integral operators, and it is not clear whether one can adapt such an argument to prove weak-type estimates; in fact in product theory we usually cannot do so.

Again the two proofs are actually very closely related to each other; they are similar in spirit in the sense that both of them requires a lifting to a higher dimensional group, and one needs to integrate away the additional variable somewhere in the argument.

Finally, let us end by mentioning that the Sobolev inequality above fails in general at the endpoint $p = Q$. In the next talk, we shall discuss a remedy of this failure, and in the last one we shall discuss some applications of these ideas to several complex variables.

References

[1] C. Carathéodory, *Untersuchungen über die Grundlagen der Thermodynamik*, Math. Ann. **67** (1909), no. 3, 355–386, DOI 10.1007/BF01450409.

[2] Wei-Liang Chow, *Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung*, Math. Ann. **117** (1939), 98–105.

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

[4] Alexander Nagel, Elias M. Stein, and Stephen Wainger, *Balls and metrics defined by vector fields. I. Basic properties*, Acta Math. **155** (1985), no. 1-2, 103–147.

[5] Linda P. Rothschild and Elias M. Stein, *Hypoelliptic differential operators and nilpotent groups*, Acta Math. **137** (1976), no. 3-4, 247–320.

[6] Po-Lam Yung, *Sobolev inequalities for* $(0, q)$ *forms on CR manifolds of finite type*, to appear in Math. Res. Letts.