Representations of the Heisenberg Group and Reproducing Kernels

Mark Hamilton

April 2016

1 Overview

- Heisenberg Group
- Characterizing its unitary IR reps
- Heisenberg Lie Algebra
- Reproducing Kernels
- Applications to Machine Learning

2 The Heisenberg Group and its Representations

Definition 2.1. Heisenberg Group

$$H = \{ \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \mid x, y, z \in \mathbb{R} \}$$

where the identity element is the identity element of matrices, and inverses are the matrix inverses (Exercise: check this). Multiplication is given by:

$$\begin{bmatrix} 1 & x_1 & z_1 \\ 0 & 1 & y_1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & x_2 & z_2 \\ 0 & 1 & y_2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & x_1 + x_2 & z_1 + z_2 + x_1 y_2 \\ 0 & 1 & y_1 + y_2 \\ 0 & 0 & 1 \end{bmatrix}$$

Remark. The center of H is:

$$\mathcal{Z} = \{ \begin{bmatrix} 1 & 0 & z \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mid z \in \mathbb{R} \}$$

Theorem 1. The Heisenberg group has a family of unitary representations $\pi_\hbar, \hbar \in \mathbb{R}\backslash 0$ on $L^2(\mathbb{R})$ given by:

let $f \in L^2(\mathbb{R})$ then $\pi_\hbar\,(g) : L^2(\mathbb{R}) \to L^2(\mathbb{R})$

$$(\pi_\hbar \left(\begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \right) f)\,(t) \equiv e^{ity} e^{i\hbar z} f(t + \hbar x)$$

Proof. First let us show that it is a representation:

$$\pi_\hbar \left(\begin{bmatrix} 1 & x_1 & z_1 \\ 0 & 1 & y_1 \\ 0 & 0 & 1 \end{bmatrix} \right) \pi_\hbar \left(\begin{bmatrix} 1 & x_2 & z_2 \\ 0 & 1 & y_2 \\ 0 & 0 & 1 \end{bmatrix} \right) f(t) =$$

$$= \pi_\hbar \left(\begin{bmatrix} 1 & x_1 & z_1 \\ 0 & 1 & y_1 \\ 0 & 0 & 1 \end{bmatrix} \right) e^{ity_2} e^{i\hbar z_2} f(t + \hbar x_2)$$

$$= e^{it(y_1+y_2)} e^{itx_1y_2} e^{i\hbar(z_1+z_2)} f(t + \hbar x_1 + \hbar x_2)$$

$$= \pi_\hbar \left(\begin{bmatrix} 1 & x_1 + x_2 & z_1 + z_2 + x_1y_2 \\ 0 & 1 & y_1 + y_2 \\ 0 & 0 & 1 \end{bmatrix} \right) f(t)$$

$$= \pi_\hbar \left(\begin{bmatrix} 1 & x_1 & z_1 \\ 0 & 1 & y_1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & x_2 & z_2 \\ 0 & 1 & y_2 \\ 0 & 0 & 1 \end{bmatrix} \right) f(t)$$

Now that we have shown it is a representation we must show that it is unitary given the standard inner product in $L^2(\mathbb{R})$:

$$(f_1, f_2) = \int_{-\infty}^{\infty} f_1(t)\overline{f_2(t)}dt$$

Let us begin by expanding the definition

$$(\pi_\hbar(g)f, \pi_\hbar(g)f) = \int_{-\infty}^{\infty} e^{ity} e^{i\hbar z} f_1(t + \hbar x)\overline{e^{ity} e^{i\hbar z} f_1(t + \hbar x)}dt$$

$$= \int_{-\infty}^{\infty} f_1(t + \hbar x)\overline{f_2(t + \hbar x)}dt = (f_1, f_2)$$

Where the final step comes from moving the translation into the bounds of the integral and noting that it does not effect the integral.

□

3 Heisenberg Lie Algebra

The relation between the Heisenberg group and its lie algebra can be seen by taking the derivative of the one parameter subgroups at the identity yielding the three generating elements:

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, H = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, Q = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

One can then check that the exponential map brings us back into the Heisenberg group:

$$exp\left(\begin{bmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{bmatrix}\right) = I + \begin{bmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 1 & x & z + \frac{xy}{2} \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix}$$

And the commutation relations are as follows:

$$[P, Q] = H$$
$$[H, P] = [H, Q] = 0$$

4 Characterizing the Unitary IR Representations of the Heisenberg Group

We have now seen that there is an infinite family of unitary representations on $L^2(\mathbb{R})$. We will now see that these are irreducible. Moreover, they are the only nontrivial IR reps of the Heisenberg group.

We begin by stating a classic theorem of Stone and Von Neumann. This theorem states that any IR unitary representation with a nontrivial action on the center is equivalent to a $\pi_\hbar$. This theorem can be further generalized to higher dimensional Heisenberg groups but it is unneeded for this discussion.

To simplify the notation we will consider the shorthand true for any representation of the heisenberg group ρ

$$\rho(x, y, z) \equiv \rho\left(\begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix}\right)$$

Theorem 2. Stone-Von Neumann Let ρ be a unitary representation of H on a Hilbert space $\mathcal{H}$ such that $\rho(0, 0, z) = e^{ihz}I$ for some $h \in \mathbb{R}\backslash 0$. Then $\mathcal{H} = \oplus\mathcal{H}_\alpha$, where the $\mathcal{H}_\alpha$'s are mutually orthogonal subspaces of $\mathcal{H}$, each invariant under ρ such that $\rho \mid \mathcal{H}_\alpha$ is unitarily equivalent to $\pi_\hbar$ for each α. In particular is ρ is IR then ρ is equivalent to $\pi_\hbar$

Corollary 2.1. Every irreducible unitary representation of H is unitarily equivalent to one and only one of the following representations:

- $\pi_\hbar(\hbar \in \mathbb{R}\backslash\{0\}$, acting on $L^2(\mathbb{R})$
- $\sigma_{a,b}(x,y,z) = e^{i(ax+by)}(a,b \in \mathbb{R})$ acting on $\mathbb{C}$

Proof. If ρ is an IR unitary representation, Shur's lemma implies that ρ must map the center $\mathcal{Z}$ of H homomorphically into $\{cI \mid |c| = 1, c \in \mathbb{C}\}$ so $\rho(0,0,z) = e^{ihz}I$ for some $h \in \mathbb{R}$. If $h \neq 0$ then it is equivalent to $\pi_\hbar$ by Stone-Von Neumann. If $h = 0$ then ρ factors through the center $H/\mathcal{Z} \simeq \mathbb{C}$. The Irreducible representations of $\mathbb{C}$ are one dimensional and by Shur's lemma they are just $(x,y) \to e^{i(ax+by)}$ □

5 Hermite Polynomials

Now that we have seen the way the Heisenberg group acts on $L^2(\mathbb{R})$, we would like to use this action to discover properties about the space. Our first goal is to decompose this space into sensible basis vectors. Two decompositions stand out immediately, one based on position, and one based on Fourier components. The Hermite polynomials are a mix of the two of them, or more specifically they are proportional to the result of projecting the space basis onto the Fourier basis. To actually calculate these special basis functions one must find a unitary representation of the lie algebra in question. One can check that:

$$a_- = \frac{1}{\sqrt{2}}\left(x + \frac{d}{dx}\right), a_+ = \frac{1}{\sqrt{2}}\left(x - \frac{d}{dx}\right), I$$

do the trick. These operators can also be interpreted as the raising and lowering operators of the quantum harmonic oscillator. To derive the Hermite functions, one play 2 equations off of each other. The first comes from noting that $a_-(f_0) = 0$ where f_0 corresponds to the zeroth Fourier harmonic. The second equation comes from expanding the lowering operator in terms of the Fourier and position basis and setting them equal. Through this process one discovers:

Definition 5.1. Hermite Polynomials

$$H_k(x) = (-1)^k \frac{d^k}{dx^k}(e^{-x^2})e^{x^2}, k \in \mathbb{Z}_{\geq 0}$$

Definition 5.2. Normalized Hermite Functions

$$h_k(x) = (2^k k! \sqrt{\pi})^{-\frac{1}{2}} H_k(x) e^{-\frac{1}{2}x^2}, k \in \mathbb{Z}_{\geq 0}$$

One can also form these polynomials by diagonalizing the operator $P^2 + Q^2$ in the Heisenberg lie algebra. The first Hermite polynomial is a constant, and all others can be derived by applying the raising operator the the previous polynomial. Furthermore, by analyzing the recurrence relations used to derive the polynomials as in [7], one can also show that:

Proposition 3.

$$\int_{-\infty}^{\infty} h_m(x)\overline{h_n(x)}dx = \delta_{mn}$$

Completeness of this set of functions comes from the fact that there is a Hermite polynomial of order n for every n, and the system $(e^{-x^2/2}x^n)_0^\infty$ is closed. For a proof see [6]

6 Reproducing Kernels

Often times, we would like to understand the structure of the inner products within a space like $L^2(\mathbb{R})$. The inner product is key to the geometry of the space, and we shall soon see that we can view the geometry of the space through its dual space. To set the stage we must familiarize ourselves with some results from the theory of dual spaces.

Theorem 4. Riesz Representation Theorem Given a Hilbert space $\mathcal{H}$ and its dual $\mathcal{H}^*$ we can construct a map $\Phi(x) : \mathcal{H} \to \mathcal{H}* = \langle \cdot, x \rangle_{\mathcal{H}}$, which is an isometric isomorphism where:

- Φ is bijective
- $|x| = |\Phi(x)|$
- $\Phi(x + y) = \Phi(x) + \Phi(y)$
- $\Phi(\lambda x) = \overline{\lambda}\Phi(x)$

This theorem shows that the dual space of continuous linear functionals has an identical metric structure to the original Hilbert space. Furthermore in spaces like $L^2(\mathbb{R})$, or other Hilbert spaces of functions on a set, one can show that the evaluation function $ev_z(f) \equiv f(z) \in \mathbb{R}$ is a continuous linear functional $L^2(\mathbb{R}) \to \mathbb{R}$ and is hence in its dual space. This means that in the dual space we can form $\langle ev_z, f \rangle_{\mathcal{H}^*} = f(z)$, and then ask the same question but mapped into the original space. When we bring the evaluation functional back down into the original Hilbert space, we get a reproducing kernel vector, $K_z \in \mathcal{H}$ such that $\langle f, K_z \rangle_{\mathcal{H}} = f(z)$.

With this in mind, we can define a function that captures much of the geometric information about our Hilbert space, $L^2(\mathbb{R})$.

Definition 6.1. Reproducing Kernel

We can define the reproducing kernel on $L^2(\mathbb{R})$ to be the function:

$$K : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, (x, y) \mapsto K_y(x)$$

Note that this map has the so called "reproducing property" where:

$$f(y) = \int_{-\infty}^{\infty} \overline{K(x, y)}f(x)dx$$

also note that by definition of K_x:

$$K(x,y) = \langle K_x, K_y \rangle_{L^2(\mathbb{R})}$$

Proposition 5. If H has a kernel an orthonormal basis $(\psi_n)_{n=0}^{\infty}$,

$$K(x,y) = \sum_{n=0}^{\infty} \psi_n(x)\overline{\psi_n(y)}$$

Proof. Expanding K_y in terms of the basis yields:

$$K_y = \sum_{n=0}^{\infty} \langle K_y, \psi_n(x) \rangle \psi_n$$

By the reproducing property of the kernel the right hand side becomes:

$$K_y = \sum_{n=0}^{\infty} \overline{\psi_n(y)}\psi_n$$

Continuity of pointwise evaluation yields

$$K(x,y) = K_y(x) = \sum_{n=0}^{\infty} \overline{\psi_n(y)}\psi_n(x)$$

□

Specializing to the case of $L^2(\mathbb{R})$, lets construct the kernel explicitly.

Theorem 6. The kernel for the Hilbert space $L^2(\mathbb{R})$ is:

$$K(x,y) = \delta(x-y)$$

Proof. By definition the Dirac delta satisfies the reproducing property:

$$f(y) = \int_{-\infty}^{\infty} \overline{K(x,y)} f(x) dx$$

□

Now consider a slight generalization of the kernel in $L^2(\mathbb{R})$ by considering a weighted sum of the Hermite functions. One can show that as long as the coefficients are in $l_2(\mathbb{R})$ this new weighted sum is indeed a kernel, but in a different Hilbert space. In fact this is often how new kernels are constructed, by multiplying the basis functions in $L^2(\mathbb{R})$ by special sequences in $l_2(\mathbb{R})$.

Definition 6.2. The Gaussian kernel can be represented as:

$$K(x,y) = exp\left(\frac{-\|x-y\|_2^2}{2\sigma^2}\right) = \sum_{n=0}^{\infty} C * b^n e_n(x)\overline{e_n(y)}$$

where

$$e_n(x) = exp\left(-(c-a)x^2)H_k(x\sqrt{2c}\right)$$

and a, b, c are functions of σ [4]

7 Applications to Machine Learning

Often we are in the case where we would like to discover patterns in data. Most of the time, these patterns are complicated, nonlinear, and not handled well by linear regression. Fortunately kernel methods allow us to transform our space into a nonlinear space by using the nonlinear space's associated kernel. To see this we must investigate linear regression in its primal and associated dual form.

7.1 Standard Form of Regression

In its primal form linear regression can be summarized as:

Given a data matrix $X \in \mathbb{R}^{N \times d}$ where N is the number of d-dimensional observations we want to learn a function

$$\hat{f}(x) = \sum_{i=1}^{d} x_i w_i = \langle x, w \rangle$$

where $x \in \mathbb{R}^d$.

To learn this function we must estimate w by finding the solution with the lowest squared error over the whole dataset:

$$\hat{w} = argmin_w(\|y - Xw\|^2)$$

7.2 Dual Form of Regression

One can also show that the best w must lie in the span of the data in $\mathbb{R}^d$. We can now reformulate linear regression using a dual viewpoint.

Given a data matrix $X \in \mathbb{R}^{N \times d}$ where N is the number of d-dimensional observations we want to learn a function

$$\hat{f}(x) = \sum_{i=1}^{N} \alpha_i \langle x, x_i \rangle$$

Where:

$$w = \sum_{i=1}^{N} \alpha_i x_i$$

7.3 Kernel Regression

Note that the problem is completely defined by the inner products of the data in $\mathbb{R}^d$. This gives us the powerful opportunity to substitute the euclidean inner products with other inner products. Using the kernel as an alternative inner product, we can now perform regression in incredibly high, and potentially infinite, dimensional Hilbert spaces without having to map the points into the infinite dimensional space first. The kernel does the heavy lifting and allows us

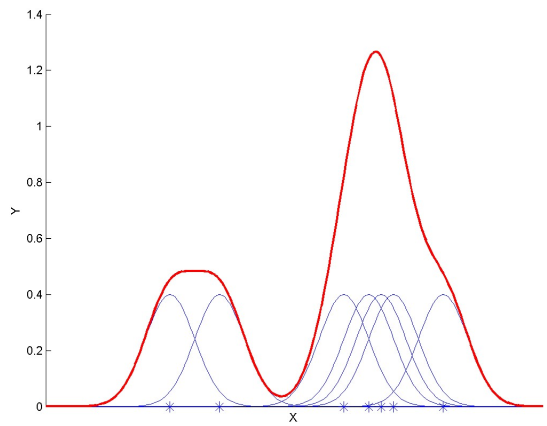

Figure 1: Gaussian kernel function estimation and components of the sum

to regress in these spaces because it captures all the metric information of the ambient space.

The kernel we have investigated is called the Gaussian kernel and allows one to regress a function by placing Gaussian bumps at each of the data-points to reconstruct the function as seen in figure 1.

The idea of transforming a linear algorithm into a nonlinear algorithm using kernels is applicable to almost any statistical technique including regression, classification, clustering, principle component analysis, and hypothesis testing.

References

[1] Nello Cristianini and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods.* Cambridge university press, 2000.

[2] Gerald B Folland. *Harmonic analysis in phase space.* Number 122. Princeton university press, 1989.

[3] Bruce Gilligan and Guy Roos. *Symmetries in Complex Analysis: Workshop on Several Complex Variables, Analysis on Complex Lie Groups, and Homogeneous Spaces, October 17-29, 2005, Zhejiang University, Hangzhou, PR China*, volume 468. American Mathematical Soc., 2008.

[4] Arthur Gretton. Introduction to rkhs, and some simple kernel algorithms.

[5] Thomas Hofmann, Bernhard Schölkopf, and Alexander J Smola. Kernel methods in machine learning. *The annals of statistics*, pages 1171–1220, 2008.

[6] Gabor Szeg. *Orthogonal polynomials*, volume 23. American Mathematical Soc., 1939.

[7] Sundaram Thangavelu. *Lectures on Hermite and Laguerre expansions*, volume 42. Princeton University Press, 1993.