Canad. J. Math. Vol. **56** (3), 2004 pp. 566–589

Geodesics in a Manifold with Heisenberg Group as Boundary

Yilong Ni

Abstract. The Heisenberg group is considered as the boundary of a manifold. A class of hypersurfaces in this manifold can be regarded as copies of the Heisenberg group. The properties of geodesics in the interior and on the hypersurfaces are worked out in detail. These properties are strongly related to those of the Heisenberg group.

1 Introduction

The Heisenberg group H_1 is the simplest non-commutative nilpotent Lie group. In this group, we have a natural analogue of the Laplacian, Δ_H, which is a sum of squares of two horizontal vector fields. The geometry associated to this sub-elliptic operator Δ_H, known as sub-Riemannian geometry, was introduced in [4]. Sub-Riemannian metric, which is also called *Carnot-Carathéodory* metric, is defined as the infimum of the length among all horizontal curves that join two points. As in the Riemannian case, we may define geodesics as locally shortest curves. The fact that the entire axis $x = 0$ is a line of conjugate points relative to the origin (see [1]) is quite different from the Riemannian case, in which, for any two points close enough to each other, there exists a unique shortest geodesic connecting them. Therefore the Heisenberg group serves as a model for the study of similarities and dissimilarities between sub-Riemannian geometry and Riemannian geometry.

The purpose of this article is to consider the Heisenberg group H_1 with its subelliptic Laplacian Δ_H as the limit of a family of Riemannian manifolds. We construct a manifold and identify H_1 with its boundary. With a group action, this manifold, is isomorphic to the direct product $H_1 \times \mathbf{R}^+$. For each $u > 0$, we endow the hypersurface $H_1 \times \{u\}$ with a Riemannian metric g_u, which degenerates to the usual sub-Riemannian metric g_0 of the Heisenberg group when $u \to 0$. Therefore each hypersurface can be regarded as an approximation of the Heisenberg group. We will choose the g_u carefully so that the metric of the interior coincides nicely with the sub-Riemannian metric of the Heisenberg group as boundary. We study the properties of geodesics in the interior as well as geodesics on the hypersurfaces, and show their relations with those of the Heisenberg group. We will show that geodesics that leave the boundary and return to the boundary have the same length as the boundary geodesics that have the same endpoints.

Also we know that in the Heisenberg group H_1, we have infinitely many geodesics connecting the origin and $(0, t)$, $\forall t \neq 0$. On the hypersurface $H_1 \times \{u\}$, this is not

Received by the editors June 26, 2002.
AMS subject classification: 53C22, 53C17.
Keywords: Heisenberg group, Hamiltonian mechanics, geodesic.

https://doi.org/10.4153/CJM-2004-026-6 Published online by Cambridge University Press

the case. When $|t| < 2\pi u$, there is a unique geodesic connecting the origin and $(0, t)$. However for $|t| \geq 2\pi u$, there will be infinitely many geodesics connecting these two points. We can see that this property reduces perfectly to the Heisenberg case as $u \to 0$.

The paper is organized as follows. In Section 2 we go over some basic facts of H_1 and construct the Riemannian manifold with Heisenberg group H_1 as boundary. In Section 3 Hamiltonian mechanics is used to study geodesics in the interior. We show that for any two points in the interior there is a unique shortest geodesic joining them. In Section 4 we obtain some properties of the geodesics on the hypersurface. These properties are compared with those of the Heisenberg group.

2 Heisenberg Group as Boundary

The 3-dimensional Heisenberg group H_1 can be coordinatized as $R^3 = (x_1, x_2, t) = (x, t)$, with group law:

$$(x, t) \circ (x', t') = (x + x', t + t' + 2ax_2x_1' - 2ax_1x_2'), \tag{1}$$

where a is a positive real parameter. The vector fields

$$X_1 = \frac{\partial}{\partial x_1} + 2ax_2\frac{\partial}{\partial t}, \quad X_2 = \frac{\partial}{\partial x_2} - 2ax_1\frac{\partial}{\partial t}, \quad T = \frac{\partial}{\partial t}. \tag{2}$$

are left invariant and generate the Lie algebra of H_1. The Lie algebra relations are

$$[X_1, X_2] = -4aT, \quad [X_1, T] = [X_2, T] = 0. \tag{3}$$

Since the vector fields X_1, X_2 Lie-generate the tangent space of H_1, Chow's Theorem tells us that any two points can be joined by a horizontal curve. The *Carnot-Carathéodory* metric is obtained by taking the infimum of the length among all horizontal curves that join two points. This metric was introduced and studied by Gaveau [3], [4]. The Heisenberg (sub-)Laplacian is the left-invariant subelliptic operator

$$\Delta_H = \frac{1}{2}(X_1^2 + X_2^2). \tag{4}$$

Geodesics, which are locally shortest curves, can be obtained as the traces of Hamiltonian paths([5]). The Hamiltonian is the symbol of Δ_H.

Consider H_1 as a subset of $\mathbf{C}^2 = \{(z, w)\}$. Introduce a group operation in $\mathbf{C}^2$ by

$$(z, w) \circ (z', w') = (z + z', w + w' + 2ia\bar{z}z'). \tag{5}$$

Use also real coordinates x_1, x_2, y_1, y_2, with

$$z = x_1 + ix_2, \quad w = y_1 + iy_2.$$

Introduce the functions

$$t = y_1, \quad u = u(z, w) = y_2 - az\bar{z}.$$

https://doi.org/10.4153/CJM-2004-026-6 Published online by Cambridge University Press

Using the coordinate $(x,t,u)=(x_1,x_2,t,u)$ the group law is

(6) $$(x,t,u)\circ(x',t',u')=\big(x+x',t+t'+2a(x_2x_1'-x_1x_2'),u+u'\big).$$

Since $u\colon \mathbf{C}^2\to(\mathbf{R},+)$ is a group homomorphism our group is isomorphic to the direct product $H_1\times\mathbf{R}$. The corresponding Lie algebra is generated by the left-invariant vector fields

$$X_1,\quad X_2,\quad T,\quad U=\frac{\partial}{\partial u}.$$

Consider the complex vector fields

$$\frac{\partial}{\partial z}=\frac{1}{2}\Big(\frac{\partial}{\partial x_1}-i\frac{\partial}{\partial x_2}\Big),\quad \frac{\partial}{\partial w}=\frac{1}{2}\Big(\frac{\partial}{\partial y_1}-i\frac{\partial}{\partial y_2}\Big),$$
$$Z=\frac{\partial}{\partial z}+2ia\bar{z}\frac{\partial}{\partial w},\quad W=\frac{\partial}{\partial w},$$

and their conjugates. The Siegel domain

$$\mathbf{C}^2_+=\{\operatorname{Im} w>az\bar{z}\}=\{u>0\}$$

is a sub-semigroup of $\mathbf{C}^2$ and if we identify H_1 with $\{u=0\}$, the boundary of $\mathbf{C}^2_+$, then H_1 is a subgroup of $\mathbf{C}^2$ that acts on $\mathbf{C}^2_+$ by left and right translations. For any choice of $b>0$ the operator

$$L=Z\bar{Z}+\bar{Z}Z+bu(W\bar{W}+\bar{W}W)+\frac{b}{2}U=\frac{1}{2}(X_1^2+X_2^2)+\frac{bu}{2}(T^2+U^2)+\frac{b}{2}U$$

is elliptic in $\mathbf{C}^2_+$, self-adjoint in $L^2(\mathbf{C}^2_+)$, and invariant with respect to the H_1 action.

For each $u>0$, the hypersurface $H_1\times\{u\}$ is invariant with respect to the H_1 action. The restriction of L to this hypersurface is given by

$$L=\frac{1}{2}(X_1^2+X_2^2)+\frac{bu}{2}T^2.$$

It degenerates to the Heisenberg sublapacian Δ_H as $u\to 0$. We then study the behavior of geodesics in the "interior", $\mathbf{C}^2_+$, associated to L, on the hypersurfaces $\{u\equiv u_0\}$, and their relations to the geodesics on the boundary H_1.

3 Geodesics in the interior

As in [2], take variables (x,t,u) and dual variables (ξ,θ,σ). Let

$$\Lambda=\begin{pmatrix}0 & 2a\\ -2a & 0\end{pmatrix};\quad \zeta=\zeta(x,\theta)=\xi+\theta\Lambda x.$$

We take as Hamiltonian the principal symbol of L

(7) $$H=\frac{1}{2}\langle\zeta,\zeta\rangle+\frac{bu}{2}(\theta^2+\sigma^2).$$

Hamilton's equations for a curve $\big(x(s), t(s), u(s), \sigma(s), \zeta(s), \theta(s)\big)$ can be organized as

$$
\begin{aligned}
&\dot{x} = \zeta; && \dot{\zeta} = 2\theta\Lambda\zeta; \\
&\dot{t} = \langle \zeta, \Lambda x\rangle + bu\theta; && \dot{\theta} = 0; \\
&\dot{u} = bu\sigma; && \dot{\sigma} = -\frac{b}{2}(\theta^2 + \sigma^2).
\end{aligned} \tag{8}
$$

We start with the last two equations, keeping in mind that θ is constant. We have

$$\frac{d}{ds}\left(\frac{\sigma}{\theta}\right) = -\frac{b\theta}{2}\left(1 + \left(\frac{\sigma}{\theta}\right)^2\right),$$

so for some choice of phase ω,

$$\sigma(s) = -\theta \tan\left(\omega + \frac{1}{2}b\theta s\right). \tag{9}$$

Then

$$\frac{d}{ds}\log u = b\sigma = -b\theta \tan\left(\omega + \frac{1}{2}b\theta s\right) = 2\frac{d}{ds}\log\cos\left(\omega + \frac{1}{2}b\theta s\right).$$

Therefore we get

$$u(s) = u_0 \cos^2\left(\omega + \frac{1}{2}b\theta s\right). \tag{10}$$

First we consider a normalized geodesic which starts from a boundary point and also makes its first return to a boundary point. Because of the invariance under the H_1 action, we take the origin as our starting point on the boundary. So we have

$$u\left(-\frac{1}{2}\right) = u\left(\frac{1}{2}\right) = 0; \quad u(s) > 0,\ -\frac{1}{2} < s < \frac{1}{2}.$$

Choosing $\omega = 0$ in (10) we have $u_0 = u(0)$, $\frac{1}{4}b\theta = \pm\frac{\pi}{2}$, or

$$b\theta = \pm 2\pi. \tag{11}$$

Integrating (8) and using the boundary conditions $x(\frac{1}{2}) = x$, $x(-\frac{1}{2}) = 0$, we obtain

$$
\begin{aligned}
\zeta(s) &= \exp(2s\theta\Lambda)\zeta(0); \\
x(s) &= (2\theta\Lambda)^{-1}\big(\exp(2s\theta\Lambda) - \exp(-\theta\Lambda)\big)\zeta(0); \\
x = x\left(\frac{1}{2}\right) &= (2\theta\Lambda)^{-1}\sinh(\theta\Lambda)\zeta(0) = \frac{\sin(2a\theta)}{2a\theta}\zeta(0).
\end{aligned} \tag{12}
$$

Because Λ is skew symmetric, the first equation of (8) implies that $\langle\zeta,\zeta\rangle$ is constant along the curve. Then from the second equation of (8) we have

$$\begin{aligned}(13)\quad \dot{t} &= \frac{1}{2\theta}\Big\langle \zeta, \zeta - \zeta\Big(-\frac{1}{2}\Big)\Big\rangle + b\theta u(0)\cos^2\Big(\omega + \frac{1}{2}b\theta s\Big)\\ &= \frac{1}{2\theta}\Big\langle \zeta\Big(-\frac{1}{2}\Big), \zeta\Big(-\frac{1}{2}\Big)\Big\rangle - \frac{1}{2\theta}\Big\langle \dot{x}, \zeta\Big(-\frac{1}{2}\Big)\Big\rangle + \frac{b\theta u(0)}{2}\big(1+\cos(b\theta s)\big).\end{aligned}$$

Integrating, and using the boundary conditions $t(\frac{1}{2}) = t$, $t(-\frac{1}{2}) = 0$, we obtain, because of (11),

$$t = \frac{1}{2\theta}\Big|\zeta\Big(-\frac{1}{2}\Big)\Big|^2 - \frac{1}{2\theta}\Big\langle x, \zeta\Big(-\frac{1}{2}\Big)\Big\rangle + \frac{b\theta u(0)}{2}$$

Consider curves for which $\zeta \equiv 0$, so that $x \equiv 0$, and $b\theta u(0) = t$. Then the action is

$$S = H\big(x(0), t(0), u(0), \xi(0), \theta(0), \sigma(0)\big) = \frac{1}{2}|\zeta(0)|^2 + \frac{bu(0)}{2}\big(\theta^2 + \sigma^2(0)\big) = t\theta$$

We note that for the boundary situation, $S_0 = \pi|t|/2a$. If we take $\theta = \operatorname{sgn}(t)\pi/2a$, then these two coincide with each other. Combining this with (11) we get

$$\theta = \operatorname{sgn}(t)\frac{2\pi}{b} = \operatorname{sgn}(t)\frac{\pi}{a}, \quad b = 4a.$$

With this choice of b, θ and ζ, geodesics that leave the origin necessarily return to the boundary at the t-axis, and have the same length as the boundary geodesics that have the same endpoints. Condition $b = 4a$ ties the metric in the interior and that of the boundary together. We then assume $b = 4a$ throughout this paper. We have the following theorem:

Theorem 1 *The geodesics that start from the origin and make their first return to a boundary point $(x, t, 0)$ necessarily return to the boundary at the t-axis, i.e., $x = 0$, and have the same length d_1, where*

$$(d_1)^2 = \frac{\pi|t|}{a}.$$

All the geodesics with this property are parametrized by a part of a paraboloid in $\mathbf{R}^3$.

Proof We now only need to consider geodesics for which $\zeta \neq 0$. $b = 4a$ and $b\theta = 2\pi\operatorname{sgn}(t)$ imply that $2a\theta = \pi\operatorname{sgn}(t)$. $x = x(\frac{1}{2}) = \frac{\sin(2a\theta)}{2a\theta}\zeta(0) = 0$. So the endpoint is on the t-axis. Also

$$(14)\quad \begin{aligned} t &= \frac{1}{2\theta}\Big|\zeta\Big(-\frac{1}{2}\Big)\Big|^2 - \frac{1}{2\theta}\Big\langle x, \zeta\Big(-\frac{1}{2}\Big)\Big\rangle + \frac{b\theta u(0)}{2} = \frac{1}{2\theta}|\zeta(0)|^2 + \frac{b\theta u(0)}{2}\\ S &= \frac{1}{2}|\zeta(0)|^2 + \frac{bu(0)}{2}\big(\theta^2 + \sigma^2(0)\big) = \frac{1}{2}|\zeta(0)|^2 + \frac{b\theta^2 u(0)}{2} = t\theta = \frac{\pi|t|}{2a}\end{aligned}$$

This shows that they have the same length, which satisfies

$$(d_1)^2 = 2S = \frac{\pi|t|}{a}.$$

Furthermore, the geodesics may be parametrized by $\big(\zeta(0), u(0)\big)$. From (14), $\big(\zeta(0), u(0)\big)$ satisfies $|\zeta(0)|^2 + \pi^2 u(0)/a = \pi|t|/a$, $u(0) > 0$, which is part of a paraboloid in $\mathbf{R}^3$. ■

Second, we consider all the normalized geodesics that start from the origin and return to the boundary(they may hit the boundary many times). We have:

Theorem 2 *The geodesics that join the origin to a point $(0, t, 0)$ have lengths $d_1, d_2, d_3, \dots$, where $d_n^2 = n\pi|t|/a$. For each length d_n, the geodesics of that length hit the boundary $n+1$ times (including the end points) and are parametrized by $\big(\zeta(0), u(0)\big)$, satisfying $|\zeta(0)|^2 + \pi^2 u(0)/a = n\pi|t|/2a$, which is part of a paraboloid in $\mathbf{R}^3$.*

Proof The boundary conditions are

$$u\Big(-\frac{1}{2}\Big) = u\Big(\frac{1}{2}\Big) = 0; \quad u(s) \geq 0,\ -\frac{1}{2} < s < \frac{1}{2}.$$

Choosing $\omega = 0$ in (10) we have $\frac{1}{4} b\theta_{2m+1} = \operatorname{sgn}(t)(\frac{1}{2}\pi + m\pi)$, $m = 0, 1, 2, \dots$ $(n = 2m+1)$. $b = 4a$ implies $2a\theta_{2m+1} = (1+2m)\pi \operatorname{sgn}(t)$. $x = x(\frac{1}{2}) = \frac{\sin(2a\theta)}{2a\theta}\zeta(0) = 0$.

$$\begin{aligned} t &= \frac{1}{2\theta_{2m+1}}\Big|\zeta\Big(-\frac{1}{2}\Big)\Big|^2 - \frac{1}{2\theta_{2m+1}}\Big\langle x, \zeta\Big(-\frac{1}{2}\Big)\Big\rangle + \frac{b\theta_{2m+1}u(0)}{2} \\ &= \frac{1}{2\theta_{2m+1}}|\zeta(0)|^2 + \frac{b\theta_{2m+1}u(0)}{2}. \end{aligned}$$

$$\begin{aligned} S_{2m+1} = H_{2m+1}(0) &= \frac{1}{2}|\zeta(0)|^2 + \frac{bu(0)}{2}\big(\theta_{2m+1}^2 + \sigma^2(0)\big) \\ &= \frac{1}{2}|\zeta(0)|^2 + \frac{b\theta_{2m+1}^2 u(0)}{2} = t\theta_{2m+1} = \frac{\pi|t|}{2a}(1+2m). \end{aligned}$$

Therefore they have the same length, which satisfies $d_{2m+1}^2 = (2m+1)\pi|t|/a$. For each m, the geodesics hit the boundary $2m+2$ times(including the end points), and may be parametrized by $\big(\zeta(0), u(0)\big)$, satisfying $|\zeta(0)|^2 + \pi^2 u(0)/a = (2m+1)\pi|t|/2a$, which is again part of a paraboloid in $\mathbf{R}^3$.

Choosing $\omega = \pi/2$ in (9) and (10) we obtain

$$\sigma(s) = \theta \cot\Big(\frac{1}{2}b\theta s\Big);$$

$$u(s) = u(0)\sin^2\Big(\frac{1}{2}b\theta s\Big).$$

With the same procedure as above, we get

$$\frac{1}{4}b\theta_{2m} = \operatorname{sgn}(t)m\pi, m = 1, 2, 3, \dots (n = 2m).$$

Note there that $m \neq 0$. For if $m = 0$, then $u(s) \equiv 0$, that means the geodesic lies on the boundary. $b = 4a$ implies $2a\theta_{2m} = 2m\pi \operatorname{sgn}(t)$. $x = x(\frac{1}{2}) = \frac{\sin(2a\theta)}{2a\theta}\zeta(0) = 0$.

$$\begin{aligned} t &= \frac{1}{2\theta_{2m}}\left|\zeta\left(-\frac{1}{2}\right)\right|^2 - \frac{1}{2\theta_{2m}}\left\langle x, \zeta\left(-\frac{1}{2}\right)\right\rangle + \frac{b\theta_{2m}u(0)}{2} \\ &= \frac{1}{2\theta_{2m}}|\zeta(0)|^2 + \frac{b\theta_{2m}u(0)}{2}. \end{aligned}$$

$$\begin{aligned} S_{2m} = H_{2m}(0) &= \frac{1}{2}|\zeta(0)|^2 + \frac{bu(0)}{2}\big(\theta_{2m}^2 + \sigma^2(0)\big) \\ &= \frac{1}{2}|\zeta(0)|^2 + \frac{b\theta_{2m}^2 u(0)}{2} = t\theta_{2m} = \frac{\pi|t|}{2a}2m. \end{aligned}$$

Therefore they have the same length, which satisfies $d_{2m}^2 = 2m\pi|t|/a$. For each m, the geodesics hit the boundary $2m + 1$ times(including the end points), and may be parametrized by $\big(\zeta(0), u(0)\big)$, satisfying $|\zeta(0)|^2 + \pi^2 u(0)/a = 2m\pi|t|/2a$, which is still a part of a paraboloid in $\mathbf{R}^3$. This completes the proof. ■

Remark This result is very similar to the result on the boundary. From Theorem 1.41 in [2], we know that in H_1 the geodesics that join the origin to a point $(0, t)$ have lengths $d_1, d_2, d_3, \dots$, where $d_n^2 = n\pi|t|/a$. For each length d_n, the geodesics of that length are parametrized by the circle S^1, which is the boundary of the above paraboloid.

Next, we consider the geodesics that start from an arbitrary point in the interior. Because of the invariance under the H_1 action, we can take the starting point to be $(0, 0, u^0)$. Therefore we need to find the Hamiltonian curves with the following boundary conditions:

$$x(0) = 0;\ x(1) = x;\ t(0) = 0;\ t(1) = t;\ u(0) = u^0;\ u(1) = u. \tag{15}$$

We use (10), with $b = 4a$ and a different phase shift, so

$$u(s) = u_0 \sin^2(\omega + 2a\theta s), \quad 0 \le s \le 1,\ 0 \le \omega < \pi, \tag{16}$$

and,

$$\sigma(s) = \frac{\dot{u}}{bu} = \theta\cot(\omega + 2a\theta s).$$

We may assume that $t \ge 0$, otherwise we only need to change $t \to -t$.

From the boundary conditions (15), integrating (8) we have

$$\zeta(s) = \exp(2s\theta\Lambda)\zeta(0);$$

$$x(s) = (2\theta\Lambda)^{-1}\big(\exp(2s\theta\Lambda) - I\big)\zeta(0);$$

$$x = x(1) = (\theta\Lambda)^{-1}\exp(\theta\Lambda)\sinh(\theta\Lambda)\zeta(0).$$

so

$$\begin{aligned} |\zeta(0)|^2 &= \Big|\theta\Lambda\big(\sinh(\theta\Lambda)\big)^{-1}x\Big|^2 \\ &= \frac{(2a\theta)^2}{\sin^2(2a\theta)}|x|^2, \end{aligned}$$

and

$$\begin{aligned} \langle x, \zeta(0)\rangle &= \Big\langle x, 2\theta\Lambda\big(\exp(2\theta\Lambda) - I\big)^{-1}x\Big\rangle \\ &= \langle x, \theta\Lambda\exp(\theta\Lambda)\sinh^{-1}(\theta\Lambda)x\rangle \\ &= \langle x, \theta\Lambda\coth(\theta\Lambda)x\rangle \\ &= 2a\theta\cot(2a\theta)|x|^2. \end{aligned}$$

Notice that $\langle\zeta,\zeta\rangle$ is constant, we then have:

$$\begin{aligned} \dot{t} &= \frac{1}{2\theta}\langle\zeta, \zeta - \zeta(0)\rangle + b\theta u(0)\sin^2(\omega + 2a\theta s) \\ &= \frac{1}{2\theta}\langle\zeta(0), \zeta(0)\rangle - \frac{1}{2\theta}\langle\dot{x}, \zeta(0)\rangle + 4a\theta u_0\sin^2(\omega + 2a\theta s). \end{aligned}$$

Integrating, and using the boundary condition (15), we obtain:

$$\begin{aligned} t &= \frac{1}{2\theta}\langle\zeta(0), \zeta(0)\rangle - \frac{1}{2\theta}\langle x, \zeta(0)\rangle + 4a\theta u_0\int_0^1 \sin^2(\omega + 2a\theta s)\,ds \\ &= \Big(\frac{2a^2\theta}{\sin^2(2a\theta)} - a\cot(2a\theta)\Big)|x|^2 + 2a\theta u_0\Big(1 - \frac{\sin(2\omega + 4a\theta s)}{4a\theta}\Big|_0^1\Big) \\ &= a\mu(2a\theta)|x|^2 + 2a\theta u_0\Big(1 - \frac{\sin(2\omega + 4a\theta s)}{4a\theta}\Big|_0^1\Big), \end{aligned} \tag{17}$$

where

$$\mu(\varphi) = \frac{\varphi}{\sin^2\varphi} - \cot\varphi.$$

Now return to (16). Set

$$\alpha = \exp(i2a\theta),\ y = \exp(i\omega),\ \lambda = \Big(\frac{u^0}{u}\Big)^{1/2}.$$

$$\begin{aligned} \lambda &= \Big(\frac{u^0}{u}\Big)^{1/2} = \Big(\frac{u_0\sin^2(\omega + 2a\theta)}{u_0\sin^2(\omega)}\Big)^{1/2} = -\delta\frac{\sin(\omega + 2a\theta)}{\sin(\omega)} \\ &= \frac{\alpha y - (\alpha y)^{-1}}{y - y^{-1}} = \frac{\alpha y^2 - \alpha^{-1}}{y^2 - 1}, \end{aligned}$$

where $\delta = 1$, if $\omega + 2a\theta > \pi$; $\delta = -1$, if $\omega + 2a\theta \le \pi$. Therefore,

$$y^2 = \frac{\delta\lambda + \alpha^{-1}}{\delta\lambda + \alpha},$$

and

$$\begin{aligned} u^0 &= u_0 \sin^2(\omega) = \frac{u_0}{2}\big(1 - \cos(2\omega)\big) \\ &= \frac{u_0}{4}\big(2 - \exp(i2\omega) - \exp(-i2\omega)\big) \\ &= \frac{u_0}{4}\frac{2 - \alpha^2 - \alpha^{-2}}{\lambda^2 + \delta\lambda(\alpha + \alpha^{-1} + 1)} \\ &= \frac{\sin^2(2a\theta)}{\lambda + 2\delta\lambda\cos(2a\theta) + 1}. \end{aligned}$$

So,

$$u_0 = \frac{u + 2\delta\sqrt{uu^0}\cos(2a\theta) + u^0}{\sin^2(2a\theta)}. \tag{18}$$

In terms of the quantities α, y, and λ,

$$\begin{aligned} \sin(2\omega + 4a\theta s)\big|_0^1 &= \frac{1}{2i}\big((\alpha y)^2 - (\alpha y)^{-2} - y^2 + y^{-2}\big) \\ &= \frac{1}{2i}\left(\frac{(\alpha^2 - 1)(\delta\lambda + \alpha^{-1})}{\delta\lambda + \alpha} + \frac{(1 - \alpha^{-2})(\delta\lambda + \alpha)}{\delta\lambda + \alpha^{-1}}\right) \\ &= \frac{1}{2i}\frac{(\alpha - \alpha^{-1})\big((\lambda^2 + 1)(\alpha + \alpha^{-1}) + 4\delta\lambda\big)}{(\delta\lambda + \alpha)(\delta\lambda + \alpha^{-1})} \\ &= 2\sin(2a\theta)\frac{(u + u^0)\cos(2a\theta) + 2\delta\sqrt{uu^0}}{u + 2\delta\sqrt{uu^0}\cos(2a\theta) + u^0}. \end{aligned}$$

Substitute this equation in (17); we obtain:

$$\begin{aligned} t &= a\mu(2a\theta)|x|^2 + 2a\theta u_0\left(1 - \frac{\sin(2a\theta)}{2a\theta}\frac{(u + u^0)\cos(2a\theta) + 2\delta\sqrt{uu^0}}{u + 2\delta\sqrt{uu^0}\cos(2a\theta) + u^0}\right) \\ &= a\mu(2a\theta)|x|^2 + (u + u^0)\mu(2a\theta) + 2\left(\frac{2a\theta\cos(2a\theta)}{\sin^2(2a\theta)} - \frac{1}{\sin(2a\theta)}\right)\delta\sqrt{uu^0}. \end{aligned} \tag{19}$$

The associated classical action is

$$
\begin{aligned}
S(x,t,u;0,0,u^0;\theta) &= \frac{1}{2}|\zeta(0)|^2 + \frac{bu(0)}{2}\big(\theta^2 + \sigma^2(0)\big) \\
&= \frac{(2a\theta)^2|x|^2}{2\sin^2(2a\theta)} + \frac{b}{2}u_0 \sin^2\omega(\theta^2 + \theta^2\cot^2\omega) \\
&= \frac{(2a\theta)^2|x|^2}{2\sin^2(2a\theta)} + 2a\theta^2 \frac{u + 2\delta\sqrt{uu^0}\cos(2a\theta) + u^0}{\sin^2(2a\theta)} \\
&= \frac{2a\theta^2}{\sin^2(2a\theta)}\big(a|x|^2 + u + 2\delta\sqrt{uu^0}\cos(2a\theta) + u^0\big),
\end{aligned} \tag{20}
$$

where θ is determined from (19).

Let $D = a|x|^2 + u + u^0$, $E = 2\delta\sqrt{uu^0}$, and $\varphi = 2a\theta$, then (19) and (20) can be rewritten as

$$
\begin{aligned}
t &= D\mu(\varphi) + E\left(\frac{\varphi\cos\varphi}{\sin^2\varphi} - \frac{1}{\sin\varphi}\right) \\
2aS &= \frac{\varphi^2}{\sin^2\varphi}(D + E\cos\varphi)
\end{aligned} \tag{21}
$$

We denote the right hand side of (19) by $F(\varphi)$. The following lemma gives us some information on the behavior of function $F(\varphi)$.

Lemma 1 *When $|x| \neq 0$ or $u \neq u^0$,* i.e., *$D > |E|$, $F(\varphi)$ is a increasing diffeomorphism of the interval $(-\pi,\pi)$ onto* $\mathbf{R}$*. On each interval $(m\pi,(m+1)\pi)$, $m = 1,2,\dots$, it has a unique critical point φ_m. On this interval it decreases strictly from $+\infty$ to its value at φ_m and then increases strictly to $+\infty$. Also the values $F(\varphi_m)$ are increasing and goes to $+\infty$ as $m \to +\infty$.*

Proof We take the second derivative of $F(\varphi)$ with respect to φ:

$$
\begin{aligned}
\frac{d^2F}{d\varphi^2} &= \frac{1}{4\sin^4\varphi}\Big(D\big(16\varphi + 8\varphi\cos(2\varphi) - 12\sin(2\varphi)\big) \\
&\qquad + E\big(23\varphi\cos\varphi + \varphi\cos(3\varphi) - 15\sin\varphi - 3\sin(3\varphi)\big)\Big) \\
&\equiv \frac{1}{4\sin^4\varphi}\big(D\cdot g(\varphi) + E\cdot h(\varphi)\big).
\end{aligned} \tag{22}
$$

Notice that

$$
\begin{aligned}
g(\varphi) + h(\varphi) &= 4(1+\cos\varphi)^2\big(\varphi(2+\cos\varphi) + 3\sin\varphi\big) \ \geq 0, \\
g(\varphi) - h(\varphi) &= 4(1-\cos\varphi)^2\big(\varphi(2-\cos\varphi) + 3\sin\varphi\big) \ \geq 0,
\end{aligned} \tag{23}
$$

for any $\varphi \in [0, +\infty)$, and $D \geq |E|$, we have

$$\frac{d^2}{d\varphi^2}F(\varphi) > 0, \quad \varphi \neq m\pi.$$

$\forall \varphi \in \big(m\pi, (m+1)\pi\big)$, we have

$$\begin{aligned} F\big(2(m+1)\pi - \varphi\big) - F(\varphi) &= D\left(\frac{2(m+1)\pi - 2\varphi}{\sin^2\varphi} + 2\frac{\cos\varphi}{\sin\varphi}\right) \\ &\quad + E\left(\frac{2(m+1)\pi - 2\varphi}{\sin^2\varphi}\cos\varphi + \frac{2}{\sin\varphi}\right) \\ &\equiv D\cdot A + E\cdot B \end{aligned}$$

And

$$A + B = 2(1+\cos\varphi)\frac{(m+1)\pi + \varphi}{\sin^2\varphi} + 2\frac{1+\cos\varphi}{\sin\varphi} > 0$$

$$\begin{aligned} A - B &= 2(1+\cos\varphi)\frac{(m+1)\pi - \varphi}{\sin^2\varphi} + 2\frac{\cos\varphi - 1}{\sin\varphi} \\ &= \frac{2}{\sin^2\varphi}(1+\cos\varphi)\big((m+1)\pi - \varphi\big) - \sin\varphi > 0 \end{aligned}$$

Since $D \geq |E|$, we have $F\big(2(m+1)\pi - \varphi\big) - F(\varphi) > 0$, therefore $F(\varphi_{m+1}) = F\big(2(m+1)\pi - \big(2(m+1)\pi - \varphi_{m+1}\big)\big) > F\big(2(m+1)\pi - \varphi_{m+1}\big) \geq F(\varphi_m)$. Also,

$$\begin{aligned} F(\varphi) &= (D + E\cos\varphi)\mu(\varphi) - E\sin\varphi, \\ F(\varphi + 2\pi) - F(\varphi) &= (D + E\cos\varphi)\big(\mu(\varphi + 2\pi) - \mu(\varphi)\big) \\ &\geq (D + E\cos\varphi)\frac{2\pi}{\sin^2\varphi} > D\pi, \end{aligned}$$

imply that

$$F(\varphi_{m+1}) > D\pi + F(\varphi_{m+1} - 2\pi) \geq D\pi + F(\varphi_{m-1}),$$

so $\lim_{m\to+\infty} F(\varphi_m) = +\infty$. ∎

Next we are going to show that the actions associated to the solutions of (19) increase strictly with θ. The argument here is very similar to the proof of Theorem 3.24 in ([2]). Let $f(\tau)$ be a complex function defined as

$$f(\tau) = -it\tau + \tau(a|x|^2 + u + u^0)\coth(2a\tau) + 2\sqrt{uu_0}\,\delta\frac{\tau}{\sinh(2a\tau)}. \tag{24}$$

We will see later that this function is in fact the modified complex action function for the interior.

We have the following lemma:

Lemma 2 *The function $f(\tau)$ has finitely many critical points on the imaginary axis; there is one critical point between the origin and the first pole of f on the positive imaginary axis, and it is a local maximum for f; there are either zero or two critical points(counting multiplicity) between each pair of poles on the positive imaginary axis; of such a pair of critical points the one nearer the origin is a local minimum and the other a local maximun for f.*

Proof Notice that

$$f(i\theta) = t\theta + \theta D\cot(2a\theta) + E\theta\csc(2a\theta)$$

and

$$\begin{aligned} \frac{d}{d\theta}f(i\theta) &= t + D\cot(2a\theta) - 2a\theta D\frac{1}{\sin^2(2a\theta)} + \frac{E}{\sin(2a\theta)} - \frac{2a\theta E\cos(2a\theta)}{\sin^2(2a\theta)} \\ &= t - F(\varphi), \end{aligned} \tag{25}$$

where $\varphi = 2a\theta$ and $F(\varphi)$ is the function defined above. Then the lemma follows from the properties of function $F(\varphi)$. ■

Lemma 3 *When $|x| \neq 0$ or $u \neq u^0$, there is exactly one branch of the set*

$$\Gamma_0 = \{\tau \,|\, \operatorname{Im} f(\tau) = 0, \quad \operatorname{Re}\tau > 0,\ \operatorname{Im}\tau > 0\}$$

that goes to ∞ in the quadrant $\operatorname{Re}\tau > 0$, $\operatorname{Im}\tau > 0$. On this branch $\operatorname{Re} f$ increases as $\tau \to \infty$.

Proof Suppose $\tau = s + i\theta$, $s, \theta > 0$ and let

$$\tilde{f}(\tau) = \tau D\coth(\tau) + \frac{2aE\tau}{\sinh(\tau)};$$

then

$$\operatorname{Im}\tilde{f}(\tau) = \frac{1}{\sinh^2 s + \sin^2\theta}\Big(\frac{D}{2}\big(\theta\sinh(2s) - s\sin(2\theta)\big) + E(\theta\cos\theta - s\sin\theta\sinh s)\Big)$$

For any fixed $\theta = \operatorname{Im}\tau > 0$,

$$\lim_{s\to+\infty}\operatorname{Im}\tilde{f}(\tau) = \frac{D\theta}{2}, \tag{26}$$

uniformly for bounded θ. Moreover, if $\sin\theta = 0$, then

$$\begin{aligned} \operatorname{Im}\tilde{f}(\tau) &= \frac{1}{\sinh^2 s}\Big(\frac{D}{2}\theta\sinh(2s) \pm E\theta\sinh s\Big) \\ &= \frac{\theta}{\sinh s}(D\cosh s \pm E) \\ &> \frac{D\theta}{s}. \end{aligned} \tag{27}$$

https://doi.org/10.4153/CJM-2004-026-6 Published online by Cambridge University Press

For $\sin\theta = 0$, since $f(\tau) = \frac{1}{2a}\tilde{f}(2a\tau) - st$,

$$\lim_{s\to 0} f(\tau) = +\infty, \quad \lim_{s\to+\infty} f(\tau) = -\infty,$$

therefore, $\exists s_\theta$ s.t. $\operatorname{Im} f(s_\theta + i\theta) = 0$. At $s_\theta + i\theta$, from (27)

$$0 = \operatorname{Im} f(s_\theta + i\theta) = \frac{1}{2a}\operatorname{Im}\tilde{f}(2a\tau) - s_\theta t > \frac{D\theta}{s} - s_\theta t.$$

This implies that

$$s_\theta > \sqrt{\frac{D\theta}{t}}. \tag{28}$$

The derivative

$$\begin{aligned}
\frac{\partial}{\partial s}\big(\operatorname{Im}\tilde{f}(\tau)\big) &= D\frac{-\theta\sinh^2 s\cos^2\theta + \theta\cosh^2 s\sin^2\theta_2 s\cosh s\sinh s\cos\theta\sin\theta}{(\sinh^2 s + \sin^2\theta)^2} \\
&\quad - \frac{D\cos\theta\sin\theta}{\sinh^2 s + \sin^2\theta} \\
&\quad - \frac{2E\sinh s\cosh s}{(\sinh^2 s + \sin^2\theta)^2}(\theta\cos\theta\sinh s - s\sin\theta\cosh s) \\
&\quad + \frac{E}{\sinh^2 s + \sin^2\theta}\cdot(\theta\cos\theta\cosh s - \sin\theta\cosh s + s\sin\theta\sinh s) \\
&= O\Big(\frac{s+\theta}{\sinh s}\Big)
\end{aligned}$$

as $s \to +\infty$. This gives the estimate for suitably large θ, and $\sin\theta = 0$,

$$\Big|\frac{\partial}{\partial s}\big(\operatorname{Im}\tilde{f}(\tau)\big)\Big| \le C\frac{s+\theta}{\sinh s} \le C_1\theta\exp(-C_2\sqrt{\theta}), \quad s > s_\theta,$$

where C, C_1, and C_2 are positive constants which depend only on a, x, t, u, u^0. This estimate implies that for suitably large θ, and $\sin\theta = 0$,

$$\frac{\partial}{\partial s}\big(\operatorname{Im} f(\tau)\big) < 0, \quad s > s_\theta,$$

and there is only one solution s_θ of $\operatorname{Im} f(\tau) = 0$. It follows from (26) that no branch of Γ_0 can escape to ∞ between two such lines $\operatorname{Im}\tau = \theta$, with $\sin(2a\theta) = 0$. Since $\operatorname{Im} f = 0$ on the imaginary τ−axis, a branch can escape from the quadrant through the imaginary τ−axis only at a critical point of f. By previous lemma, for large θ, there is no critical point of f, and therefore no such escape. Suppose there is a branching of Γ_0 between two consecutive $\operatorname{Im}\tau = \theta$. Two such branches must join at the two s_θ points, since there are unique. This implies the existence of a bounded region Ω on which $\operatorname{Im} f$ is harmonic and non-constant and vanishes on $\partial\Omega$, which is a contradiction. Thus, for large $|\tau|$, there is exactly one branch of Γ_0 that goes to infinity within the quadrant.

In particular, f has no critical points on this single branch of Γ_0. Therefore, on this branch, Re f must increase or decrease, since it can not have a stationary point. Using (28), for θ such that $\sin(2a\theta) = 0$, we have

$$\operatorname{Re} f(s_\theta + i\theta) = \theta t + Ds_\theta \coth(2as_\theta) \pm E\frac{s_\theta}{\sinh(2as_\theta)} \to +\infty, \quad \text{as } \theta \to +\infty.$$

Therefore, Re f increases on the branch of Γ_0 which goes off to infinity. ■

Similar to the Lemma 3.45 of [2], we have

Lemma 4 *Assume* $|x| \neq 0$ *or* $u \neq u^0$. *Let the critical points of* f *on the positive imaginary axis, counted according to the multiplicity, be* $i\theta_1, \dots, i\theta_{am+1}$, *with*

$$\theta_1 < \theta_2 \leq \theta_3 < \cdots < \theta_{2m} \leq \theta_{2m+1}. \tag{29}$$

Let Γ *be the union of* Γ_0 *and the closed intervals*

$$[0, i\theta_1], [i\theta_2, i\theta_3], \dots, [i\theta_{2m}, i\theta_{2m+1}].$$

Then Γ, *oriented in the direction of increasing* Re f, *is a simply connected curve from* 0 *to* ∞.

Proof The proof of Lemma 3.45 of [2] also applies here. ■

Theorem 3 *For any two points in the interior, there is a unique shortest geodesic connecting them.*

Proof Because of the invariance under the H_1 action, we can take the starting point to be $(0, 0, u^0)$. Suppose the ending point is (x, t, u). We continue to assume that t is positive. Then every solution θ of (19) corresponds to a geodesic connecting these two points. The square of the length of the corresponding geodesic is $S(\theta)$, which is given by (20). By (25), the solutions of (19) are in one-to-one correspondence with the critical points of f. Let $i\theta_k$ be the critical points of f on the positive imaginary axis, numbered as in (29). Since Lemma 4 says that Γ has no self-intersection, the critical points occur in the order on the oriented curve Γ. Therefore:

$$f(i\theta_1) < f(i\theta_2) \leq f(i\theta_3) < \cdots < f(i\theta_{2m}) \leq f(i\theta_{2m+1}),$$

with strict inequality where the corresponding inequality in 29 is strict. At critical point $i\theta$, we have

$$t = F(\theta) = D\mu(2a\theta) + E\left(\frac{2a\theta\cos(2a\theta)}{\sin^2(2a\theta)} - \frac{1}{\sin(2a\theta)}\right),$$

https://doi.org/10.4153/CJM-2004-026-6 Published online by Cambridge University Press

and

$$\begin{aligned}2af(i\theta) &= 2at\theta + 2a\theta D\cot(2a\theta) + 2aE\theta\csc(2a\theta)\\ &= D\left(\frac{\varphi^2}{\sin^2\varphi} - \frac{\varphi\cos\varphi}{\sin\varphi}\right) + E\left(\frac{\varphi^2\cos\varphi}{\sin^2\varphi} - \frac{\varphi}{\sin\varphi}\right)\\ &\qquad + \frac{D\varphi\cos\varphi}{\sin\varphi} + \frac{E\varphi}{\sin\varphi}\\ &= \frac{\varphi^2}{\sin^2\varphi}(D + E\cos\varphi)\\ &= 2aS\end{aligned}$$

by (21), where $\varphi = 2a\theta$. Therefore the corresponding action (and therefore the length) of the geodesic increases strictly with θ. In the interval $0 \neq \varphi < \pi$, for either case ($\delta = 1$ or $\delta = -1$), we have a unique solution of (19). These two solutions correspond to two geodesics. The shorter of these two will give us the shortest geodesic.

In order to determine which one is shorter, we only need to look the associated action. Write $\varphi_\delta = 2a\theta_\delta$, and $D = a|x|^2 + u + u^0$. From (19), we have

$$2\delta\sqrt{uu^0} = \frac{t - D\mu(\varphi)}{\frac{\varphi\cos(\varphi)}{\sin^2\varphi} - \frac{1}{\sin(\varphi)}}.$$

Substituting this in (20), we get:

$$\begin{aligned}2aS &= \frac{\varphi^2}{2a\sin^2\varphi}\left(D + \cos\varphi\frac{t - D\mu(\varphi)}{\frac{\varphi\cos\varphi}{\sin^2\varphi} - \frac{1}{\sin\varphi}}\right)\\ &= \frac{\varphi^2 D}{1 - \varphi\cot\varphi} + \frac{\varphi^2 t}{\varphi - \tan\varphi}.\end{aligned}$$

Note that $\varphi_{-1} \leq \varphi_{+1}$, and equality holds if and only if $\varphi_{-1} = \varphi_{+1} = 0$, which implies $t = 0$ or $u^0 = 0$, and the two geodesics coincide with each other.

Take the derivative of $2aS$ with respect to φ:

$$\begin{aligned}\frac{d}{d\varphi}(2aS) = \frac{\varphi}{(\varphi - \tan\varphi)^2}\Bigg(&\left(2\tan^2\varphi - \varphi\tan\varphi - \frac{\varphi^2}{\cos^2\varphi}\right)D\\ &+ (\varphi\tan^2\varphi + 2\varphi - 2\tan\varphi)t\Bigg).\end{aligned} \tag{30}$$

We want to prove that $\frac{d}{d\varphi}(2aS) \geq 0$, for $\forall\varphi \leq \varphi_{+1}$.

φ_{+1} satisfies the equation

$$t = D\mu(\varphi_{+1}) + 2\left(\frac{\varphi_{+1}\cos(\varphi_{+1})}{\sin^2(\varphi_{+1})} - \frac{1}{\sin(\varphi_{+1})}\right)\delta\sqrt{uu^0}.$$

For any $\varphi \leq \varphi_{+1}$, it satisfies the same equation, with a different $t' \leq t$.

$$2\sqrt{uu^0} \leq u + u^0 \leq D \quad \text{and} \quad \frac{\varphi\cos\varphi}{\sin^2\varphi} - \frac{1}{\sin\varphi} \leq 0$$

imply that

$$t' \geq D\left(\mu(\varphi) + \frac{\varphi\cos(\varphi)}{\sin^2(\varphi)} - \frac{1}{\sin(\varphi)}\right).$$

Substituting this in (30), we have

$$\begin{aligned}
&\frac{\varphi}{(\varphi-\tan\varphi)^2 D}\frac{d}{d\varphi}(2aS)|_{\varphi} \\
&\quad \geq -\varphi\tan\varphi + 2\tan^2\varphi - \frac{\varphi^2}{\cos^2\varphi} \\
&\qquad + \frac{t}{t'}\left(\frac{\varphi - \sin\varphi\cos\varphi}{\sin^2\varphi} + \frac{\varphi\cos\varphi - \sin\varphi}{\sin^2\varphi}\right)(2\varphi - 2\tan\varphi + \varphi\tan^2\varphi) \\
&\quad \geq \cos^2\varphi\sin^2\varphi\big(\varphi^2(2\cos^2\varphi + \cos^3\varphi + \cos\varphi) \\
&\qquad + \varphi(-3\cos\varphi\sin\varphi - \sin\varphi - 3\cos^2\varphi\sin\varphi - \cos\varphi\sin^3\varphi) \\
&\qquad - 2\cos^2\varphi\sin\varphi + 2\cos\varphi\sin^2\varphi + 2\cos^2\varphi\sin^2\varphi + 2\sin^4\varphi\big) \\
&\quad = \cos^2\varphi\sin^2\varphi(1+\cos\varphi)(\varphi\cos\varphi - \sin\varphi)\big((1+\cos\varphi)\varphi - 2\sin\varphi\big) \\
&\quad \geq 0.
\end{aligned}$$

Therefore we have $\frac{d}{d\varphi}(2aS) \geq 0$, for all $\varphi \leq \varphi_{+1}$, which implies $(2aS)|_{\varphi_{-1}} < (2aS)|_{\varphi_{+1}}$. Note that $\delta = +1$ means $\omega + 2a\theta > \pi$, so the geodesic corresponding to φ_{+1} hits the boundary. The geodesic correspond to φ_{-1} lies completely in the interior, and gives the shortest geodesic connecting $(0,0,u^0)$ and (x,t,u). ■

Remark If t is large enough (19) may also have finite solutions outside of interval $|2a\theta| < \pi$, which means we can have finitely many geodesics joining $(0,0,u^0)$ to (x,t,u). But for these geodesics, $|2a\theta| > \pi$, and (16) shows they will hit the boundary $(u=0)$.

Theorem 4 *There is a unique shortest geodesic connecting a boundary point and an interior point. This geodesic lies in the interior except the starting point.*

Proof Because of the invariance under the H_1 action, we can take the starting point to be $(0,0,0)$. Taking $u^0 = 0$, and following the steps of the proof of Theorem 3, we have:

$$u(s) = u_0\sin^2(\omega + 2a\theta s), \quad s \in [0,1], \ \omega \in [0,\pi)$$

$u(0) = 0$ implies

$$u_0\sin^2(\omega) = 0. \tag{31}$$

https://doi.org/10.4153/CJM-2004-026-6 Published online by Cambridge University Press

Since $u_0 \neq 0$, $\omega = 0$. Using the boundary condition $u(1) = u$, we have $u_0 = u/\sin^2(2a\theta)$. From (19) and (20), θ is the solution of

$$t = (a|x|^2 + u)\mu(2a\theta)$$

and the corresponding action

$$S = \frac{2a\theta^2}{\sin^2(2a\theta)}(a|x|^2 + u).$$

The unique shortest geodesic corresponds to the unique solution of θ that lies in the interval $[0, \pi/2a)$. ■

For any point (x, t, u) in the interior, denote by $\gamma(s; x, t, u)$ the shortest geodesic that connects the origin and this point. If $t = 0$, then $\theta = 0$ which is trivial. Therefore we assume that $t > 0$. An interesting problem is to consider the limit of $\gamma(s; x, t, u)$ as $u \to 0$. Two different cases emerge:

First Case: $x \neq 0$ In this case, as $u \to 0$, $\theta(x, t, u)$ goes to θ_0, which is the solution of $t = a|x|^2\mu(2a\theta)$ that lies in the interval $[0, \pi/2a)$. With $u = 0$, (31) implies $u_0 = 0$, $u(s) \equiv 0$. Thus the limiting geodesic lies on the boundary.

Second Case: $x = 0$ As $u \to 0$, $\theta(0, t, u)$ goes to $\theta_0 = \pi/2a$. Take the limit of (31):

$$\begin{aligned}
\lim_{u\to 0} u_0(0, t, u) &= \lim_{u\to 0} \frac{u}{\sin^2(2a\theta)} \\
&= \lim_{u\to 0} \frac{t}{\mu(2a\theta)\sin^2(2a\theta)} \\
&= \lim_{\theta\to\pi/2a} \frac{t}{2a\theta - \cos(2a\theta)\sin(2a\theta)} \\
&= \frac{t}{\pi}
\end{aligned}$$

Therefore the limiting geodesic can be described as:

$$x(s) = 0,\ t(s) = \frac{t}{2\pi}\big(2\pi s - \sin(2\pi s)\big),\ u(s) = \frac{t}{\pi}\sin^2(\pi s) \quad s \in [0, 1].$$

This corresponds to the case $\zeta(0) = 0$ in Theorem 1.

Remark This verifies the fact that a geodesic that starts from the origin and returns to the boundary necessarily returns to the boundary at the t-axis (see Theorem 1).

4 Geodesics on the Hypersurface $u \equiv u_0$

In this section we deal with geodesics on the hypersurface $u \equiv u_0$. We restrict the Hamiltonian $H = \frac{1}{2}\langle\zeta,\zeta\rangle + \frac{1}{2}bu(\theta^2+\sigma^2)$ to the hypersurface and get $H_u = \frac{1}{2}\langle\zeta,\zeta\rangle + \frac{1}{2}bu_0\theta^2$. Then Hamilton's equations for a curve $\big(x(s), t(s), \zeta(s), \theta(s)\big)$ can be written as

$$\begin{aligned} \dot{x} &= \zeta; \quad \dot{\zeta} = 2\theta\Lambda\zeta; \\ \dot{t} &= \langle\zeta,\Lambda x\rangle + bu_0\theta; \quad \dot{\theta} = 0. \end{aligned} \tag{32}$$

First we consider an arbitrary normalized geodesic. Because of the invariance under the H_1 action, we take the origin $(0,0)$ as the starting point, *i.e.*, we have the following boundary conditions:

$$x(0) = 0,\ t(0) = 0,\ x(1) = x,\ t(1) = t.$$

Integrating (32) and using the boundary condition, we obtain:

$$\begin{aligned} \zeta(s) &= \exp(2s\theta\Lambda)\zeta(0) \\ x(s) &= (2\theta\Lambda)^{-1}\big(\exp(2s\theta\Lambda) - I\big)\zeta(0). \end{aligned} \tag{33}$$

As before we have that $\langle\zeta,\zeta\rangle$ is constant along the curve. Then (32) implies:

$$\dot{t} = \frac{1}{2\theta}\langle\zeta,\zeta-\zeta(0)\rangle + bu\theta = \frac{1}{2\theta}\langle\zeta(0),\zeta(0)\rangle - \frac{1}{2\theta}\langle\dot{x},\zeta(0)\rangle + bu_0\theta. \tag{34}$$

We integrate (34) and use the boundary condition:

$$t = \frac{1}{2\theta}|\zeta(0)|^2 - \frac{1}{2\theta}\langle x,\zeta(0)\rangle + bu_0\theta. \tag{35}$$

From (33) we have:

$$\begin{aligned} \langle\zeta(0),\zeta(0)\rangle &= \Big\langle \big(\exp(2\theta\Lambda) - I\big)^{-1} 2\theta\Lambda x, \big(\exp(2\theta\Lambda) - I\big)^{-1} 2\theta\Lambda x\Big\rangle \\ &= \frac{(2a\theta)^2}{\sin^2(2a\theta)}|x|^2, \end{aligned}$$

$$\begin{aligned} \langle x,\zeta(0)\rangle &= \zeta^t(0)(2\theta\Lambda)^{-1}\big(\exp(2s\theta\Lambda) - I\big)\zeta(0) \\ &= \frac{1}{4a\theta}\sin(4a\theta)|\zeta(0)|^2 = 2a\theta\cot(2a\theta)|x|^2. \end{aligned}$$

Therefore we have

$$t = \frac{1}{2\theta}\frac{(2a\theta)^2}{\sin^2(2a\theta)}|x|^2 - \frac{1}{2\theta}2a\theta\cot(2a\theta)|x|^2 + bu_0\theta = a\mu(2a\theta)|x|^2 + bu_0\theta.$$

The action is

$$S = H = H\big(x(0), t(0), \zeta(0), \theta(0)\big)$$
$$= \frac{1}{2}|\zeta(0)|^2 + \frac{1}{2}bu_0\theta^2 = \frac{2(a\theta)^2}{\sin^2(2a\theta)}|x|^2 + \frac{1}{2}bu_0\theta^2.$$

Similar to the Theorem 1.36 of [2], we have

Theorem 5 *Assume that $x \neq 0$. There are only finitely many geodesics that join the origin to (x, t). These geodesics are parametrized by the solutions θ of:*

$$|t| = a\mu(2a\theta)|x|^2 + bu_0\theta, \tag{36}$$

and their lengths increase strictly with θ.

The square of the length of the geodesic associated to a solution θ of (36) is

$$2S(x, |t|, 1; \theta) = \nu(2a\theta)\Big(\frac{t - b\theta u_0}{a} + |x|^2\Big) + u_0 b\theta^2, \tag{37}$$

where $\nu(0) = 1$ and otherwise

$$\nu(\varphi) = \frac{\varphi^2}{\varphi + \sin^2\varphi - \sin\varphi\cos\varphi}.$$

Consequently, if $2a\theta \in \big(k\pi, (k+1)\pi\big)$ the length d_θ of the geodesic satisfies

$$\frac{(k+1)^2\pi^2}{k\pi}\Big(\frac{|t| - k\pi u_0}{a} + |x|^2\Big) < d_\theta^2 < \frac{k^2\pi^2}{(k+3/4)\pi + 1}\Big(\frac{t - \big((k+1/4)\pi - 1\big)u_0}{a} + |x|^2\Big). \tag{38}$$

Proof The geodesics that join the origin to (x, t) correspond exactly to the solution of (36) if $t \geq 0$, and to the negatives of the solutions if $t < 0$. Therefore the enumeration of the geodesics follows easily from the properties of the function $\mu(\varphi)$ (see Lemma 1.33 in [2]). The expression of associated action is

$$\begin{aligned}
2S(x, |t|, 1; \theta) &= \frac{(2a\theta)^2}{\sin^2(2a\theta)}|x|^2 + bu_0\theta^2 \\
&= \frac{(2a\theta)^2|x|^2}{\sin^2(2a\theta)\big(|x|^2 + (|t| - b\theta u_0)/a\big)}\big(|x|^2 + (|t| - b\theta u_0)/a\big) + bu_0\theta^2 \\
&= \frac{(2a\theta)^2|x|^2}{\sin^2(2a\theta)(1 + \mu(2a\theta)}\big(|x|^2 + (|t| - b\theta u_0)/a\big) + bu_0\theta^2 \\
&= \nu(2a\theta)\Big(\frac{|t| - b\theta u_0}{a} + |x|^2\Big) + u_0 b\theta^2.
\end{aligned}$$

To get the estimate (38), we consider the denominator of $\nu(\varphi)$, namely $\varphi + \sin^2\varphi - \sin(\varphi)\cos(\varphi)$, on the interval $[m\pi, (m+1)\pi]$. It is easy to see that the minimum occurs at the point $m\pi$, and the maximum occurs at the point $m\pi + \frac{3}{4}\pi$. Therefore

$$m\pi \le \varphi + \sin^2\varphi - \sin(\varphi)\cos(\varphi) \le \Big(m + \frac{3}{4}\Big)\pi + 1,$$

on $[m\pi, (m+1)\pi]$.

$$\begin{aligned}(d_\theta)^2 = 2S &= \nu(2a\theta)\Big(\frac{|t| - b\theta u_0}{a} + |x|^2\Big) + u_0 b\theta^2\\ &< \frac{(k+1)^2\pi^2}{k\pi}\Big(\frac{|t|}{a} - 4\theta u_0 + |x|^2\Big) + 4u_0 a\Big(\frac{(k+1)\pi}{2a}\Big)^2\\ &= (k+1)^2\pi^2\Big(\frac{1}{k\pi}\Big(\frac{|t|}{a} - 4\theta u_0 + |x|^2\Big) + \frac{u_0}{a}\Big)\\ &\le (k+1)^2\pi^2\Big(\frac{1}{k\pi}\Big(\frac{|t|}{a} - \frac{2k\pi u_0}{a} + |x|^2\Big) + \frac{u_0}{a}\Big)\\ &= \frac{(k+1)^2\pi^2}{k\pi}\Big(\frac{|t| - k\pi u_0}{a} + |x|^2\Big),\end{aligned}$$

and

$$\begin{aligned}(d_\theta)^2 = 2S &= \nu(2a\theta)\Big(\frac{|t| - b\theta u_0}{a} + |x|^2\Big) + u_0 b\theta^2\\ &> \frac{k^2\pi^2}{(k+3/4)\pi + 1}\Big(\frac{|t|}{a} - 4\theta u_0 + |x|^2\Big) + 4u_0 a\Big(\frac{(k+1)\pi}{2a}\Big)^2\\ &= k^2\pi^2\Big(\frac{1}{(k+3/4)\pi + 1}\Big(\frac{|t|}{a} - 4\theta u_0 + |x|^2\Big) + \frac{u_0}{a}\Big)\\ &\ge k^2\pi^2\Big(\frac{1}{(k+3/4)\pi + 1}\Big(\frac{|t|}{a} - \frac{2(k+1)\pi u_0}{a} + |x|^2\Big) + \frac{u_0}{a}\Big)\\ &= \frac{k^2\pi^2}{(k+3/4)\pi + 1}\Big(\frac{t - \big((k+1/4)\pi - 1\big)u_0}{a} + |x|^2\Big).\end{aligned}$$

Finally, we prove that the lengths increase strictly with θ. Let

$$\begin{aligned}f(\varphi) &= |x|^2\mu(\varphi) + 2\frac{u_0}{a}\varphi,\\ g(\varphi) &= \frac{\varphi^2}{\sin^2\varphi}|x|^2 + \frac{u_0}{a}\varphi^2,\\ h(\varphi) &= \frac{\varphi^2}{\sin^2\varphi}.\end{aligned}$$

Then we have $t/a = f(2a\theta), 2S(x,|t|,1;\theta) = g(2a\theta)$, and $\frac{dg}{d\varphi} = \varphi\frac{df}{d\varphi}$. Suppose first that there are two solutions $\theta_1 < \theta_2$ of (36) in the same interval $\big(m\pi/(2a), (m+1)\pi/(2a)\big)$. Since $f''(\varphi) = |x|^2\mu''(\varphi) > 0$ in the interval, and $f(\varphi) \to +\infty$ as $\varphi \to m\pi^+$ or $\varphi \to (m+1)\pi^-$, $f(\varphi)$ has a unique critical point ϕ_m on the interval $\big(m\pi, (m+1)\pi\big)$, and the two solutions θ_1 and θ_2 satisfy $2a\theta_1 < \phi_m < 2a\theta_2$. Noticing that $g'(\varphi) = \varphi f'(\varphi)$,

$$g(2a\theta_2) - g(\phi_m) = \int_{\phi_m}^{2a\theta_2} g'(t)\,dt = \int_{\phi_m}^{2a\theta_2} tf'(t)\,dt = t_2(f(2a\theta_2 - f(\phi_m)), \tag{39}$$

where $t_2 \in (\phi_m, 2a\theta_2)$. And similarly

$$g(2a\theta_1) - g(\phi_m) = \int_{\phi_m}^{2a\theta_2} tf'(t)\,dt = t_1\big(f(2a\theta_1) - f(\phi_m)\big), \tag{40}$$

where $t_1 \in (2a\theta_1, \phi_m)$. Since $t_1 < t_2$ and $f(2a\theta_1) = t/a = f(2a\theta_2)$, it follows from (39) and (40) that:

$$2S(x,|t|,1;\theta_1) = g(2a\theta_1) < g(2a\theta_2) = 2S(x,|t|,1;\theta_2).$$

Now turn to the case that θ_1 and θ_2 are consecutive solutions of (36) that lie on either side of $m\pi/2a$. From the properties of the function $\mu(\varphi)$ (see [2]), on each interval $\big(m\pi, (m+1)\pi\big)$, $m = 1, 2, 3\dots,$ μ has a unique critical point φ_m.

$$\begin{aligned}
h(2a\theta_1) &= h(\varphi_m) + \int_{\varphi_m}^{2a\theta_1} t\mu'(t)\,dt \\
&= \frac{\varphi_m^2}{\sin^2(\varphi_m)} + 2a\theta_1\mu(2a\theta_1) - \varphi_m\mu(\varphi_m) - \int_{\varphi_m}^{2a\theta_1}\mu(t)\,dt \\
&= 1 + 2a\theta_1\mu(2a\theta_1) - \int_{\varphi_m}^{2a\theta_1}\mu(t)\,dt.
\end{aligned}$$

Therefore,

$$\begin{aligned}
g(2a\theta_1) &= h(2a\theta_1)|x|^2 + \frac{u(2a\theta_1)^2}{a} \\
&= |x|^2 + 2a\theta_1 f(2a\theta_1) - \frac{u_0(2a\theta_1)^2}{a} - |x|^2\int_{\varphi_m}^{2a\theta_1}\mu(t)\,dt,
\end{aligned}$$

and similarly

$$g(2a\theta_2) = |x|^2 + 2a\theta_2 f(2a\theta_2) - \frac{u_0(2a\theta_2)^2}{a} - |x|^2\int_{\varphi_{(m+1)}}^{2a\theta_2}\mu(t)\,dt.$$

Subtract the above two equations and notice that $f(2a\theta_1) = f(2a\theta_2) = t/a$:

$$g(2a\theta_2) - g(2a\theta_1) = (2a\theta_2 - 2a\theta_1)\frac{t}{a} - \frac{u_0}{a}\big((2a\theta_2)^2 - (2a\theta_1)^2\big) + |x|^2\left(\int_{2a\theta_2}^{\varphi_{m+1}} - \int_{2a\theta_1}^{\varphi_m}\right)\mu(t)\,dt. \tag{41}$$

Since $t/a = h(2a\theta_2) + u_0(2a\theta_2)^2/a$,

$$(2a\theta_2 - 2a\theta_1)\frac{t}{a} - \frac{u_0}{a}\big((2a\theta_2)^2 - (2a\theta_1)^2\big) = \frac{2a\theta_2 - 2a\theta_1}{a}\big(t - u_0(2a\theta_1 + 2a\theta_2)\big) > 0.$$

It is obvious that $2a\theta_2 \le \phi_{m+1} < \varphi_{m+1}$. Moreover

$$\begin{aligned} f(\varphi_m) &= \mu(\varphi_m)|x|^2 + \frac{2u_0\varphi_m}{a} < |x|^2\mu(\varphi_{m+1}) - \pi + 2\frac{u_0\phi_{m+1}}{a} \\ &< |x|^2\mu(2a\theta_2) - \pi + 2\frac{u_0}{a}\varphi_m = |x|^2\mu(2a\theta_1) + \frac{2u_0}{a}(2a\theta_1 - 2a\theta_2) + \frac{2u_0}{a}\varphi_m \\ &= f(2a\theta_1) + \frac{2u_0}{a}(\varphi_m - 2a\theta_2) - \pi < f(2a\theta_1) \end{aligned}$$

imply that $\varphi_m < 2a\theta_1$. It follows from (41) that $g(2a\theta_1) < g(2a\theta_2)$. ■

Next, we consider the geodesic which starts from the origin and ends at a point on the $t-$axis. The boundary conditions we have now are

$$x(0) = 0,\ t(0) = 0,\ x(1) = 0,\ t(1) = t. \tag{42}$$

Without loss of generality, we may assume that $t > 0$. We have the following theorem:

Theorem 6 *If $t < 2\pi u_0$, there is a unique geodesic joining the origin and $(0, t)$, and this geodesic coincides with the line segment $[0, t]$ on the $t-$axis. If $2n\pi u_0 \le t < 2(n+1)\pi u_0$, $n \in \mathbf{N}$, besides the geodesic, which coincides with the line segment $[0, t]$ on the $t-$axis, for each $m \le n$, $n \in \mathbf{N}$, we have a family of geodesics, which are parametrized by the circle S^1, join the origin to $(0, t)$, and they have the same length $d_m = \sqrt{\frac{m\pi}{a}(t - m\pi u_0)}$.*

Proof We have two cases: (1) $\zeta(0) \neq 0$. In this case, (33) and the boundary condition (42) imply $\exp(2s\theta\Lambda) - I = 0$, so $\exp(2a\theta) = I$. We get $\zeta(s) = \zeta(0)$ and $2a\theta = m\pi, m = 1, 2, 3 \ldots$. The Hamiltonian is constant along the curve:

$$H = H\big(x(0), t(0), \zeta(0), \theta(0)\big) = \frac{1}{2}|\zeta(0)|^2 + \frac{1}{2}bu_0\theta^2. \tag{43}$$

(36) is not applicable in this case, therefore we use (35) instead and get:

$$t = \frac{1}{2\theta}|\zeta(0)|^2 + bu_0\theta = \frac{H}{\theta} + \frac{1}{2}bu_0\theta.$$

From (43) $H \geq \frac{1}{2}bu_0\theta^2$, we therefore have a restriction for t: $t \geq bu_0\theta = 2m\pi u_0$. The length of such geodesic is

$$L = \sqrt{2H} = \sqrt{2\Big(t - \frac{1}{2}bu_0\theta\Big)\,\theta} = \sqrt{\frac{m\pi}{a}(t - m\pi u_0)}.$$

Noticing that $t \geq 2m\pi u_0$, we have $L \geq m\pi\sqrt{\frac{u_0}{a}}$. For each m,

$$\begin{aligned} x^m(s) &= (2\theta\Lambda)^{-1}\big(\exp(2s\theta\Lambda) - I\big)\,\zeta(0) \\ &= \Big(\frac{m\pi}{a}\Lambda\Big)^{-1}\Big(\exp\Big(s\frac{m\pi}{a}\Lambda\Big) - I\Big)\,\zeta(0). \end{aligned}$$

Integrating (34) yields

$$\begin{aligned} t^m(s) &= \int_0^s \frac{1}{2\theta}\langle\zeta(0),\zeta(0)\rangle - \frac{1}{2\theta}\langle\dot{x}(r),\zeta(0)\rangle + bu_0\theta\,dr \\ &= \frac{as}{m\pi}|\zeta(0)|^2 - \frac{a}{m\pi}\langle x(s),\zeta(0)\rangle + 2m\pi u_0 s \\ &= \frac{as}{m\pi}|\zeta(0)|^2 - \frac{a}{m\pi}\Big\langle\Big(\frac{m\pi}{a}\Lambda\Big)^{-1}\Big(\exp\Big(s\frac{m\pi}{a}\Lambda\Big) - I\Big)\,\zeta(0),\zeta(0)\Big\rangle \\ &\quad + 2m\pi u_0 s \\ &= \frac{as}{m\pi}|\zeta(0)|^2 - \frac{\sin(2m\pi s)}{2m\pi}|\zeta(0)|^2 + 2m\pi u_0 s. \end{aligned}$$

Along such geodesics, $|\zeta(0)|^2 = m\pi(t - 2m\pi u_0)/a$ is a constant. These show that for each m, the geodesics $\big(x^m(s), t^m(s)\big)$ may be parametrized by $\zeta(0) \in S^1$.

(2) $\zeta(0) = 0$. In this case, we obtain, from (33), $x(s) \equiv 0$, $\zeta(s) \equiv 0$. Therefore we have $t(s) = bu_0\theta s$, $t = bu_0\theta$. So, in fact the geodesic coincides with the line segment $[0,t]$ on the t-axis.

$$H = \frac{1}{2}bu_0\theta^2 = \frac{1}{2}bu_0\Big(\frac{t}{bu_0}\Big)^2 = \frac{t^2}{2bu_0},$$
$$L = \sqrt{2H} = \frac{t}{\sqrt{bu_0}}.$$

If $t < 2\pi u_0$, then case (1) cannot occur, therefore we have a unique geodesic joining the origin and $(0,t)$, and this geodesic coincides with the line segment $[0,t]$ on the t-axis. If $2n\pi u_0 \leq t < 2(n+1)\pi u_0$, both cases occur. Since in case (1) we have the restriction $t \geq 2m\pi u_0$, m can only take values $1, 2, \dots, n$. ■

Remark When $t = 2\pi u_0$ these two kinds of geodesics have the same length $\pi\sqrt{\frac{u_0}{a}}$. And as t increases, the case (2) geodesic will no longer be the shortest one. Its length $t/(2\sqrt{au_0})$ is greater than the length of the geodesics in case (1) for $m = 1$. For small t, case (1) can not occur and we have a unique geodesic. This result is different from the Heisenberg group H_1, where we have infinitely many geodesics connecting the origin and $(0, t)$. But this is not surprising; because the hypersurface is a Riemannian manifold, for any two points which are close enough to each other, there is only one geodesic joining them. As $u_0 \to 0$, we can see this result reduces nicely to the Heisenberg case.

Acknowledgment The basic idea of this paper was suggested by my advisor Professor R. Beals. I am very grateful to him for useful conversations, and for his support and encouragement.

References

[1] R. Beals, *Geometry and PDE on the Heisenberg group: a case study*. In: The geometrical study of differential equations (Washington, DC, 2000), Contemp. Math., 285, Amer. Math. Soc., Providence, RI, pp. 21–27.

[2] R. Beals, B. Gaveau and P. C. Greiner, *Hamilton-Jacobi theory and the heat kernel on Heisenberg groups*. J. Math. Pures Appl. (7) **79**(2000), 633–689.

[3] B. Gaveau, *Principe de moindre action, propagation de la chaleur et estimées sous-elliptiques sur certains groupes nilpotents*. Acta Math. **139**(1977), 95–153.

[4] ———, *Systèmes dynamiques associés a certains operateurs hypoelliptiques*. Bull. Sci. Math. **102**(1978), 203–229.

[5] R. Strichartz, *Sub-Riemannian geometry*. J. Differential Geom. **24**(1986), 221–263; correction, ibid. **30**(1989), 595–596.

Department of Mathematics and Statistics
McMaster University
1280 Main Street West
Hamilton, ON
L8S 4K1
e-mail: yilong.ni@aya.yale.edu