HAMILTON–JACOBI THEORY AND THE HEAT KERNEL ON HEISENBERG GROUPS *

Richard BEALS a, Bernard GAVEAU b, Peter C. GREINER c

a Department of Mathematics, Yale University, 10 Hillhouse Avenue, P.O. Box 208283, New Haven, USA
b Université Pierre et Marie Curie (Paris-VI), 4, Place Jussieu, 75252 Paris cedex 05, France
c University of Toronto, Toronto, Canada

Manuscript received 2 October 1999

ABSTRACT. – The subelliptic geometry of Heisenberg groups is worked out in detail and related to complex Hamiltonian mechanics. The two geometric pictures are essential for complete understanding of the heat equation for the subelliptic Laplacian. We give a complete description of the geodesics and obtain precise global estimates and small-time asymptotics of the heat kernel. © 2000 Éditions scientifiques et médicales Elsevier SAS

Keywords: Heisenberg group, Sub-Riemannian geometry, Carnot–Caratheodory metric, Subelliptic diffusion, Hamiltonian mechanics


Introduction

The purpose of this article and its continuation [8] is to construct the heat kernel of the subelliptic Laplacian on contact manifolds, and to estimate it and find its small time asymptotic behaviour in terms of the subelliptic and sub-Riemannian geometries of the underlying manifold. In the present paper we study the situation on Heisenberg groups.

The Heisenberg group $H_n$, of $2n + 1$ real dimensions, is the simplest non-commutative nilpotent Lie group. The subelliptic Laplacian $\Delta_H$ on $H_n$ is a sum of squares of $2n$ “horizontal” vector fields, and is therefore not elliptic, but a well-known result of Hörmander [25] implies that nevertheless it is hypoelliptic. Given a left-invariant metric on the horizontal subspace of $TH_n$, $\Delta_H$ may be written as a sum of squares of the vector fields in any orthonormal frame of the horizontal subspace.

Heisenberg groups and their subelliptic Laplacians are at the cross-roads of many domains of analysis and geometry: nilpotent Lie group theory, hypoelliptic second order partial differential equations, strictly pseudoconvex domains in complex analysis, probability theory of degenerate diffusion processes, sub-Riemannian geometry, control theory and semiclassical analysis of quantum mechanics, see [4–7,10,19–21,24,26,28,35–38].

An explicit formula for the heat kernel of $\Delta_H$ was derived by Hulanicki [28], using representation theory, and by Gaveau [20], using probability theory. In [9] the Laguerre calculus

* Research partially supported by NSF Grant DMS-9800605, by E.U. Grant “Capital humain et mobilité” and by NSERC Grant OGP0003017.
is used to construct the heat kernel. The formula was extended in [4] to any step 2 nilpotent Lie group via complex Hamilton–Jacobi theory; see also [3,30].

The heat kernel with source at the origin has the form:

\[
P(x, t; u) = \frac{1}{(2\pi u)^{n+1}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{f(x, t, \tau)}{u} \right\} V(\tau) \, d\tau,
\]

where \((x, t) \in \mathbb{R}^{2n} \times \mathbb{R} = H_n, \ u \in \mathbb{R}_+\) represents time, and \(\tau\) is the momentum conjugate to \(t\). The function \(f(x, t, \tau)\) is a complex action; it satisfies a Hamilton–Jacobi equation with the symbol of \(\Delta_H\) as Hamiltonian, but is constructed with complex bicharacteristics. The function \(V\) is a volume element on the characteristic variety of \(\Delta_H\), obtained by integrating a first order transport equation along the bicharacteristics. Thus the heat kernel (1) is given by an integral over a one parameter family of complex actions, the integrand being the usual semiclassical entity, and the formula is exact. The probabilistic calculation in [20] shows that

\[
P(x, t; u) > 0.
\]

The pointwise asymptotic behaviour of the heat kernel in \(H_1\) was also derived in [20]. With source at the origin it is

\[
P(x, t; u) \sim \frac{1}{u^{\alpha(x,t)}} C(x,t) \exp \left\{ -\frac{S(x,t)}{u} \right\},
\]

where \(S\) is the minimal action for the (real) bicharacteristics associated to the symbol of the operator, and the exponent \(\alpha(x,t)\) is 3/2 for \(x \neq 0\) and 2 for \(x = 0, \ t \neq 0\). In fact the latter points are all conjugate to the origin, and the asymptotics (4) cannot be uniform in a neighbourhood of the origin. This is in contrast to the elliptic case, for which the asymptotics are uniform in a neighbourhood of the diagonal but break down at conjugate points; see [13,16,29]. Nevertheless, (1) is an exact formula, and the integral provides a uniformization of the asymptotic behaviour (reminiscent of the representation of Airy functions). These results were obtained by a detailed study of the bicharacteristics in \(H_1\).

A general geometric setting for a large class of subelliptic operators, now known as “sub-Riemannian geometry”, was introduced in [21] and related to isoperimetric and control problems. In this paper we derive in detail the sub-Riemannian (“Carnot–Carathéodory”) structure of the general Heisenberg group and relate it to the small time asymptotics of the heat kernel. Among other things, the results indicate how complicated a control problem can become, even in the simplest situations. The non-uniformity of the asymptotics, and the related fact that the classical action \(S\) is not a smooth function, show that sub-Riemannian concepts are not adequate for the study of subelliptic Laplacians, although they were initially introduced for that purpose. Non-uniformity also raises the question of what are the correct estimates for the heat kernel. We obtain global estimates of the form:

\[
P(x, t; u) \leq F(x, t, u) \exp \left\{ -\frac{S(x, t, u)}{u} \right\},
\]

where \(F\) is an explicit function that has piecewise algebraic growth. These estimates are proved by careful analysis of the exact formula (1), and they are sharp.

The questions considered in this paper have been discussed, sometimes in much greater generality, in many other places; see [1,2,10,14,15,27,36–38] and references therein. It was shown by
Gaveau [21] that \(-2u \log P \geq d^2\), where \(d\) is the Carnot–Carathéodory distance. Equality was conjectured by Bismut [14] and proved by Ben Arous [11] and Léandre [31] using the Malliavin calculus [32]. However, as some of the preceding references indicate, the literature is not without controversy, even in the elliptic case. Very general estimates have been proved for general nilpotent groups by Varopoulos [38], developing a method exploited in the elliptic case by Davies [17], but the method does not yield optimal exponentials or optimal powers of the time variable. Optimal estimates are asserted in [37], but the main statement, Theorem 2, is misleadingly phrased, and the argument, based on an idea of Kannai [29] in the elliptic case, makes sketchy use of a number of delicate results from the literature. In view of this situation, it seems useful to have a complete and precise working out of (classical and complex) Hamiltonian geometry and the heat kernel for the basic model cases of subellipticity: the isotropic Heisenberg and general Heisenberg models. This should include a complete picture of the geodesics and shortest curves, the small-time asymptotics, global estimates, and the interconnections among them.

The paper is organized as follows. In Section 1 we study general Hamiltonian mechanics on the three-dimensional group \(H_1\). Specializing to the classical case leads to a detailed study of the classical action \(S\) and the Carnot–Carathéodory distance. (For example, for any point \(p\), any positive integer \(m\), and any neighborhood \(U\) of \(p\), there are points that can be joined to \(p\) by exactly \(m\) geodesics, having different lengths.) Specializing to the complex version leads to the complex action function \(f\) above. The classical and complex actions are identical precisely at the critical points of \(f\) on the imaginary axis. The heat kernel on \(H_1\) is studied in Section 2. After a brief justification of the formula (1) from the point of view of complex Hamiltonian mechanics, we derive a sharp global estimate and prove the exact small time asymptotics; Yakar Kannai helped with the estimate on \(H_1\).

Sections 3, 4 and 5 are devoted to the generalization to \(H_n\). In Section 3 we obtain the classical and complex actions on \(H_n\); much of this section is devoted to finding the shortest geodesic, which yields the Carnot–Carathéodory distance. Next we obtain sharp global estimates and small time asymptotics for the heat kernel on isotropic \(H_n\), in Section 4, and on general \(H_n\), in Section 5. We obtain precise estimates on isotropic \(H_n\) with a simpler proof. On general anisotropic \(H_n\) one of our main concerns is that the estimates should be uniform with respect to the eigenvalues of the Levi form; this is essential for precise applications to general contact manifolds, to be considered later.

1. Hamiltonian mechanics and geometry on the 3-dimensional Heisenberg group \(H_1\)

The 3-dimensional Heisenberg group \(H_1\) may be realized as \(R^2 \times R = \{(x,t)\}\) with the group law:

\[(x,t) \circ (x',t') = (x + x', t + t' + 2ax_2x_1' - 2ax_1x_2'),\]

where \(a\) is a positive parameter. The vector fields

\[X_1 = \frac{\partial}{\partial x_1} + 2ax_2 \frac{\partial}{\partial t},\]

\[X_2 = \frac{\partial}{\partial x_2} - 2ax_1 \frac{\partial}{\partial t},\]

are left invariant and generate the Lie algebra of \(H_1\).

The associated Heisenberg (sub-)Laplacian is the left-invariant subelliptic operator

\[\Delta_H = \frac{1}{2}(X_1^2 + X_2^2).\]
We shall study the associated subelliptic geometry and obtain precise estimates and asymptotics for the associated heat kernel.

Another structure on \( H^1 \) associated to the vector fields \( \{X_j\} \) is a Carnot–Caratheodory distance or sub-Riemannian distance described as follows. A piecewise \( C^1 \) curve \( \gamma \) in \( H^1 \) is horizontal if its tangent vectors lie in the subbundle of the tangent bundle generated by the \( \{X_j\} \). We take \( X_1, X_2 \) to be an orthonormal frame. Any two points of \( H^1 \) can be joined by a horizontal curve and the Carnot–Caratheodory distance is the infimum of the lengths of such curves.

Both the heat kernel and the Carnot–Caratheodory metric are closely associated to Hamiltonian mechanics on \( H^1 \). The formula for the heat kernel that was found by Gaveau [20] and Hulanicki [28] can be interpreted in terms of an action function associated to complex Hamiltonian mechanics [4]. The Carnot–Caratheodory geodesics are traces of real Hamiltonian paths and the associated action function gives the square of the distance. In each case the Hamiltonian function is just the symbol of \( -\Delta_H \). Up to a point, real and complex Hamiltonian mechanics can be treated simultaneously.

**General Hamiltonian mechanics on \( H^1 \)**

The symbol of \( -\Delta_H \) is:

\[
H(x, \xi, \theta) = \frac{1}{2} \left\{ (\xi_1 + 2ax_2\theta)^2 + (\xi_2 - 2ax_1\theta)^2 \right\}.
\]

(1.4)

It will be convenient in what follows to adopt vector notation. In particular \( \partial/\partial x \) will denote the gradient in \( \mathbb{R}^2 \) and \( \langle \cdot, \cdot \rangle \) will denote the inner product

\[
\langle x, y \rangle = x_1y_1 + x_2y_2,
\]

and also its (bilinear) extension to \( C^2 \). Then we may write the group law (1.1) in the form

\[
(x, t) \circ (x', t') = (x + x', t + t' + \langle Ax, x' \rangle),
\]

(1.5)

where

\[
A = \begin{bmatrix} 0 & 2a \\ -2a & 0 \end{bmatrix}.
\]

(1.6)

Note that

\[
A' = -A, \quad A^2 = -4a^2I.
\]

(1.7)

Similarly the vector fields (1.2) may be denoted collectively by

\[
X = \frac{\partial}{\partial x} + Ax \frac{\partial}{\partial t},
\]

(1.8)

and the Hamiltonian (1.4) by

\[
H(x, \xi, \theta) = \frac{1}{2} \langle \xi + \theta Ax, \xi + \theta Ax \rangle = \frac{1}{2} \langle \xi, \xi \rangle,
\]

(1.9)

where \( \xi = \xi + \theta Ax \).
In this notation Hamilton’s equations for a curve \((x(s), t(s), \xi(s), \theta(s))\) take the form:

\[
\begin{align*}
\dot{x}(s) &= \frac{\partial H}{\partial \xi} = \xi(s), \\
\dot{\xi}(s) &= -\frac{\partial H}{\partial x} = \theta \Lambda \xi(s), \\
\dot{t}(s) &= \frac{\partial H}{\partial \theta} = \{\xi(s), \Lambda x(s)\}, \\
\dot{\theta}(s) &= -\frac{\partial H}{\partial t} = 0, \quad \text{i.e. } \theta(s) = \theta(0),
\end{align*}
\]  

(1.10)

where the dot denotes \(d/ds\). We let \(s\) run along the ray from 0 to a point \(\tau \in \mathbb{C}\).

Because of group invariance we consider paths relative to the origin and a point \((x, t)\), and assume boundary conditions

\[
x(0) = 0, \quad x(\tau) = x, \quad t(\tau) = t.
\]  

(1.11)

(For \(\tau\) or \(\theta\) not real, the curves may need to be taken in the complexification of \(H_n\).) As usual, the Hamiltonian itself is constant along a curve (1.10):

\[
H(x(s), \xi(s), \theta) = H_0 = \frac{1}{2}\{\xi(0), \xi(0)\}.
\]  

(1.12)

The boundary conditions (1.11) leave one parameter free. By (1.10), \(\theta\) is constant, and we take it to be the free parameter. Then the system (1.10) is easily integrated. In fact the equations (1.10) imply that:

\[
\dot{\xi}(s) = \dot{\xi}(s) + \theta \Lambda \xi(s) = 2\theta \Lambda \xi(s),
\]  

(1.13)

so one can solve for \(x(s)\) as a function of \(x\), \(\tau\) and \(\theta\), and then solve for \(t(s)\) as a function of \(x\), \(t\), \(\tau\) and \(\theta\). We proceed as follows:

\[
\begin{align*}
\xi(s) &= e^{2\theta \Lambda} \xi(0), \\
x(s) &= \int_0^s \xi(r) \, dr = (2\theta \Lambda)^{-1} \left\{ e^{2\theta \Lambda} - I \right\} \xi(0) = (2\theta \Lambda)^{-1} \left\{ \xi(s) - \xi(0) \right\}, \\
ts(s) &= t(0) + \int_0^s \{\xi(r), \Lambda x(r)\} \, dr = t(0) + \frac{1}{2\theta} \int_0^s \{\xi(r), \xi(r) - \xi(0)\} \, dr
\end{align*}
\]

(1.14, 1.15, 1.16)

Therefore

\[
\begin{align*}
\xi(0) &= 2\theta \Lambda \left( e^{2\theta \Lambda} - I \right)^{-1} x = \frac{\theta \Lambda e^{-\tau \theta \Lambda}}{\sinh(\tau \theta \Lambda)} x, \\
x(s) &= e^{(t-s)\theta \Lambda} \frac{\sinh(s \theta \Lambda)}{\sinh(\tau \theta \Lambda)} x.
\end{align*}
\]  

Also,

\[
\begin{align*}
\theta = 0 &\Rightarrow \xi(s) = \xi(0), \quad x(s) = \xi(0)s \quad \text{and} \quad t(s) = t(0).
\end{align*}
\]  

(1.17, 1.18)
Note that in any expression
\[ \{ F(A)x, x \} \]
we may replace \( F \) by its even part
\[ \frac{1}{2} \left[ F(A) + F(-A) \right] = \frac{1}{2} \left[ F(A) + F(-A) \right]. \]

so
\[ (1.19) \quad \{ x(s), \xi(0) \} = \frac{1}{2} \left[ \frac{\sinh(2s \theta A)}{\sinh^2(\tau \theta A)} \theta Ax, x \right]. \]

Now (1.12)–(1.19) and
\[ \exp \begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix} = \begin{bmatrix} \cos b & \sin b \\ -\sin b & \cos b \end{bmatrix} \]
yield the following

**THEOREM 1.21.** – The solution of (1.10) with boundary conditions (1.11) is:
\[ (1.22) \quad x(s) = e^{(s-\tau)\theta A} \frac{\sinh(s \theta A)}{\sinh(\tau \theta A)} x = e^{(s-\tau)\theta A} \frac{\sin(2s \theta)}{\sin(2\tau \theta)} x, \]
\[ t - t(s) = \frac{\tau - s}{2\theta} \frac{(\theta A)^2}{\sinh^2(\tau \theta A)} x, x \]
\[ (1.23) \quad -\frac{1}{4\theta} \left[ [\sinh(2\tau \theta A) - \sinh(2s \theta A)] \theta A \right] x, x \]
\[ = (\tau - s) \frac{2a^2 \theta}{\sin^2(2\tau \theta)} (x, x) - a \frac{\sin(4\tau \theta) - \sin(4s \theta)}{2\sin^2(2\tau \theta)} (x, x). \]

The value of the Hamiltonian \( H \) on this path is:
\[ (1.24) \quad H_0 = \frac{1}{2} \frac{(\theta A)^2}{\sinh^2(\tau \theta A)} x, x \]
\[ = \frac{2a^2 \theta^2}{\sin^2(2\tau \theta)} (x, x). \]

A particular consequence of (1.23) is
\[ (1.25) \quad t - t(0) = a \mu(2\tau \theta) (x, x), \]
where we have set
\[ (1.26) \quad \mu(\varphi) = \frac{\varphi}{\sin^2 \varphi} - \cot \varphi. \]

see [20]. The action integral associated to the Hamiltonian curve is:
\[ (1.27) \quad S(x, t, \tau; \theta) = \int_0^\tau \left( [\xi(s), \dot{\xi}(s)] + \theta i(s) - H(x(s), \xi(s), \theta) \right) ds. \]

Because \( H \) is homogeneous of degree 2 with respect to \( (\xi, \theta) \), we observe that
\[ (1.28) \quad S = \int_0^\tau \left\{ \left[ \xi, \frac{\partial H}{\partial \xi} \right] + \theta \frac{\partial H}{\partial \theta} - H \right\} ds = \int_0^\tau (2H - H) ds = \tau H_0. \]
From (1.23), (1.24) and (1.28) we obtain various forms for the action.

**Theorem 1.29.** – The action integral \( S(x,t;\tau;\theta) \) is given by:

\[
S(x,t;\tau;\theta) = \frac{1}{2}(\tau(\theta A)^2 - [\sinh(2\theta A)]^2) \sin 2\theta A \sinh \frac{1}{2}x^2 x^2 \coth(\theta A)x^2 x^2
\]

(1.30)

These forms of \( S \) suggest that it is at least equally natural to consider a modified action which depends on \( \tau \) and \( \theta \) only through the product \( \tau \theta \):

(1.31) \[ F(x,t,\tau \theta) = \tau S(x,t,\tau;\theta) = \left[ I - t(0) \right] \theta + \frac{1}{2}(\theta A \coth(\theta A)x^2 x^2, \quad \theta \in [0,\pi/2a), \]

The classical action and the Carnot–Caratheodory distance

Classical Hamiltonian mechanics with respect to the Hamiltonian is obtained by taking \( \tau \) and \( \theta \) to be real and requiring the curve to join two specified points. We take the points to be the origin and a point \((x,t)\); thus we complete the boundary conditions (1.11) with the condition \( t(0) = 0 \).

It is convenient to fix \( \tau = 1 \) here. Then the Hamiltonian paths are determined entirely by \( \theta \) and, in view of (1.25), \( \theta \) must satisfy

(1.32) \[ t = a\mu(2a\theta)\|x\|^2. \]

Note that \( \mu \) is an odd function. We need some further information on its behavior in the range \( \varphi > -\frac{\pi}{2} \).

**Lemma 1.33.** – \( \mu \) is a monotone increasing diffeomorphism of the interval \((-\pi,\pi)\) onto \( \mathbb{R} \). On each interval \( (m\pi, (m+1)\pi) \), \( m = 1, 2, \ldots, \mu \) has a unique critical point \( \varphi_m \). On this interval \( \mu \) decreases strictly from \( +\infty \) to \( \mu(\varphi_m) \) and then increases strictly from \( \mu(\varphi_m) \) to \( +\infty \).

Moreover

(1.34) \[ \mu(\varphi_m) + \pi < \mu(\varphi_{m+1}), \quad m = 1, 2, \ldots, \]

\[ 0 < \left( m + \frac{1}{2} \right) \pi - \varphi_m < \frac{1}{m\pi}. \]

**Proof.** – We note that \( \sin \varphi - \varphi \cos \varphi \) vanishes at \( \varphi = 0 \) and is an increasing function in \([0,\pi]\).

Therefore:

(1.35) \[ \frac{1}{2} \mu'(\varphi) = \frac{\sin \varphi - \varphi \cos \varphi}{\sin^2 \varphi} \begin{cases} 1/3, & \varphi = 0, \\ > 1/3, & \varphi \in (0,\pi). \end{cases} \]

The second inequality holds because

\[ \frac{1}{2} \mu''(\varphi) = \frac{\varphi + 2\varphi \cos^2 \varphi - 3\cos \varphi \sin \varphi}{\sin^4 \varphi} > 0; \]

note that the numerator vanishes at \( \varphi = 0 \), and its derivative is

\[ 4 \sin \varphi (\sin \varphi - \varphi \cos \varphi) > 0, \quad \varphi \in (0,\pi). \]
Therefore $\mu$ is a diffeomorphism of the interval $(-\pi, \pi)$ onto $\mathbb{R}$. In the interval $(m\pi, (m+1)\pi)$, $\mu$ approaches $+\infty$ at the endpoints. The derivative $\mu'(\varphi)$ has the same sign as $1 - \varphi \cot \varphi$, so the unique critical point is the solution of $\varphi_m = \tan \varphi_m$ in the interval. Note that

$$
\mu(\varphi + \pi) = \mu(\varphi) + \frac{\pi}{\sin^2 \varphi},
$$

so the successive minimum values increase by more than $\pi$. Finally, $d\cot \varphi/d\varphi < -1$ and we have $\tan \varphi_m > m\pi$, so $0 < (m + \frac{1}{2})\pi - \varphi_m < \cot \varphi_m - \cot(m + \frac{1}{2})\pi < 1/m\pi$ and the second part of (1.34) is a consequence.

By a geodesic for the Carnot–Caratheodory metric we mean a curve that is locally length-minimizing. Each geodesic in $H_1$ is the trace of a Hamiltonian path; this is a special case of a result of Bismut [14, p. 38] and Strichartz [36]. It is enough to understand the geodesics that start from the origin. The complete picture is given in the following two theorems:

**Theorem 1.36.** There are finitely many geodesics that join the origin to $(x, t)$ if and only if $x \neq 0$. These geodesics are parametrized by the solutions $\theta$ of:

$$
a\mu(2a\theta) \|x\|^2 = |t|, \tag{1.37}
$$

and their lengths increase strictly with $\theta$. There is exactly one such geodesic if and only if:

$$
|t| < a\mu(\varphi_1) \|x\|^2; \tag{1.38}
$$

the number of geodesics increases without bound as $|t|/a \|x\|^2 \to \infty$.

The square of the length of the geodesic associated to a solution $\theta$ of (1.37) is

$$
2S(x, |t|, 1; \theta) = v(2a\theta) \left( \frac{|t|}{a} + \|x\|^2 \right), \tag{1.39}
$$

where $v(0) = 2$ and otherwise

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

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

$$
\frac{k^2 \pi^2}{(k+1)\pi + 2 \left( \frac{|t|}{a} + \|x\|^2 \right)} < (d_\theta)^2 < \frac{(k+1)^2 \pi^2}{k\pi} \left( \frac{|t|}{a} + \|x\|^2 \right). \tag{1.40}
$$

A special case of (1.38) is the square of the Carnot–Caratheodory distance, $d(x, t)^2$:

$$
d(x, t)^2 = 2S(x, |t|, 1; \theta_c) = v(2a\theta_c) \left( \frac{|t|}{a} + \|x\|^2 \right), \tag{1.41}
$$

where $\theta_c = \theta_c(x, t)$ is the unique solution of (1.37) in the interval $[0, \pi/2a]$.

**Theorem 1.41.** The geodesics that join the origin to a point $(0, t)$ have lengths $d_1, d_2, d_3, \ldots$, where

$$
(d_m)^2 = \frac{m\pi |t|}{a}. \tag{1.42}
$$
In particular, the square of the Carnot–Caratheodory distance to the origin is \( \pi |t|/a \). For each length \( d_m \), the geodesics of that length are parametrized by the circle \( S^1 \).

Proof of Theorem 1.36. – As noted, each geodesic \( \gamma \) is the trace of a Hamiltonian path, which may be assumed to be parametrized by the unit interval. The tangent vector is:

\[
\dot{\gamma} = \dot{x}_1 + \dot{x}_2 + \frac{t}{\dot{t}} \frac{\partial}{\partial t}
\]

(1.43)

\[
= \xi_1 \frac{\partial}{\partial x_1} + \xi_2 \frac{\partial}{\partial x_2} + \langle \xi, \lambda x \rangle \frac{\partial}{\partial t}
\]

\[
= \xi_1 x_1 + \xi_2 x_2.
\]

The \( X_j \) are orthogonal with length 1, so \( \dot{\gamma} \) has constant length \( k_1 \). The expression (1.38) for the action comes from (1.30) and (1.37):

\[
2S(x, |t|, 1; \theta) = \frac{(2a\theta)^2}{\sin^2(2a\theta)} \|x\|^2
\]

\[
= \frac{(2a\theta)^2}{\sin^2(2a\theta)} \|x\|^2 + |t|/a \left( \frac{|t|}{a} + \|x\|^2 \right)
\]

(1.44)

\[
= \frac{(2a\theta)^2}{\sin^2(2a\theta)} \frac{1}{1 + \mu(2a\theta)} \left( \frac{|t|}{a} + \|x\|^2 \right)
\]

\[
= \nu(2a\theta) \left( \frac{|t|}{a} + \|x\|^2 \right).
\]

To get the estimate (1.39), we note that the minimum of the denominator of \( \nu(\phi) \), \( \sin^2 \phi [1 + \mu(\phi)] \), on the interval \([m\pi, (m + 1)\pi]\) occurs at the points \( \phi = m\pi \); this is true when \( m = 0 \), otherwise use \( \sin^2(\phi + m\pi)[1 + \mu(\phi + m\pi)] = \sin^2 \phi(1 + \mu(\phi)) + m\pi \). Therefore

\[
m\pi < \phi + \sin^2 \phi - \sin \phi \cos \phi < (m + 1)\pi + 2
\]

on this interval. The estimate (1.39) follows from this and trivial estimates for the numerator of \( \nu \).

Finally, we show that the lengths are strictly monotonically increasing in \( \theta \). Suppose first that there are two solutions \( \theta_m < \theta_m' \) of (1.37) in the same interval \((m\pi/2a, (m + 1)\pi/2a)\); then

\[
\frac{d}{d\phi} \left( \frac{\phi^2}{\sin^2 \phi} \right) = \phi' \mu'(\phi), \quad \phi = 2a\theta,
\]

implies that both these functions are minimal at \( \varphi_m \) on the interval in question, while \( \varphi^2 / \sin^2 \varphi \) increases faster relative to \( \mu \) as one moves to the right from \( \varphi_m \), compared to movement to the left. More precisely we have

\[
\frac{(2a\theta_m)^2}{\sin^2(2a\theta_m')} = \frac{\varphi_m^2}{\sin^2 \varphi_m} = \int_{\varphi_m}^{\varphi_m'} \frac{t\mu'(t) dt}{\varphi_m} = \int_{\varphi_m}^{\varphi_m'} \left[ \mu(2a\theta_m') - \mu(\varphi_m') \right],
\]

(1.46)
with $t'_m \in (\varphi_m, 2a\theta'_m)$, and a similar formula for $(2a\theta_m)^2 / \sin^2(2a\theta_m)$ with $t_m \in (2a\theta_m, \varphi_m)$. Since $t_m < t'_m$ and $\mu(2a\theta_m) = \mu(2a\theta'_m)$, it follows that:

\begin{equation}
\frac{(2a\theta_m)^2}{\sin^2(2a\theta_m)} < \frac{(2a\theta'_m)^2}{\sin^2(2a\theta'_m)},
\end{equation}

and thus that the actions are in the same relation. A similar argument applies if $\theta'_m$ and $\theta'_{m+1}$ are consecutive solutions of (1.37) that lie on either side of $m \pi/2a$. In particular, we have:

\begin{equation}
\frac{(2a\theta'_{m+1})^2}{\sin^2(2a\theta'_{m+1})} - \frac{\varphi^2_{m+1}}{\sin^2 \varphi_{m+1}} = t'_{m+1} \left[ \mu(2a\theta'_{m+1}) - \mu(\varphi_{m+1}) \right],
\end{equation}

$t'_{m+1} \in (2a\theta'_{m+1}, \varphi_{m+1})$. To compare (1.46) and (1.48) we note that

\begin{equation}
\frac{\varphi^2_m}{\sin^2 \varphi_m} = \varphi_m \mu(\varphi_m) + 1, \quad m = 1, 2, \ldots.
\end{equation}

Hence (1.46) becomes

\begin{equation}
\frac{(2a\theta_m)^2}{\sin^2(2a\theta_m)} - \varphi_m \mu(\varphi_m) - 1 = t'_m \mu(2a\theta'_m) - t'_m \mu(\varphi_m),
\end{equation}

and, since $\varphi_m < t'_m$, one has

\begin{equation}
\frac{(2a\theta'_m)^2}{\sin^2(2a\theta'_m)} < t'_m \mu(2a\theta'_m) + 1.
\end{equation}

Analogously, (1.48) implies

\begin{equation}
\frac{(2a\theta'_{m+1})^2}{\sin^2(2a\theta'_{m+1})} > t'_{m+1} \mu(2a\theta'_{m+1}) + 1
\end{equation}

which yields

\begin{equation}
\frac{(2a\theta'_m)^2}{\sin^2(2a\theta'_m)} < \frac{(2a\theta'_{m+1})^2}{\sin^2(2a\theta'_{m+1})}.
\end{equation}

Consequently in all cases the action increases with $\theta$ for solutions of (1.37). This completes the proof. \qed

\textbf{Proof of Theorem 1.41.} – We need to find the Hamiltonian paths connecting $(0, 0)$ to $(0, t)$, i.e. $x(1) = 0$, $t(1) = t$. Formula (1.15) implies that $\xi(1) = \xi(0)$, otherwise arbitrary, and then (1.14) implies that $e^{2\theta A} = I$. This and (1.20) yield

\begin{equation}
2a\theta = m \pi, \quad m = 1, 2, \ldots;
\end{equation}

note that according to (1.16) we have

\begin{equation}
t = \frac{H_0}{\theta}.
\end{equation}
therefore \( \theta \neq 0 \) and \( m \neq 0 \) in (1.52). Also

\[
d_m^2 = 2H_0 = \frac{m\pi|t|}{a},
\]

depends on \( |t| \). Since \( x = 0 \) we use \( \zeta(0) \) and (1.15) to obtain the geodesics:

\[
x^{(m)}(s) = \left( \frac{t}{am\pi} \right)^{1/2} \exp \left( \frac{m\pi}{2a} \Delta x \right) \frac{\zeta(0)}{\|\zeta(0)\|}, \quad \|\zeta(0)\| = \sqrt{2H_0},
\]

where \( \zeta(0) \in \mathbb{R}^2 \) is arbitrary, and substituting (1.54) into (1.16) yields

\[
t^{(m)}(s) = \left[ 2m\pi s - \sin(2m\pi s) \right] \frac{t}{2m\pi}.
\]

This shows that for each fixed \( m, m = 1, 2, \ldots \), the geodesics \( (x^{(m)}(s), t^{(m)}(s)) \) may be parametrized by \( \zeta(0)/\|\zeta(0)\| \in \mathbb{S}^1 \). (Note that these curves lie in a cylinder around the \( t \)-axis whose radius is \( O(1/\sqrt{m}) \).) This completes the proof of Theorem 1.41.

**Theorem 1.56.** – The Carnot–Caratheodory distance \( d(x, t) \) is real analytic at points \( (x, t) \), \( x \neq 0 \), and is homogeneous of degree 1 with respect to the dilations:

\[
(x, t) \rightarrow (\lambda x, \lambda^2 t), \quad \lambda > 0.
\]

The gradient \( \partial d/\partial x \) is discontinuous at every point \( (0, t), t \neq 0 \), although \( d(x, t) \) is continuous at all such points. \( d(x, t) \) increases to \( d(0, t) \) as \( x \rightarrow 0 \).

**Proof.** – The function \( \mu \) has an analytic inverse on the interval \((-\pi, \pi)\), so \( \theta_c \) depends analytically on \( (x, t) \), \( x \neq 0 \). It follows from (1.40) that \( d(x, t) \) is analytic for \( x \neq 0 \). The quotient \( |t|/a\|x\|^2 \) is homogeneous of degree 0 with respect to the dilations (1.57), so \( \theta_c(x, t) \) is homogeneous of degree 0 and \( d(x, t) \) is homogeneous of degree 1.

The identity (1.37) implies that the gradient \( \partial \theta_c/\partial x \) satisfies

\[
a\|x\|^2 \mu'(2a\theta_c) \frac{\partial \theta_c}{\partial x} + \mu(2a\theta_c)x = 0.
\]

Combining this with the third identity in (1.30) gives

\[
\frac{\partial S}{\partial x} = (2a\theta_c) \cot(2a\theta_c)x.
\]

As \( x \rightarrow 0 \) with \( t \neq 0 \) fixed, \( 2a\theta_c \rightarrow \pi \) and (1.37) shows that

\[
\sin(2a\theta_c) \sim \left( \frac{a\pi}{|t|} \right)^{1/2} \|x\|.
\]

It follows that

\[
\frac{\partial S}{\partial x}(x, t) = -\left( \frac{\pi|t|}{a} \right)^{1/2} \|x\|^{-1}x + o(1), \quad x \rightarrow 0.
\]

This implies that the limit of the gradient of \( d(x, t) \) at \( x = 0 \) depends on the direction of approach. Again, letting \( x \rightarrow 0 \), \( 2a\theta_c \rightarrow \pi \) and (1.38) converges to \( d(0, t)^2 \), see (1.42). This proves the
continuity of \( d(x, t) \) at points \((0, t), t \neq 0\). Finally, (1.61) shows that \( S \) decreases in every \( x \)-direction away from the \( t \)-axis, and we have completed the proof of Theorem 1.56. \( \square \)

The complex action

The condition \( t(0) = 0 \) of the previous section determines the classical action \( S \). For reasons that will become clear in the next section, here we consider a complex version of Hamiltonian mechanics. Instead of the condition \( t(0) = 0 \), we simply fix

\[
\theta = -i.
\]

We modify (1.27) and take the complex action integral to be

\[
g(x, t, \tau) = -it + \int_0^\tau \{ i\dot{x}, \xi \} - H \} \, ds
\]

\[
= \int_0^\tau \{ i\dot{x}, \xi \} + i(-i) - H \} \, ds + t(0)(-i)
\]

\[
= S(x, t, \tau; -i) + t(0)(-i)
\]

\[
= -it + \frac{1}{2} \{ iA \coth(iA)x, x \}
\]

\[
= -it + a \coth(2at)\|x\|^2,
\]

where we made use of (1.30). Like the classical action, the complex action \( g \) satisfies the Hamilton–Jacobi equation:

\[
0 = \frac{\partial g}{\partial \tau} + H \left( x, \frac{\partial g}{\partial x} \right) = \frac{\partial g}{\partial \tau} + \frac{1}{2} (X_1 g)^2 + \frac{1}{2} (X_2 g)^2.
\]

This identity follows from a direct calculation, or from an adaptation of the classical argument [5]. Moreover, as in the case of the classical action, the derivatives of the complex action are given by:

\[
\frac{\partial g}{\partial x} = \xi(\tau), \quad \frac{\partial g}{\partial t} = \theta.
\]

The second identity is self-evident. To obtain the first identity we use (1.9), (1.15) and (1.14) to obtain \( \xi(\tau) \) as follows:

\[
\xi(\tau) = \xi(\tau) - \theta Ax(\tau) = \xi(\tau) - \frac{1}{2} (\xi(\tau) - \xi(0))
\]

\[
= \frac{1}{2} (e^{2i\tau A} + 1) \xi(0) = \theta A \coth(\tau A)\|x\|
\]

\[
= 2a \coth(2at)\|x\|
\]

which agrees with \( \partial g/\partial x \) since \( \theta = -i \). Therefore \( H(\cdot, \cdot, \theta) = \frac{1}{2} (\xi(\tau), \xi(\tau)) = H_0 \), so the Hamilton–Jacobi equation becomes

\[
0 = \frac{\partial g}{\partial \tau} + H_0 = \frac{\partial g}{\partial \tau} + \frac{S}{\tau} = \frac{\partial g}{\partial \tau} + \frac{1}{\tau} \left\{ g(x, t, \tau) + it(0) \right\}.
\]
see (1.28) and (1.62). Note that \( g(x,t,\tau) \) is a meromorphic function of \( \tau \) in the complex plane, with poles only on the imaginary axis.

**Theorem 1.66.** Suppose \( x \neq 0 \). Then the unique critical point with respect to \( \tau \) of the modified complex action function

\[
f(x,t,\tau) = \tau g(x,t,\tau) = -\text{i} \tau t + a \tau \coth(2a\tau)\|x\|^2
\]

in the strip \( \{|\Im \tau| < \pi/2a\} \) is the point \( \tau_c(x,t) = \text{i}\theta_c(x,t) \), where \( \theta_c \) is the solution of (1.32) in this interval. At the critical point

\[
f(x,t,\tau_c(x,t)) = S(x,t,1;\theta_c) = \frac{1}{2} d(x,t)^2.
\]

The identity (1.68) is also valid at points \((0,t), \ t \neq 0\).

**Proof.** The second equality in (1.68) is just (1.40). Also (1.65) yields

\[
\frac{\partial f}{\partial \tau} = \frac{\partial g}{\partial \tau} = g + \tau \frac{\partial g}{\partial t} = g - \left\{ g + \text{i} \tau (0;x,t,\tau,-\text{i}) \right\} = -\text{i}[t-a\mu(2a(-\text{i})\|x\|^2)]
\]

(1.69)

\[
= -\text{i}[t-a\mu(2a(-\text{i})\|x\|^2)]
\]

see (1.25), and the critical points yield Hamiltonian paths with \( t(0) = 0 \). When \( t(0) = 0 \), then the third and fourth equality in (1.62) implies:

\[
f(x,t,\tau_c) = S(x,t,\tau_c; -\text{i}) = S(x,t,1,-\text{i}\tau_c).
\]

If \( \tau_c \) is purely imaginary, then (1.69) shows that \( \tau_c(x,t) = \text{i}\theta_c(x,t) \), where \( \theta_c(x,t) \) is the unique solution of (1.32) in \((-\pi/2a, \pi/2a)\). In this case (1.70) reduces to the first equality in (1.68).

To see that critical points in the strip can occur only on the imaginary axis, we shall show that for real \( \tau \) and fixed imaginary \( \text{i}b \) with \(|2ab| < \pi\), the real part of \( f(x,t,\tau + \text{i}b) \) increases strictly with \(|\tau|\). In particular

\[
\Re \left[ f(x,t,\tau + \text{i}b) - f(x,t,\text{i}b) \right] = \Re \left[ a(\text{i}b) \coth(2a(\text{i}b)) - \text{i}ab \coth(2ab) \right] \|x\|^2.
\]

(1.71)

We note that

\[
\coth(u + \text{i}v) = \frac{\cosh(u \sinh(v) - \text{i} \sin(v) \cos(v))}{\sinh^2(u) + \sin^2(v)}; \qquad \text{i}v \coth(\text{i}v) = v \cot(v);
\]

then

\[
\Re \left[ (u + \text{i}v) \coth(u + \text{i}v) - \text{i}v \coth(\text{i}v) \right] = u \cosh(u \sinh(u) + \v \sin(v) \cos(v)) - v \cot(v)
\]

\[
= \frac{\sinh^2(u)}{\sinh^2(u) + \sin^2(v)} (u \coth(u) - v \cot(v)).
\]

(1.72)
Now \( v \cot v \leq 1 \) on the interval \( |v| \leq \pi \), while \( u \coth u \geq 1 \) for real \( u \), so the second factor on the right is positive except at \( u = v = 0 \). Both factors are strictly increasing with respect to \( u \). We substitute \( 2a(\tau + ib) \) for \( u + iv \) and multiply (1.72) by \( \| x \|^2 / 2 \) to obtain the desired result. □

An explicit calculation yields the following:

**Corollary 1.73.** All critical points of \( f(\tau) = f(x, t, \tau) \) on the imaginary \( \tau \)-axis are of the form \( \tau = i\theta \), where \( \theta \) is a solution of (1.32).

**2. The heat kernel on \( H_1 \)**

We write \( u \) for the time variable and consider the heat operator:

\[
\frac{\partial}{\partial u} - \Delta_H = \frac{\partial}{\partial u} - \frac{1}{2}(X_1^2 + X_2^2).
\]

The fundamental solution at the origin is the function \( P(x, t; u) \) defined on \( H_1 \times R_+ \) such that

\[
\frac{\partial P}{\partial u} - \Delta_H P = 0, \quad u > 0,
\]

\[
\lim_{u \to 0} P(x, t; u) = \delta(x)\delta(t).
\]

It is known that \( P \) has the form

\[
P(x, t; u) = \frac{1}{(2\pi u)^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-f(x, t, \tau)/u} V(\tau) \, d\tau,
\]

where \( f \) is the modified complex action function of (1.67) and

\[
V(\tau) = \frac{2a\tau}{\sinh(2a\tau)}.
\]

Versions of this result are proved in [9,20,28]. A more general result is derived from the point of view of complex Hamiltonian mechanics in [5]. We sketch the derivation here. It is based on the Hamilton–Jacobi equation (1.63) for \( g = f/\tau \), which implies:

\[
H\left(x, \frac{\partial f}{\partial x}\right) = \tau^2 H\left(x, \frac{\partial g}{\partial x}\right) = -\tau^2 \frac{\partial g}{\partial \tau} = f - \tau \frac{\partial f}{\partial \tau}.
\]

Then a straightforward calculation using (2.6) yields

\[
\begin{align*}
\left(\Delta_H - \frac{\partial}{\partial u}\right) e^{-f/u} / u^2 &= e^{-f/u} / u^2 \left[H\left(x, \frac{\partial f}{\partial x}\right) - f\right] - e^{-f/u} / u^3 [\Delta_H f - 2] \\
&= -\frac{e^{-f/u}}{u^3} \tau \frac{\partial f}{\partial \tau} - \frac{e^{-f/u}}{u^3} [\Delta_H f - 2],
\end{align*}
\]

which gives:
\[
\left( \Delta_H - \frac{\partial}{\partial u} \right) u^{-2} e^{-f/u} V
\]
\[
= -u^{-3} e^{-f/u} \left[ \frac{dV}{d\tau} + (\Delta_H f - 1)V \right] + \frac{\partial}{\partial \tau} \left[ u^{-3} e^{-f/u} \tau V \right].
\]

Thus, assuming integrability of the last expression and the vanishing of \( u^{-3} e^{-f/u} \tau V \) at \( \tau = \pm \infty \), the integral (2.4) will satisfy the heat equation provided that
\[
\tau \frac{dV}{d\tau} + (\Delta_H f - 1)V = 0.
\]

If we set \( V = \tau v \) then (2.9) becomes
\[
\frac{dv}{d\tau} + (\Delta_H g)v = 0.
\]

It is easily verified that (2.9), together with the initial condition \( V(0) = 1 \), gives
\[
v(\tau) = \frac{2a}{\sinh(2a\tau)},
\]
see (2.5.8) of [5]. Then (2.8) can be integrated to show that \( P \) defined by (2.4) satisfies the heat equation (2.2).

To prove (2.3) we note first that if \( 0 < |2a\gamma| < \pi \), then
\[
\text{Re } f(x, t, \tau + i\gamma \text{sgn } t) \geq \varepsilon(\gamma)(\|x\|^4 + t^2)^{1/2}
\]
for some \( \varepsilon(\gamma) > 0 \); see (2.7.7) of [5]. Consequently \( P \) can be put in the form:
\[
P(x, t; u) = \frac{1}{(2\pi u)^2} \int_{-\infty}^{\infty} e^{-f(x, t, \tau)/u} V(\tau) d\tau.
\]

Note that \( V \) is integrable (in fact exponentially decreasing) on the contour in (2.13). The following is a consequence of this fact and of (2.12).

**Lemma 2.14.** – The function \( P \) is smooth on \( H_1 \times \mathbb{R}_+ \) and satisfies
\[
\lim_{u \to 0} \left( \frac{d}{dt} \right)^m P(x, t; u) = 0,
\]
uniformly in \( (x, t) \) in compact subsets of \( H_1 \setminus \{0\} \), for each \( m = 0, 1, 2, \ldots \).

To derive (2.3) we need to compute
\[
\lim_{u \to 0} \int_{\mathbb{R}^3} P(x, t; u) \, dx \, dt.
\]
This was done in [5], see (2.8.13)–(2.8.15). In fact
\[
\int_{\mathbb{R}^3} P(x, t; u) \, dx \, dt = 1.
\]
Finally we note that the probabilistic derivation of (2.4), given in [20] shows that $P$ is positive.

We have proved that $P(x, t; u)$ is the heat kernel relative to the origin. It follows from left invariance of $\Delta_H$ that the full heat kernel is obtained by left translations. Equivalently, Heisenberg convolution with $P$ gives the solution of the usual initial-value problem for the heat operator (2.1) in $H_1$.

A sharp global estimate for the heat kernel

In this section we prove the following estimate. The asymptotic results that are obtained in the next section show that the estimate is essentially sharp.

Theorem 2.17. – The heat kernel $P(x, t; u)$ satisfies the estimate:

$$P(x, t; u) \leq C \frac{e^{-d(x, t)^2 / 2u}}{u^2} \min \left\{ 1, \left( \frac{u}{\|x\| d(x, t)} \right)^{1/2} \right\}, \quad (x, t, u) \in H_1 \times \mathbb{R}_+.$$  \hspace{1cm} (2.18)

By continuity, it is enough to prove the estimate (2.18) when $x \neq 0$. For convenience, we assume also that $t > 0$. As in Theorem 1.36 we denote by $\theta_c = \theta_c(x, t)$ the unique solution of (1.37) in the interval $[0, \pi / 2a)$. It follows from the identity (1.32) that

$$(\pi - 2a\theta_c)^2 \sim (\pi - 2a\theta_c)^2 \left[ 1 + O(\pi - 2a\theta_c) \right] = \frac{\pi a \|x\|^2}{t}, \quad \text{as } \theta_c \to \frac{\pi}{2a}. \hspace{1cm} (2.19)$$

Lemma 2.20. – For any $\varepsilon_0 > 0$, the estimate (2.18) is valid for $(x, t)$ such that

$$2a\theta_c(x, t) \leq \pi - \varepsilon_0; \hspace{1cm} (2.21)$$

that is $t / \|x\|^2 \leq C = C(\varepsilon_0) < \infty$, with a constant depending only on $\varepsilon_0$.

Proof. – The contour for the integral (2.4) may be moved to the line $\text{Im} \tau = \theta_c$:

$$P(x, t; u) = \frac{1}{(2\pi u)^2} \int_{\text{Im} \tau = \theta_c} e^{-f(x, t, \tau)/u} V(\tau) \, d\tau. \hspace{1cm} (2.22)$$

We know from the proof of Theorem 1.66 that, on this contour, $\text{Re} f$ has a strict minimum at $\tau = i\theta_c$, and that its value there is $d(x, t)^2 / 2$. Therefore we have:

$$P(x, t; u) \leq \frac{e^{-d(x, t)^2 / 2u}}{(2\pi u)^2} \int_{\text{Re} \tau = 0} \left| V(x + i\theta_c) \right| \, dx. \hspace{1cm} (2.23)$$

To improve the estimate when $\|x\|^2 / u$ is large, we observe that the first derivative of $f$ vanishes at $\tau = i\theta_c$, while the second derivative is

$$\frac{\partial^2 f}{\partial \tau^2} \bigg|_{\tau = 0} = \frac{4a^2}{\sin^2(2a\theta_c)} \left\{ 1 - 2a\theta_c \cot(2a\theta_c) \right\} \|x\|^2 \geq \frac{\|2ax\|^2}{3}. \hspace{1cm} (2.24)$$

see (1.35). It follows that

$$\text{Re} f(x, t, \tau) \geq \frac{d(x, t)^2}{2} + \frac{\|2ax\|^2}{4}. \hspace{1cm} (2.25)$$
over an interval \([-\delta, \delta]\), \(\delta = \delta(\epsilon_0) > 0\). Therefore, in view of (2.25) and the discussion leading up to (1.71), we have an estimate:

\[
P(x, t; u) \leq \frac{e^{-d(x, t)^2/2u}}{(2\pi u)^2} \left\{ \int_{-\delta}^{\delta} e^{-x^2/4u} |V(s + i\theta_x)| ds + \int_{|x| > \delta} e^{-|x|^3/6u} |V(s + i\theta_x)| ds \right\}
\]

(2.26)

\[
\leq C e^{-d(x, t)^2/2u} \left\{ \int e^{-x^2/4u} ds + e^{-|x|^3/6u} \right\}
\]

(2.27)

where \(c = c(\epsilon_0)\) is positive. In the region under consideration (1.14) implies that \(\|x\|\) is comparable to \(\|x\|d(x, t)\|^{1/2}\), and thus (2.26) and (2.23) imply (2.18).

Proof of Theorem 2.17. – The estimates obtained in the proof of Lemma 2.20 blow up as \(2a\theta_x \to \pi\) or, equivalently, as \(t/\|x\|^2 \to \infty\) because both \(f\) and \(V\) have a pole at \(2a\pi = i\pi\). To control the situation we replace the contour in (2.22) by a circle around \(\pi i/2a\) of radius \(\pi/2a - \theta_x\), together with the line \(\pi i/2a\Im\tau = \lambda\pi\):

\[
P(x, t; u) = \frac{1}{(2\pi u)^2} \left\{ \int_{|\tau - \pi i/2a| = |\theta_x - \pi i/2a|} e^{-f(x, t, \tau)/u} V(\tau) d\tau + \int_{2a\Im\tau = \lambda\pi} e^{-f(x, t, \tau)/u} V(\tau) d\tau \right\}
\]

(2.27)

First we estimate \(P_1\), defined by the second integral in (2.27). We choose \(\lambda\) such that:

\[
1 < \lambda \leq \frac{3}{2a}
\]

so that the line and circle in (2.27) have no point in common. Assume that \(t \geq 0\). Then (1.72) yields

\[
\Re f \left( x, t, s + \frac{i\pi}{2a} \right) = \frac{2a s \sinh(4as) + \lambda \pi \sin(2\lambda \pi) |s|^2}{\sinh^2(2as) + \sin^2(\lambda \pi)} + \frac{\lambda \pi t}{2a} \geq \frac{\lambda \pi t}{2a}
\]

(2.29)

It follows that

\[
P_1(x, t; u) \leq \frac{e^{-\lambda \pi t/2au}}{(2\pi u)^2} \int_{\mathbb{R}} |V(s + i\lambda \pi/2a)| ds.
\]

(2.30)

Now \(2a\theta_x \to \pi\) if and only if \(t/\|x\|^2 \to \infty\), and by continuity of \(d(x, t)\) this implies that \(d(x, t)/d(0, t) = d(x/\sqrt{T}, 1)/d(0, 1)\) increases to 1 as \((x, t) \to (0, t)\), uniformly in \((x, t)\) as
By Theorem 1.41 one has \( d(0, t)^2 = \pi t/a \). Therefore,

\[
P_1(x, t; u) \leq C e^{-d(x, t)^2/2u} e^{-(\lambda - 1)d(x, t)^2/2u}.
\]

Here we note that \( d(x, t)^2 = 2S(x, t; 1; \theta_c) \) and then Theorem 1.29 implies that \( d(x, t) \geq \|x\| \), with equality at \( t = 0 \). Therefore when \( \|x\|d(x, t)/u \) is large, it follows that \( d^2/u \) is large and \( \exp(-d^2/u) \) is dominated by any positive power of \( u/d^2 \), which, in turn, is dominated by \( u/\|x\|d \). Thus we obtain an estimate of the form (2.18) for \( P_1 \).

We turn now to estimating \( P_0 \), defined by integration over a circle around \( i\pi/2a \). Here we set:

\[
2a\tau = \pi - i\zeta, \quad F = \frac{\pi \|x\|^2}{2},
\]

and write

\[
f = -\frac{F}{\zeta} + G(\zeta) + \frac{t}{2a}(\pi - \zeta),
\]

so that \( G(\zeta) = G(\zeta, x) = O(\|x\|^2) \) is holomorphic for \( \|\zeta\| < \pi \). The circle of integration is

\[
|\zeta| = \varepsilon \equiv \pi - 2a\theta_c.
\]

Set \( \zeta_c = \varepsilon \) so that \( 2a\tau_c = i\pi - i\zeta_c \); thus

\[
0 = \frac{\partial f}{\partial \tau}(\tau_c) = \frac{F}{\zeta_c} + G'(\zeta_c) - \frac{t}{2a}.
\]

It follows that:

\[
f - f_c = -\frac{F}{\zeta} + \frac{F}{\zeta_c} + G(\zeta) - G(\zeta_c) - \frac{t}{2a}(\zeta - \zeta_c)
\]

\[
= -\frac{F}{\zeta} + \frac{F}{\zeta_c} + \left(G'(\zeta_c) - \frac{t}{2a}\right)(\zeta - \zeta_c) + O(\|x\|^2|\zeta - \zeta_c|)
\]

\[
= -\frac{F}{\zeta} + \frac{F}{\zeta_c} - \frac{F}{\zeta_c^2}(\zeta - \zeta_c) + O(\|x\|^2|\zeta - \zeta_c|)
\]

\[
= \frac{F}{\zeta_c} \left(1 - \frac{\zeta}{\zeta_c}\right) \left(1 - \frac{\zeta_c}{\zeta}\right) + O(\|x\|^2|\zeta - \zeta_c|),
\]

uniformly for \( \varepsilon \leq \pi/2 \). We set \( \zeta = \varepsilon e^{i\varphi} \) so that (2.36) becomes

\[
f - f_c = \frac{\pi \|x\|^2}{\varepsilon} (1 - \cos \varphi) + O(\|x\|^2\varepsilon^2(1 - \cos \varphi)).
\]

By (2.19), \( \varepsilon^2 \sim a\pi \|x\|^2/t \), so (2.37) can be written as

\[
f - f_c \sim \left(\frac{\varepsilon}{a\pi} + O(\varepsilon^4)\right) t (1 - \cos \varphi).
\]

In particular, we conclude that for some \( \varepsilon_0 > 0 \):

\[
\Re f \geq f_c \text{ if } |\pi - 2a\tau| = |i\pi - 2a\tau_c| \leq \varepsilon_0.
\]
Although $|V(\tau)| = O(1/\varepsilon)$ on the circle $|i\pi - 2a\pi| = \varepsilon$, the circle itself has length $2\pi\varepsilon$.
Therefore we obtain the estimate
\begin{equation}
P_0(x, t; u) \leq C e^{-d(x, t)^2/2u}.
\end{equation}

Once again, this can be improved if $\|x\|d(x, t)$ is large. In fact $1 - \cos \varphi$ is bounded from above by $\frac{\varphi^2}{2}$ on $|\varphi| \leq \pi$, so (2.37) implies that the integral that occurs in the definition of $P_0$ is dominated by:
\begin{align}
e^{-d(x, t)^2/2u} \int_{\mathbb{R}} e^{-c\|x\|^2\varphi^2/2u} |d\tau| &\leq e^{-d(x, t)^2/2u} \int_{-\pi}^{\pi} e^{-c\|x\|^2\varphi^2/2u} d\varphi \\
&\leq C e^{-d(x, t)^2/2u} \sqrt{8u} \
\end{align}

In this range $\epsilon^2 \sim \|x\|^2/t \sim \|x\|^2/d(x, t)^2$ by (2.19) and (1.42), so we obtain an estimate of the form (2.18) for $P_0$. The proof of Theorem 2.17 is complete. \(\square\)

**Small time behavior of the heat kernel**

To describe the small time behaviour of the heat kernel it is necessary to distinguish two cases.

**Theorem 2.42.** Given a fixed point $(x, t)$, $x \neq 0$, let $\theta_c$ denote the solution of (1.37) in the interval $[0, \pi/2a]$; then
\begin{equation}
P(x, t; u) = \frac{1}{(2\pi u)^{3/2}} e^{-d(x, t)^2/2u} \left\{ \Theta(x, t) \sqrt{2\pi u} + O(u) \right\},
\end{equation}
as $u \to 0^+$, where
\begin{equation}
\Theta(x, t) = \left(\frac{1}{f''(i\theta_c)}\right)^{1/2} V(i\theta_c) = \frac{\theta_c}{\sqrt{|1 - 2a\theta_c \cot(2a\theta_c)|^{1/2}}}.
\end{equation}

\(f''(\tau) = d^2 f/d\tau^2\). When $|t|/\|x\|^2$ is large $\Theta(x, t)$ has the following behaviour
\begin{equation}
\Theta(x, t) = \frac{\pi^{3/4}}{2a^{3/4}\|x\|^{1/2}|t|^{1/4}} \left\{ 1 + O\left(\frac{\|x\|}{|t|^{1/4}}\right) \right\}.
\end{equation}

We note that when $|t|/\|x\|^2$ is large, $\|x\|^{1/2}|t|^{1/4}$ is comparable to $\|x\|^{1/2}d(x, t)^{1/2}$, so the asymptotics (2.43) correspond exactly to the estimate (2.18).

**Theorem 2.46.** At points $(0, t)$, $t \neq 0$,
\begin{equation}
P(0, t; u) = \frac{1}{4au^{3/2}} e^{-d(0, t)^2/2u} \left\{ 1 + O\left(e^{-d(0, t)^2/2u}\right) \right\}
\end{equation}
as $u \to 0^+$.

**Proof of Theorem 2.42.** This result is a consequence of the representation (2.22), the identity (2.24), and a stationary phase argument. Indeed we know that on the line $\tau = s + i\theta_c$, $s \in \mathbb{R}$, the
real part of $f$ attains its global minimum only at $s = 0$; it is a strictly increasing function of $|s|$. To be more precise, we set:

\[(2.48) \quad \Phi(s) = f(x, t, s + i\theta_c) - f(x, t, i\theta_c),\]

and note that (1.72) yields the estimate

\[\text{Re } \Phi(s) \geq C |s|^2 \|x\|^2, \quad |s| \leq 1,\]

with a positive constant $c$. With this information in hand we consider the following integral:

\[(2.49) \quad I = \int_{-\infty}^{\infty} e^{-\Phi(s)/u} V(s + i\theta_c) \, ds = \left\{ \int_{\delta}^{\delta} + \int_{|s| > \delta} \right\} e^{-\Phi(s)/u} V(s + i\theta_c) \, ds = I_\delta + I_\delta',\]

where $\delta$ is to be chosen, $0 < \delta \leq 1$. First

\[(2.50) \quad |I_\delta'| \leq e^{-\text{Re } \Phi(\delta)/u} \int_{\mathbb{R}} |V(s + i\theta_c)| \, ds \leq C e^{-\text{Re } \Phi(\delta)/u}, \quad C = C(|t|/\|x\|^2) > 0.\]

To evaluate the limit of $I_\delta$ as $u \to 0+$, we use the method of stationary phase. To this end we note that $\Phi(0) = 0$, $\Phi'(0) = f'(i\theta_c) = 0$, and

\[(2.51) \quad \Phi''(0) = f''(i\theta_c) \geq \frac{2a\|x\|^2}{3},\]

which is just (2.24). Consequently,

\[(2.52) \quad \Phi(s) = \Phi''(0) \frac{s^2}{2} + O(|s|^3 \|x\|^2) = \Phi''(0) \frac{s^2}{2} (1 + O(|s|)).\]

Thus we may choose a $\delta > 0$, $0 < \delta < \pi/2a - \theta_c$, and introduce a new variable $z = s + O(s^2)$, such that:

\[\Phi(s) = \Phi''(0) \frac{z^2}{2}, \quad |z| < \delta;\]

then

\[(2.53) \quad I_\delta = \int_{z(-\delta)}^{z(\delta)} \exp(-\Phi''(0)z^2/2u) V(s(z) + i\theta_c) \frac{ds}{dz} \, dz.\]

The contour of integration in (2.53) may be complex. On the other hand the integrand is holomorphic in $z$, so by moving the contour to the real axis, the error we commit is exponentially small in $u$; we note that if $z = \sigma + i\gamma$ is on the path of integration in (2.53), then $|\gamma| < c\sigma^2$. In fact we have

\[I_\delta = \int_{-\delta}^{\delta} \exp(-\Phi''(0)z^2/2u) V(s(z) + i\theta_c) \frac{ds}{dz} \, dz + O(\exp(-c/\sigma))\]
\[ P(x, t; u) = \frac{e^{-d(x, t)^2/2u}}{(2\pi u)^{1/2}} \left( \frac{2\pi u}{f''(i\theta_c)} \right)^{1/2} V(i\theta_c) + O(u) \]
Therefore (2.4) is the sum of the residues:

\[ P(0, t; u) = \frac{1}{(2\pi u)^2} \int_{\mathbb{R}} e^{it/u} V(\tau) \, d\tau \]

\[ = \frac{2\pi i}{(2\pi u)^2} \sum_{m=1}^{\infty} (-1)^m \frac{i m \pi}{2a} e^{-m\pi t/2au} \]

\[ = \frac{1}{4au^2} \left( 1 + e^{-\pi t/2au} \right)^2 \]

\[ = \frac{1}{4au^2} e^{-\pi t/2au} \left[ 1 + O(e^{-\pi t/2au}) \right]. \]

(2.59)

According to Theorem 1.41, \( \pi |t|/a = d(0, t)^2 \).

3. Hamiltonian mechanics and geometry on higher dimensional Heisenberg groups

We turn now to the study of the \((2n + 1)\)-dimensional Heisenberg group \( H_n \), normalized as follows. We equip \( \mathbb{R}^{2n} \times \mathbb{R} \) with the group law

\[ (x, t) \circ (x', t') = \left( x + x', t + t' + 2 \sum_{j=1}^{n} a_j \left[ x_{2j} x'_{2j-1} - x_{2j-1} x'_{2j} \right] \right), \]

where \( a_1, a_2, \ldots, a_n \) are positive constants, numbered so that

\[ 0 < a_1 \leq a_2 \leq \cdots \leq a_p < a_{p+1} = \cdots = a_n. \]

The vector fields

\[ X_{2j-1} = \frac{\partial}{\partial x_{2j-1}} + 2a_j x_{2j} \frac{\partial}{\partial t}, \]

\[ X_{2j} = \frac{\partial}{\partial x_{2j}} - 2a_j x_{2j-1} \frac{\partial}{\partial t}, \]

are left invariant and generate the Lie algebra. The associated Heisenberg (sub-)Laplacian is

\[ \Delta_H = \frac{1}{2} \sum_{j=1}^{2n} X_j^2. \]

We consider Hamiltonian mechanics associated to the symbol of \(-\Delta_H:\)

\[ H(x, \xi, \theta) = \frac{1}{2} \sum_{j=1}^{n} \left[ (\xi_{2j-1} + 2a_j x_{2j} \theta)^2 + (\xi_{2j} - 2a_j x_{2j-1} \theta)^2 \right]. \]

(3.5)

Again we adopt vector notation: \( \partial/\partial x \) will denote the gradient in \( \mathbb{R}^{2n} \) and \( \langle , \rangle \) will denote the inner product by:

\[ \langle x, y \rangle = \sum_{j=1}^{2n} x_j y_j. \]
and also its (bilinear) extension to $C^{2n}$.

Setting

$$A = \text{diag}[A_1, \ldots, A_n], \quad A_j = \begin{pmatrix} 0 & 2a_j \\ -2a_j & 0 \end{pmatrix},$$

we may write the group law (3.1) in the form

$$\langle x, t \rangle \circ \langle x', t' \rangle = \langle x + x', t + t' + \langle Ax, x' \rangle \rangle,$$

where we note in particular that

$$A + A' = 0,$$

$$A^2 = \text{diag}(-4a_1^2, -4a_2^2, \ldots, -4a_n^2).$$

Similarly the vector fields (3.3) may be denoted collectively by

$$X = \frac{\partial}{\partial x} + Ax \frac{\partial}{\partial t},$$

and the Hamiltonian (3.5) by

$$H(x, \xi, \theta) = \frac{1}{2} \langle \xi + \theta Ax, \xi + \theta Ax \rangle = \frac{1}{2} \langle \xi, \xi \rangle,$$

where $\xi = \xi + \theta Ax$.

Hamilton’s equations for a curve take the form (1.10) of Section 1, and the solution with boundary conditions (1.11) also takes essentially the same form.

**Theorem 3.12.** – The solution of (1.10) with boundary conditions (1.11) is:

$$x(s) = e^{(s - \tau) A} \sinh(s \theta A) \frac{\sinh(\tau \theta A)}{\sinh(\tau \theta A)} x;$$

$$t - t(s) = \frac{\tau - s}{2\theta} \left( \frac{\langle A \rangle^2}{\sinh^2(\tau \theta A)} x, x \right) - \frac{1}{4\theta} \left[ \frac{\sinh(2\tau \theta A) - \sinh(2s \theta A)}{\sinh^2(\tau \theta A)} \right] x, x$$

$$= (\tau - s) \sum_{j=1}^{n} \frac{2a_j^2 \theta}{\sin^2(2a_j \tau \theta)} r_j^2$$

$$- \sum_{j=1}^{n} \frac{a_j}{2} \frac{\sin(4a_j \tau \theta) - \sin(4a_j s \theta)}{\sin^2(2a_j \tau \theta)} \frac{r_j^2}{r_j^2}, \quad r_j^2 = x_{j-1}^2 + x_j^2.$$
where $\mu$ is defined by (1.26). The action integral is:

$$S(x, t; \tau; \theta) = \int_0^\tau \left(\|\dot{\xi}(s)\|^2 + \theta \dot{i}(s) - H(x(s), \xi(s), \theta)\right) ds = \tau H_0.$$ (3.17)

Again there are various forms for the action.

**Theorem 3.18.** – The action integral $S(x, t; \tau; \theta)$ is given by:

$$S(x, t; \tau; \theta) = \frac{1}{2} \left(\theta \Lambda^2 \left[\sinh(\theta \Lambda)\right]^{-2} x, x\right)$$

$$= \left[t - t(0)\right] \theta + \frac{1}{2} \left[\theta \Lambda \coth(\theta \Lambda) x, x\right]$$

$$= \sum_{j=1}^n \frac{\tau(2a_j \theta)^2}{2 \sin^2(2a_j \tau \theta)} r_j^2, \quad \theta \in [0, \pi/2a_n),$$

$$= \left[t - t(0)\right] \theta + \sum_{j=1}^n a_j \theta \cot(2a_j \tau \theta) r_j^2.$$ (3.19)

**The classical action and the Carnot–Carathéodory distance**

Once again we obtain classical Hamiltonian mechanics by completing the boundary conditions (1.11) with the condition $t(0) = 0$, and setting $\tau = 1$. Then necessarily

$$t = \sum_{j=1}^n a_j \mu(2a_j \theta) r_j^2.$$ (3.20)

The detailed behavior of the term on the right, as a function of $\theta$ depends both on the $r_j$ and on the distribution of the $a_j$. In the isotropic case $a_1 = a_2 = \cdots = a_n$, the results of Section 1 carry over with no change, except that each family of geodesics from $(0, 0)$ to $(0, t)$ is parametrized by the $(2n - 1)$-sphere.

It is convenient to adopt the following convention concerning the function $\mu$:

$$\mu(m \pi) = +\infty, \quad m = 1, 2, 3, \ldots, \text{ and } 0 \cdot \infty = 0.$$ (3.21)

In particular this means that:

$$\sum_{j=1}^n a_j \mu(2a_j \pi/2a_n) r_j^2 < +\infty \quad \text{if and only if} \quad r_j = 0 \text{ when } a_j = a_n.$$ (3.22)

We shall also set

$$x = (x', x''), \quad x'' = (x_{2p+1}, x_{2p+2}, \ldots, x_{2n}),$$

where $p$ is the index in (3.2).
Suppose $x'' \neq 0$. Then there are finitely many geodesics from $(0,0)$ to $(x,t)$. The geodesics are indexed by the solutions of
\begin{equation}
|t| = \sum_{j=1}^{n} a_j \mu(2a_j \theta) r_j^2,
\end{equation}
and their lengths increase with $\theta$. In particular, the Carnot–Caratheodory distance from $(x,t)$ to the origin is
\begin{equation}
d(x,t)^2 = 2S(x, |t|, 1; \theta),
\end{equation}
where $\theta_0$ is the unique solution of (3.25) in the interval $[0, \pi/2a]$. 

Proof. – Once again, the fact that geodesics are traces of Hamiltonian paths is a special case of results of Bismut [14], Strichartz [36], and the associated action is one-half the square of the length. Suppose that $x'' \neq 0$. The function on the right side of (3.25) is strictly increasing on $[0, \pi/2a]$, so there is a unique solution $\theta_0$ of (3.25) in the interval $[0, \pi/2a]$. In every other pole free interval the function on the right side of (3.25) decreases from $\pm \infty$ to a positive minimum at $\theta_m$, $m = 1, 2, \ldots$, then it increases to $\pm \infty$. Furthermore the right side of (3.25) is bounded from below by a straight line through the origin with a positive slope, since this is true for the individual summands, hence there are finitely many solutions of (3.25). We note that the minima of the right side do not necessarily increase with $m$. This is easily seen on $H_2$: choosing $a_1$ sufficiently near $a_2$, $a_1 < a_2$, we can arrange that:
\begin{equation}
\sum_{j=1}^{2} a_j \mu(2a_j \theta_1 r_j^2) > \sum_{j=1}^{2} a_j \mu(2a_j \theta_2 r_j^2).
\end{equation}

To complete the proof, we need to establish that the lengths of the geodesic associated to a solution of (3.25) increases with $\theta$. We did this directly for $H_1$ and it can be done directly for isotropic $H_n$, because the $\theta$ dependence in (3.25) is much simpler when the $a_j$ are identical. In the general case the detailed behaviour depends crucially on the relative sizes of the $a_j$ and on the relative sizes of the $r_j$. We approach the general case through a preliminary study of the modified complex action $f$.

Lemma 3.28. – For any $(x,t)$ with $x'' \neq 0$, and $t \geq 0$, the function $f(\tau) = f(x,t,\tau)$ 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 maximum for $f$.

Proof. – We set
\begin{equation}
F(\theta) = f(x,t,i\theta) = \sum_{j=1}^{n} a_j \theta \cot(2a_j \theta) r_j^2 + t \theta,
\end{equation}
and note that
\begin{equation}
F'(\theta) = t - \sum_{j=1}^{n} a_j \mu(2a_j \theta) r_j^2.
\end{equation}
Then Lemma 3.28 follows from properties of the function \( \mu \).

**Lemma 3.31.** For any \((x, t)\) with \(x'' \neq 0\) and \(t > 0\), there is exactly one branch of the set

\[
\Gamma_0 = \{ \tau : \text{Im} f(x, t, \tau) = 0, \ \text{Re} \tau > 0, \ \text{Im} \tau > 0 \}
\]

that goes to \(\infty\) in the quadrant \(\text{Re} \tau > 0, \ \text{Im} \tau > 0\). On this branch \(\text{Re} f\) increases as \(\tau \to \infty\).

**Proof.** Suppose \(s = i\theta, s > 0\); then

\[
\text{Im}(\tau \coth \tau) = \frac{\theta \sinh 2s - s \sin 2\theta}{2 \sinh^2 s + 2 \sin^2 \theta} > 0,
\]

because

\[
\theta \sinh 2s - s \sin 2\theta = \theta(\sinh 2s - 2s) + s(2\theta - \sin 2\theta) > 0.
\]

For any fixed \(\theta = \text{Im} \tau > 0\),

\[
\lim_{s \to +\infty} \text{Im}(\tau \coth \tau) = \theta, \quad \text{uniformly for bounded } \theta.
\]

Moreover, if \(\sin \theta = 0\), then

\[
\text{Im}(\tau \coth \tau) = \theta \coth s \geq \frac{\theta}{s}.
\]

The derivative

\[
\frac{\partial}{\partial s} \text{Im}(\tau \coth \tau) = -\frac{\cos \theta \sin \theta}{\sinh^2 s + \sin^2 \theta} \left[ -\theta \sinh^2 s \cos^2 \theta + \theta \cosh^2 s \sin^2 \theta + 2s \cosh s \sinh s \cos \theta \sin \theta \right] + O\left( \frac{s + \theta}{\sinh^2 s} \right)
\]

as \(s \to \infty\). Now,

\[
\text{Im} f(x, t, \tau) = -st + \text{Im} \sum_{j=1}^{n} 2a_j \tau \coth (2a_j \tau) \frac{r_j^2}{2}
\]

We choose a \(j\), such that \(r_j \neq 0\) and a very large \(\theta\), such that \(\sin(2a_j \theta) = 0\) and \(\coth(2a_j r_j \theta/2t)^{1/2} < 3/2\). In view of (3.34), (3.35) and (3.37), there will be at least one point \(\tau = s_0 + i\theta\), when \(\text{Im} f(x, t, \tau) = 0\). It follows from (3.35) and (3.37) that

\[
\frac{\theta \cdot r_j^2}{s_0} \leq s_0 t
\]

and (3.34), (3.37) and (3.38) imply that

\[
 s_0 t \leq \frac{3}{2} \sum_{k=1}^{n} 2a_k \theta \frac{r_j^2}{2}.
\]
so \( s_0 \) belongs to the following interval:

\[
(3.40) \quad \frac{r_j}{\sqrt{2t}} \sqrt{\theta} \leq s_0 \leq \frac{3}{2} \left( \sum_{k=1}^{n} a_k \frac{r_k^2}{t} \right) \theta.
\]

Now (3.36) and (3.40) yield the estimate

\[
(3.41) \quad \left| \frac{\partial}{\partial s} \text{Im} \left( \sum_{j=1}^{n} 2a_j \tau \coth(2a_j \tau) \frac{r_j^2}{2} \right) \right| \leq C \frac{s + \theta}{\sinh^2(2a_1s)} \leq C_1 \theta e^{-C_2 \sqrt{\theta}}, \quad s > s_0,
\]

where \( C, C_1 \) and \( C_2 \) are positive constants which depend only on the \( a_j \)'s and on \((x, t)\).

Consequently for suitably large \( \theta \)

\[
(3.42) \quad \frac{\partial}{\partial s} f(x, t, s + i\theta) < 0, \quad s > s_0,
\]

and there is only one solution \( s_0 \) of \( \text{Im} f(x, t, s + i\theta) = 0 \). It follows from (3.34) that no branch of \( \Gamma_0 \) can escape to \( \infty \) between two consecutive such lines \( \text{Im} \tau = \theta \), with \( \sin(2a_j \theta) = 0 \). Since \( \text{Im} f = 0 \) on the imaginary \( \tau \)-axis, a branch can escape from the quadrant through the imaginary \( \tau \)-axis only at a critical point of \( f \). By Lemma 3.28, there is no such escape for large \( \theta \). This still leaves the question of a possible branching of \( \Gamma_0 \) between two such consecutive \( \text{Im} \tau = \theta \) lines. Two such branches must join at the two \( s_0 \) points, since they are unique. This implies the existence of a bounded region \( \Omega \) on which \( \text{Im} f \) is harmonic and non-constant and vanishes on \( \partial \Omega \), which is a contradiction. Thus, for large \( |\tau| \), there is exactly one branch of \( \Gamma_0 \) that goes to infinity within the quadrant.

In particular, \( f \) has no critical points on this single branch of \( \Gamma_0 \). Therefore, on this branch, \( \text{Re} f \) must increase or decrease, it cannot have a stationary point. Now (1.72) implies that when \( \sin \theta = 0 \), then:

\[
(3.43) \quad \text{Re}(\tau \coth \tau) = s \coth s.
\]

Therefore, (3.40) and (3.43) yield

\[
(3.44) \quad \text{Re} f(x, t, s_0 + i\theta) = \theta t + \sum_{k=1}^{n} 2a_k s_0 \coth(2a_k s_0) \frac{r_k^2}{2},
\]

which increases as \( \theta \to \infty \). Consequently, \( \text{Re} f \) increases on the branch of \( \Gamma_0 \) which goes off to infinity, and we have derived Lemma 3.31.

**Lemma 3.45.** Assume \( x'' \neq 0 \) and \( t > 0 \). Let the critical points of \( f \) on the positive imaginary axis, counted according to multiplicity, be \( i\theta_1, \ldots, i\theta_{2m+1} \), with

\[
(3.46) \quad \theta_1 < \theta_2 \leq \theta_3 < \cdots < \theta_{2m} \leq \theta_{2m+1}.
\]

Let \( \Gamma' \) be the union of \( \Gamma_0 \) and the closed intervals

\[
(3.47) \quad [0, i\theta_1], [i\theta_2, i\theta_3], \ldots, [i\theta_{2m}, i\theta_{2m+1}].
\]

Then \( \Gamma' \), oriented in the direction of increasing \( \text{Re} f \), is a simple connected curve from 0 to \( \infty \).
Proof. – The set \( \Gamma \) is a union of analytic arcs that meet only at critical points of \( f \). At any critical point \( \tau \neq 0 \) there will be at least one branch that is oriented toward \( \tau \) and one that is oriented away from \( \tau \). The real part \( \text{Re} f \) increases along the branch of \( \Gamma' \) that leads to \( \infty \). It follows that any point of \( \Gamma \) is contained in at least one maximal oriented contour \( \Gamma_1 \subset \Gamma \). Moreover, \( \Gamma_1 \) must begin at 0 and eventually contain the branch that leads to \( \infty \). If \( \Gamma_2 \) is another such maximal oriented contour, either it coincides with \( \Gamma_1 \) or the union of the two contains a simple closed contour \( \Gamma_3 \). Suppose the latter; then \( \Gamma_3 \) encloses a bounded region \( \Omega \). But then \( \text{Im} f \) is harmonic and non-constant on \( \Omega \) and vanishes on the boundary, a contradiction. Thus there is a unique such maximal contour, and it coincides with \( \Gamma' \) itself. \( \Box \)

COROLLARY 3.48. – If \( f'(\tau) = 0 \) and \( \text{Im} f(\tau) = 0 \), then \( \tau \) is imaginary.

Proof of Theorem 3.24 concluded. – We continue to assume that \( t \) is positive and \( x_0 \neq 0 \). A solution of (3.25) corresponds to a critical point \( \theta = i\theta \) of the modified complex action function \( f \). The square of the length of the associated geodesic is \( 2f(\theta) \). Let \( \{i\theta_k\} \) be the critical point of \( f \) on the positive imaginary axis, numbered as in (3.46). To complete the proof we need to show that:

\[
(3.49) \quad f(i\theta_1) < f(i\theta_2) < \ldots < f(i\theta_{2m}) \leq f(i\theta_{2m+1}),
\]

with strict inequality where the corresponding inequality in (3.46) is strict. This is the same as saying that the critical points occur in this order on the oriented curve \( \Gamma' \) of Lemma 3.45. Any other ordering of the critical points along \( \Gamma \) would imply that \( \Gamma' \) has self-intersections, which contradicts Lemma 3.45. \( \Box \)

Remark 3.50. – One can show directly, that on every line \( \text{Im} \tau = \theta \), with \( i\theta \) not in any of the intervals (3.47) and \( \sin \theta \neq 0 \), there is an \( s \in (0, \infty) \), such that \( \text{Im} f(s + i\theta) = 0 \). In fact, assuming \( \sin \theta \neq 0 \), (3.36) yields

\[
(3.51) \quad \frac{\partial}{\partial s} \text{Im} f(x,t,s + i\theta) \bigg|_{s=0} = \sum_{j=1}^{n} a_j \mu(2a_j \theta) r_j^2 - t > 0, \quad \theta \notin (3.47).
\]

Therefore, if \( \theta \notin (3.47) \), \( \text{Im} f(x,t,s + i\theta) > 0 \) for small \( s \). On the other hand, by (3.34), \( \text{Im} f(x,t,s + i\theta) < 0 \) for large \( s \), so it must vanish for some \( s \in (0, \infty) \).

Next we study the geodesics when \( x'' = 0 \).

THEOREM 3.52. – Suppose \( x' \neq 0 \), \( x'' = 0 \).

(i) If

\[
(3.53) \quad |t| < \sum_{j=1}^{n} a_j \mu(2a_j \pi/2a_n) r_j^2,
\]

then the Carnot–Caratheodory distance \( d(x,t) \) is given by:

\[
(3.54) \quad d(x,t)^2 = 2S(x,t,1;\theta) = 2\theta \left[ |t| + \sum_{j=1}^{n} a_j \cot(2a_j \theta) r_j^2 \right],
\]

where \( \theta \) is the unique solution of (3.25) in \( (0, \pi/2a_n) \).
(ii) If (3.53) fails, then there are infinitely many shortest geodesics of the same length from 
(0, 0) to (x, t), parametrized by the sphere 

\[ \frac{1}{2} \| \zeta''(0) \| = \frac{\pi}{2a_n} \left[ |t| - \sum_{j=1}^{n} a_j \mu(2a_j \pi/2a_n) r_j \right]. \]  

where we set \( \zeta = (\zeta', \zeta''), \) \( \zeta'' = (\xi_{2p+1}, \ldots, \xi_{2p}). \) The length of these geodesics yield the 
Carnot–Caratheodory distance of \( (x, t) \) from the origin: 

\[ d(x, t) = 2S(x, t, 1; \pi/2a_n) \] 

\[ = \frac{2\pi}{2a_n} \left[ |t| + \sum_{j=1}^{n} a_j \cot(2a_j \pi/2a_n) r_j^2 \right]. \]

The distance \( d(x, t) \) is homogeneous of degree 1 with respect to the dilations (1.57), it is 
bounded above and below by multiples of \( |x| + |t|^{1/2} \) and it is a continuous function of 
\( (x, t), \) although it is not differentiable.

We note that it suffices to prove the result for \( t > 0, \) since for \( t < 0 \) we simply switch to 
\(-\pi/2a_n) in our arguments. We also note that formulas (3.55) and (3.56) are justified by the 
convention (3.21), because \( r_{p+1} = \cdots = r_n = 0. \) The proof of Theorem 3.52 amounts to a careful 
enumeration of all geodesics connecting \( (0, 0) \) and \( (x, t) \) plus a comparison of their lengths. We 
shall do this in several steps.

**Lemma 3.57.** Suppose \( x' \neq 0, x'' = 0 \) and \( t > 0. \) Then there exist a finite number of 
geodesics connecting \( (0, 0) \) and \( (x, t) \) which are indexed by the solutions of 

\[ t = \sum_{j=1}^{n} a_j \mu(2a_j \theta) r_j^2, \]  

and their lengths increase with \( \theta. \) The square of the length of the geodesic associated to a solution 
of (3.58) is \( 2S(x, t, 1; \theta). \n
**Proof.** If we replace \( a_n \) by the largest \( a_j \) such that \( r_j \neq 0, \) then Lemma 3.57 is an immediate 
consequence of Theorem 3.24. If \( \theta_1 < \theta_2 \leq \cdots \) denote solutions of (3.58), then \( 0 < \theta_1 < \pi/2a_1. \)

Next we shall find geodesics which are not associated to a solution \( \theta \) of (3.58). We assumed 
that \( x' \neq 0. \) Then \( t > 0 \) implies that \( \theta > 0, \) and then (1.15) yields 

\[ \zeta_j(0) = \zeta_j(1), \quad j = 2p + 1, \ldots, 2n. \]

Therefore (1.14) and (1.20) imply:

\[ \begin{bmatrix} \cos(4a_n \theta) & \sin(4a_n \theta) \\ -\sin(4a_n \theta) & \cos(4a_n \theta) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \]

so we must have 

\[ 2a_n \theta = m\pi, \quad m = 1, 2, \ldots. \]
Assuming \( t(0) = 0 \), (1.16) yields

\[
\theta t = \frac{1}{2} \| \xi(0)'' \|^2 + \frac{1}{2} \| \xi(0)' \|^2 - \frac{1}{2} \langle x', \xi(0)' \rangle.
\]

We note that

\[
\frac{1}{2} \left[ \| \xi^{(j)}(0) \|^2 - \langle x^{(j)}, \xi^{(j)}(0) \rangle \right] = a_j \theta \mu(2a_j \theta) r_j^2,
\]

where we set:

\[
\xi^{(j)} = (\xi_{2j-1}, \xi_{2j}), \quad x^{(j)} = (x_{2j-1}, x_{2j}), \quad j = 1, 2, \ldots, n.
\]

Therefore \( \theta \) must be a solution of

\[
\theta t = \frac{1}{2} \| \xi(0)'' \|^2 + \sum_{j=1}^{n} a_j \theta \mu(2a_j \theta) r_j^2.
\]

This can be turned around by saying that we may set \( t(0) = 0 \) for a given \( \theta \) if we choose \( \xi(0)'' \) so that

\[
\frac{1}{2} \| \xi(0)'' \|^2 = \theta \left[ t - \sum_{j=1}^{n} a_j \mu(2a_j \theta) r_j^2 \right].
\]

The length \( d_m(x, t) \) of such a geodesic may be obtained from (1.16):

\[
d_m(x, t)^2 = 2H_0 = 2t \theta + \langle x, \xi(0) \rangle = 2\theta \left[ t + \sum_{j=1}^{n} a_j \cot(2a_j \theta) r_j^2 \right].
\]

We collect these results in the following:

**Lemma 3.68.** Given \( (x, t) \in H_n, x' \neq 0, x'' = 0 \) and \( t \neq 0 \), let \( m \) denote a positive integer, such that

\[
|t| \geq \sum_{j=1}^{n} a_j \mu(2a_j m \pi/2a_n) r_j^2.
\]

Then there are infinitely many geodesics of the same length \( d_m(x, t) \),

\[
d_m(x, t)^2 = 2S(x, |t|, 1; m\pi/2a_n) = 2 \frac{m\pi}{2a_n} \left[ |t| + \sum_{j=1}^{n} a_j \cot(2a_j m \pi/2a_n) r_j^2 \right],
\]

between \((0, 0)\) and \((x, t)\). They are parametrized by the sphere

\[
\frac{1}{2} \| \xi(0)'' \|^2 = \frac{m\pi}{2a_n} \left[ |t| - \sum_{j=1}^{n} a_j \mu(2a_j m \pi/2a_n) r_j^2 \right], \quad \xi(0)'' \in \mathbb{R}^{2n-2\rho}.
\]

We note that (3.69) automatically implies that \( x'' = 0 \) and that \( a_n \neq a_j m \), when \( r_j \neq 0 \).
For application to Heisenberg manifolds we need the length of the shortest geodesic between
\((0,0)\) and \((x,t)\), the so-called Carnot–Caratheodory distance \(d(x,t)\). This will turn out to be
associated to a \(\theta\) in the interval \((0,\pi/2a_n]\). A geodesic associated to such a \(\theta\) always exists.

(i) If we have

\[
(3.72) \quad t < \sum_{j=1}^{n} a_j \mu (2a_j \pi / 2a_n) r_j^2, \quad t > 0,
\]

then the proof of Theorem 3.24 yields the existence of a geodesic with length \(d(x,t)\),

\[
(3.73) \quad d(x,t)^2 = 2\theta \left[ t + \sum_{j=1}^{n} a_j \cot(2a_j \theta) r_j^2 \right].
\]

which is associated to \(\theta\), the unique solution of

\[
(3.74) \quad t = \sum_{j=1}^{n} a_j \mu (2a_j \pi / 2a_n) r_j^2
\]

in the interval \((0,\pi/2a_n]\).

(ii) If

\[
(3.75) \quad t \geq \sum_{j=1}^{n} a_j \mu (2a_j \pi / 2a_n) r_j^2,
\]

then Lemma 3.68 yields a geodesic with length \(d(x,t)\),

\[
(3.76) \quad d(x,t)^2 = 2 \frac{\pi}{2a_n} \left[ t - \sum_{j=1}^{n} a_j \cot(2a_j \pi / 2a_n) r_j^2 + \sum_{j=1}^{n} a_j \frac{2a_j \pi / 2a_n}{\sin^2(2a_j \pi / 2a_n)} r_j^2 \right]
\]

\[
= \|\xi(0)'\|^2 + \sum_{j=1}^{n} \frac{(2a_j \pi / 2a_n)^2}{\sin^2(2a_j \pi / 2a_n)} r_j^2.
\]

In particular, \(d(x,t)\) always has the form (3.73) with the understanding that \(\theta = \pi/2a_n\) when
(3.75) holds.

**Lemma 3.77.** – Let \(x' \neq 0\), \(x'' = 0\), \(t > 0\) and suppose that (3.69) holds for some positive
integer \(m\); then

\[
(3.78) \quad d(x,t) \leq d_m(x,t),
\]

\[
(3.79) \quad \lim_{x'' \to 0} d(x',x'',t;\theta'') = d_k(x',0,t)^2,
\]

where \(k = 1\) or \(m\), and \(\theta'' \in ((k-1)\pi/2a_n, k\pi/2a_n)\) is that solution of (3.74) which increases as
\(x'' \to 0\).
(i) If \(k = 1\) and (3.72) holds, then (3.79) follows from the behaviour of \(\mu\) in \([0, \pi/2a_n]\).

(ii) Otherwise we assume that (3.69) holds, and \(d_k(x, t), k = 1\) or \(m\), is given by (3.70). With nonzero \(x^0\) we assume that \(\theta''\) is that solution of:

\[
\lim_{x^0 \to 0} a_n \mu(2a_n \theta'') \|\theta''\|^2 = \lim_{x^0 \to 0} \frac{2a_n \theta''}{\sin^2(2a_n \theta'')} \|\theta''\|^2 = t - \sum_{j=1}^{p} a_j \mu(2a_j k\pi/2a_n) r_j^2.
\]

in the interval \(((k - 1)\pi/2a_n, k\pi/2a_n)\), which increases to \(k\pi/2a_n\) as \(x'' \to 0\); then

\[
\sum_{j=1}^{p} a_j \mu(2a_j k\pi/2a_n) r_j^2 = t - \sum_{j=1}^{p} a_j \mu(2a_j k\pi/2a_n) r_j^2.
\]

The square of the distance associated to \(x''\) is given by

\[
2S(x', x'', t, 1; \theta'') = 2 \sum_{j=1}^{p} \frac{(2a_j \theta'')^2}{2 \sin^2(2a_j \theta'')} r_j^2 + \frac{(2a_n \theta'')^2}{2 \sin^2(2a_n \theta'')} \|\theta''\|^2,
\]

and (3.81) yields

\[
\lim_{x'' \to 0} 2S(x', x'', t, 1; \theta'') = 2 \sum_{j=1}^{p} \frac{(2a_j k\pi/2a_n)^2}{2 \sin^2(2a_j k\pi/2a_n)} r_j^2 + \frac{k\pi}{2a_n} \left[ t - \sum_{j=1}^{p} a_j \mu(2a_j k\pi/2a_n) r_j^2 \right]
\]

\[
= 2 \frac{k\pi}{2a_n} \left[ t + \sum_{j=1}^{p} a_j \cot(2a_j k\pi/2a_n) r_j^2 \right]
\]

\[
= d_k(x', 0, t)^2 = d_k(x, t)^2.
\]

Therefore the distances \(d_1(x, t)\) and \(d_m(x, t), x'' = 0\), are limits of distances \(d(y, t), y'' \neq 0\), as \(y'' \to 0\). When \(y'' \neq 0\), the distances are increasing with increasing \(\theta\), see Theorem 3.24. Hence this is also true for the limits which implies (3.78).

To obtain all geodesics between \((0, 0)\) and a given point \((x, t), x \neq 0, t > 0\), we need more terminology and notation. Let \(-4a_1^2, -4a_2^2, \ldots, -4a_m^2\) denote the distinct eigenvalues of \(A^2\), \(a_1 < a_2 < \cdots < a_m\). Let \(V_k\) denote that eigenspace of \(A^2\) which is associated to the eigenvalue \(-4a_k^2\). We set:

\[
W_k = \bigcup_{j=1}^{k} V_k, \quad R_k^2 = \sum_{a_j = a_k} r_j^2.
\]

Then the arguments of Lemmas 3.68 and 3.77 also yield.
Lemma 3.85. – Let $x \in W_k$, $x \notin W_{k-1}$. Suppose there is an integer $q$, $k < q \leq n$, and a positive integer $m$, such that:

$$t \geq \sum_{j=1}^{n} a_j \mu(2a_jm\pi/2a_q)r_j^2.$$  \hspace{1cm} (3.86)

Then there exist infinitely many geodesics of the same length, connecting $(0, 0)$ and $(x, t)$; there is only one geodesic if (3.86) is an equality. This set of geodesics is parametrized by the sphere:

$$\sum_{a_j \neq a_k, a_jm = ca_q} \left( |\xi_{a_j}(0)|^2 + |\xi_{a_j}(0)|^2 \right) = \frac{m\pi}{2a_q} \left[ |t| - \sum_{j=1}^{n} a_j \mu(2a_jm\pi/2a_q)r_j^2 \right].$$  \hspace{1cm} (3.87)

where $\ell$ is an arbitrary positive integer, and the square of their length is given by

$$2S(x, t, 1; m\pi/2a_q) = \frac{m\pi}{2a_q} \left[ |t| + \sum_{j=1}^{n} a_j \cot(2a_jm\pi/2a_q)r_j^2 \right].$$  \hspace{1cm} (3.88)

Moreover, if $m = m_1$ and $m = m_2$ both satisfy (3.86), then $m_1 < m_2$ implies

$$2S(x, t, 1; m_1\pi/2a_q) \leq 2S(x, t, 1; m_2\pi/2a_q).$$  \hspace{1cm} (3.89)

Proof. – We only need to prove (3.89). This follows from taking the limit of $S(y, t, 1; \theta)$ as $y \to x$, where $y$ has a nonzero projection into $V_q$ and $\theta$ is a solution of

$$t = \sum_{j=1}^{n} a_j \mu(2a_j\theta)r_j^2, \quad \theta \in \left((m-1)\pi/2a_q, m\pi/2a_q\right).$$  \hspace{1cm} (3.90)

which increases to $m\pi/2a_q$ as $y \to x$. \qed

Now that we have accounted for all the geodesics, we need the shortest one. Suppose $x \in W_k$, $x \notin W_{k-1}$. Among the geodesics of Lemma 3.57, the shortest is associated to $\theta \in (0, \pi/2a_k)$. Among the geodesics constructed in Lemmas 3.68 and 3.85 the shortest occurs when $m = 1$, and it is associated to a $\theta$, $\theta = \pi/2a_q$, $q > k$. Therefore the shortest geodesic is associated to a $\theta \in (0, \pi/2a_k)$.

Lemma 3.91. – Suppose $x \neq 0$, $x \in W_k$, $x \notin W_{k-1}$. Then the shortest geodesic connecting $(0, 0)$ and $(x, t)$ is associated to a $\theta \in (0, \pi/2a_q)$, and its length, the Carnot–Caratheodory distance of $(x, t)$ from the origin, is given by $d(x, t)$, where:

$$d(x, t)^2 = 2\theta \left[ |t| + \sum_{j=1}^{n} a_j \cot(2a_j\theta)r_j^2 \right].$$  \hspace{1cm} (3.92)

Proof. – (i) Suppose $|t| > 0$ is large, so that (3.69) holds with $m = 1$, and also (3.86) holds with $m = 1$ and with $q > k$. We need to show that

$$2S(x, t, 1; \pi/2a_n) < 2S(x, t, 1; \pi/2a_q),$$

that is:
\[
2 \frac{\pi}{2a_n} |t| + \sum_{j=1}^{n} \left[ \frac{(2a_j \pi / 2a_n)^2}{\sin^2(2a_j \pi / 2a_n)} - \frac{2a_j \pi}{2a_n} \mu(2a_j \pi / 2a_n) \right] r_j^2 \leq \frac{2 \pi}{2a_q} |t| + \sum_{j=1}^{n} \left[ \frac{(2a_j \pi / 2a_q)^2}{\sin^2(2a_j \pi / 2a_q)} - \frac{2a_j \pi}{2a_q} \mu(2a_j \pi / 2a_q) \right] r_j^2.
\]
(3.93)

see (3.76). We move \((2\pi/2a_n)|t|\) to the right-hand side and replace \(|t|\) by the smaller quantity:
\[
\sum_{j=1}^{n} a_j \mu(2a_j \pi / 2a_q) r_j^2.
\]

Then (3.93) will follow from
\[
\frac{(2a_j \pi / 2a_q)^2}{\sin^2(2a_j \pi / 2a_q)} \leq \frac{2a_j \pi}{2a_q} \mu(2a_j \pi / 2a_q).
\]
(3.94)

We set \(\alpha = 2a_j \pi / 2a_n\) and \(x = 2a_j \pi / 2a_q\). Then (3.94) takes the form:
\[
\frac{\alpha^2 - \alpha(\alpha - \sin \alpha \cos \alpha)}{\sin^2 \alpha} < \frac{x^2 - \alpha(x - \sin x \cos x)}{\sin^2 \alpha}, \quad \alpha < x.
\]
(3.95)

We set
\[
\psi(x) = \frac{2x^2 - \alpha(2x - \sin 2x)}{2 \sin^2 x},
\]
(3.96)

and note that
\[
\psi'(x) = (x - \alpha)\mu'(x) > 0 \quad \text{if} \quad \alpha < x < \pi.
\]
(3.97)

This proves (3.95).

(ii) The only other possibility we need to consider is that (3.69) holds with \(m = 1\), and there is a \(\theta \in (\pi/2a_n, \pi/2a_q)\) such that
\[
|t| = \sum_{j=1}^{n} a_j \mu(2a_j \theta) r_j^2.
\]
(3.98)

In this case we would like to show that
\[
2 \frac{\pi}{2a_n} |t| + \sum_{j=1}^{n} \left[ \frac{(2a_j \pi / 2a_n)^2}{\sin^2(2a_j \pi / 2a_n)} - \frac{2a_j \pi}{2a_n} \mu(2a_j \pi / 2a_n) \right] r_j^2 \leq \frac{2 \pi}{2a_q} |t| + \sum_{j=1}^{n} \left[ \frac{(2a_j \pi / 2a_q)^2}{\sin^2(2a_j \pi / 2a_q)} - \frac{2a_j \pi}{2a_q} \mu(2a_j \pi / 2a_q) \right] r_j^2,
\]
(3.99)

which may be written as
\[
\sum_{j=1}^{n} \left[ \frac{(2a_j \pi / 2a_q)^2}{\sin^2(2a_j \pi / 2a_q)} - \frac{2a_j \pi}{2a_q} \mu(2a_j \pi / 2a_q) \right] r_j^2 \leq \sum_{j=1}^{n} \left[ \frac{(2a_j \pi / 2a_q)^2}{\sin^2(2a_j \pi / 2a_q)} - \frac{2a_j \pi}{2a_q} \mu(2a_j \pi / 2a_q) \right] r_j^2.
\]
(3.100)

Again this follows from (3.95), and we have proved Lemma 3.91. \( \square \)
COROLLARY 3.101. – Let \( x = 0, t \neq 0 \), and fix a pair of integers \((j, m)\), \( j = 1, 2, \ldots, n \), \( m = 1, 2, 3, \ldots \). There are infinitely many geodesics of the same length \( d_{j m}(0, t) \) between \((0, 0)\) and \((0, t)\), where

\[
d_{j m}(0, t)^2 = \frac{2m \pi |t|}{2a_j}.
\]

The Carnot–Caratheodory distance of \((0, t)\) from the origin is

\[
d(0, t)^2 = d_{n1}(0, t)^2 = \frac{\pi |t|}{a_n}.
\]

We note that different pairs \((j, m)\) may yield the same geodesic.

Proof of Theorem 3.52 concluded. – Lemma 3.57 proves the statement (3.56), and Lemmas 3.68, 3.77, 3.85 and 3.91 imply (3.55) and (3.56). To prove that \( d(x, t) \) is continuous on \( H_n \) we only need to show that

\[
\lim_{x'' \to 0} d(x', x'', t) = d(x', 0, t).
\]

This limit has already been established in part (ii) of the proof of Lemma 3.77, if there we set \( k = 1 \).

Homogeneity with respect to the dilations (1.57) is again a consequence of the fact that \( \theta_c(x, t) \) is homogeneous of degree 0. The quotient \( d(x, t)/(|x| + |t|^{1/2}) \) is homogeneous of degree 0 and positive on \( H_n \setminus \{0\} \), so it is bounded above and below. Thus to finish the proof of Theorem 3.52, we are left with showing that \( d(x, t) \) is not differentiable. Let

\[
E_p(x, t) = |t| - \sum_{j=1}^n a_j \mu(2a_j \pi / 2a_n) x_j^2
\]

denote the \textit{excess} function defined on

\[
H_p = \{(x, t) \in H_n; \ x = (x', 0)\}.
\]

Lemmas 3.68 and 3.85 show that for points in \( H_p \), the initial momentum \( \xi(0) \) is not uniquely defined, and we obtain a multitude of bicharacteristics; we may refer to such points as \textit{caustic} points.

Lemma 3.107. – All those first derivatives of \( S(x, t; 1; \theta_c) \) which are normal to \( H_p \) are discontinuous on \( \{(x, t) \in H_n; \ E_p(x, t) > 0\} \).

Proof. – It suffices to prove Lemma 3.107 when \( t > 0 \), which we shall assume in the rest of the argument. Let \((x, t) \in H_p \) and let \( h \) denote one of the variables \( x_{2p+1}, \ldots, x_{2n} \). Then \( S \) is given by:

\[
S(h) = S(x_1, \ldots, x_{2p}, 0, \ldots, 0, h, 0, \ldots, 0, t, 1, \theta_h)
\]

\[
= \frac{1}{2} \sum_{j=1}^p \left( \frac{(2a_j \theta_h)^2}{\sin^2(2a_j \theta_h)} \right) h^2 + \left( \frac{(2a_j \theta_h)^2}{2 \sin^2(2a_j \theta_h)} \right) h^2,
\]
where
\begin{equation}
(3.109) 
    t = \sum_{j=1}^{p} a_j \mu (2a_j \theta_h) r_j^2 + a_n \mu (2a_n \theta_h) h^2.
\end{equation}

We shall calculate \( \lim_{h \to 0} S'(h) \). Using (1.45) we obtain:
\begin{equation}
(3.110) 
    S'(h) = 2\theta_h \left[ \sum_{j=1}^{p} a_j^2 \mu' (2a_j \theta_h) \frac{d\theta_h}{dh} r_j^2 + a_n \mu (2a_n \theta_h) h \right. \\
    \left. + a_n \cot(2a_n \theta_h) h + a_n^2 \mu (2a_n \theta_h) \frac{d\theta_h}{dh} h^2 \right].
\end{equation}

Next we differentiate (3.109) with respect to \( h \),
\begin{equation}
(3.111) 
    0 = \sum_{j=1}^{p} a_j^2 \mu' (2a_j \theta_h) \frac{d\theta_h}{dh} r_j^2 + a_n \mu (2a_n \theta_h) h + a_n^2 \mu' (2a_n \theta_h) \frac{d\theta_h}{dh} h^2.
\end{equation}

and substituting (3.111) into (3.110) we find
\begin{equation}
(3.112) 
    S'(h) = 2a_n \theta_h \cot(2a_n \theta_h) h.
\end{equation}

As \( h \to 0, \theta_h \to \pi / 2a_n \), so (3.108) yields:
\begin{equation}
(3.113) 
    \frac{2a_n \theta_h}{\sin(2a_n \theta_h)} \cdot \frac{h}{\sin(2a_n \theta_h)} = 2 S(h) - \sum_{j=1}^{p} \frac{(2a_j \theta_h)^2}{\sin^2(2a_j \theta_h)} r_j^2 \\
    \lim_{h \to 0} \frac{\pi}{2a_n} \left[ t + \sum_{j=1}^{p} a_j \cot(2a_j \pi / 2a_n) r_j^2 \right] - \sum_{j=1}^{p} \frac{(2a_j \pi / 2a_n)^2}{\sin^2(2a_j \pi / 2a_n)} r_j^2 \\
    = \frac{\pi}{a_n} \left[ t - \sum_{j=1}^{p} a_j \mu (2a_j \pi / 2a_n) r_j^2 \right] \\
    = \frac{\pi}{a_n} E_p.
\end{equation}

Since \( \theta_h \) is symmetric in \( h \), we have
\begin{equation}
(3.114) 
    \lim_{h \to 0^+} \frac{dS(h)}{dh} = \pm \left( \frac{\pi E_p}{a_n} \right)^{1/2} \neq 0,
\end{equation}

when \( E_p > 0 \), hence \( S'(h) \) is discontinuous at \( h = 0 \). This proves Lemma 3.107.

With Lemma 3.107 we have completed the proof of Theorem 3.52.

We make a few remarks about the set of caustics:
\begin{equation}
(3.115) 
    C_p = \{(x,t) \in H_p: E_p(x,t) > 0 \}.
\end{equation}
Let $S_p(x, t)$ denote the classical action on $H_p$:

\[
S_p(x, t) = \theta_p|t| + \sum_{j=1}^{p} a_j \theta_p \cot(2a_j \theta_p) r_j^2, \quad (x, t) \in H_p,
\]

whenever $\theta_p \in (0, \pi/2a_p)$ is a solution of

\[
|t| = \sum_{j=1}^{p} a_j \mu (2a_j \pi/2a_p) r_j^2;
\]

otherwise we extend (3.116) to all of $H_p$ by continuity. Then (ii) in the proof of Lemma 3.91 implies that

\[
S_p(x, t) > S(x, t), \quad \text{if } (x, t) \in H_p \text{ and } E_p(x, t) > 0.
\]

On the other hand we always have

\[
S_p(x, t) = S(x, t), \quad \text{if } (x, t) \in H_p \text{ and } E_p(x, t) < 0.
\]

Consequently, one has

\[
S_p(x, t) \geq S(x, t), \quad (x, t) \in H_p.
\]

Continuing in this manner, suppose that

\[
0 < a_1 \leq \cdots \leq a_{p_2} < a_{p_2+1} = \cdots = a_{p_1} < a_{p_1+1} = \cdots = a_p < \cdots.
\]

Again we set

\[
H_{p_1} = \left\{ (x, t) \in H_p; \sum_{j=p_1+1}^{n} r_j^2 = 0 \right\},
\]

\[
E_{p_1}(x, t) = |t| - \sum_{j=1}^{p_1} a_j \mu (2a_j \pi/2a_p) r_j^2, \quad (x, t) \in H_{p_1},
\]

and

\[
C_{p_1} = \left\{ (x, t) \in H_{p_1}; E_{p_1}(x, t) > 0 \right\}.
\]

Then

\[
C_p \supset C_{p_1},
\]

since $E_p(x, t) > E_{p_1}(x, t)$ on $H_{p_1}$. We say that the points of $C_{p_1}$ are more caustic than the points of $C_p \setminus C_{p_1}$, in the sense that they have more indeterminacy in their initial momenta, which leads to a larger set of bicharacteristics. Let $S_{p_1}(x, t)$ denote the classical (continuous) action on $H_{p_1}$:

\[
S_{p_1}(x, t) = \theta_{p_1}|t| + \sum_{j=1}^{p_1} a_j \theta_{p_1} \cot(2a_j \theta_{p_1}) r_j^2, \quad (x, t) \in H_{p_1},
\]
whenever $\theta_{p_1} \in (0, \pi/2a_{p_1})$ is a solution of

$$|t| = \sum_{j=1}^{p_1} a_j \mu(2a_j \theta_{p_1}) r_j^2;$$

otherwise $S_{p_1}$ is extended by continuity to all of $H_{p_1}$. Again the proof of Lemma 3.91 applies and we have:

$$S(x, t) \subseteq S_p(x, t) \subseteq S_{p_1}(x, t) \subseteq \cdots,$$

where $(x, t)$ belongs to the appropriate domain of definition of the corresponding actions; (3.128) may be continued to $S_0$, the action on $[(0, t) ; t \in \mathbb{R}]$. For obvious reasons we refer to $S(x, t)$ as the **minimal action**. Extending (3.125) we obtain a decreasing sequence of sets:

$$C_p \supset C_{p_1} \supset C_{p_2} \supset \cdots,$$

whose intersection is the $t$-axis.

### The complex action

As in Section 1, we introduce complex Hamiltonian mechanics into the picture by setting

$$\theta = -i.$$

Again the complex action integral is taken to be:

$$g(x, t, \tau) = -it + \int_0^\tau \left\{ \langle \dot{x}, \dot{\xi} \rangle - H \right\} dt$$

$$= -it + \frac{1}{2} \left\{ iA \coth(iA) x, x \right\}$$

$$= -it + \sum_{j=1}^n a_j \coth(2a_j \tau) r_j^2.$$

It satisfies the Hamilton–Jacobi equation

$$0 = \frac{\partial g}{\partial \tau} + H \left( x, \frac{\partial g}{\partial x} \right) = \frac{\partial g}{\partial \tau} + \frac{1}{2} \sum_{j=1}^n (X_j g)^2,$$

and its derivatives are given by:

$$\frac{\partial g}{\partial x} = \dot{\xi}(\tau), \quad \frac{\partial g}{\partial t} = \theta,$$

so

$$0 = \frac{\partial g}{\partial \tau} + \frac{1}{\tau} \left\{ g(x, t, \tau) + it(0) \right\}.$$
\textbf{Theorem 3.135.} – Suppose \((x, t) \in H_n, x'' \neq 0.\) Then the unique critical point with respect to \(\tau\) of the modified complex action function

\begin{equation}
\tag{3.136}
f(x, t, \tau) = \tau g(x, t, \tau) = -i\tau t + \sum_{j=1}^{n} a_j \coth(2a_j \tau) r_j^2
\end{equation}

in the strip \(|\text{Im}\, \tau| < \pi/2a_n\) is the point \(\tau(x, t) = i\theta_c(x, t);\) as before, \(\theta_c\) is the solution of (3.20) in this interval. At the critical point

\begin{equation}
\tag{3.137}
f(x, t, \tau_c(x, t)) = S(x, t, 1; \theta_c) = \frac{1}{2} d(x, t)^2.
\end{equation}

The identity (3.137) remains valid at caustics.

The proof of Theorem 3.135 is a slight modification of the proof of Theorem 1.66.

\section{4. Geometry and the heat kernel on isotropic \(H_n\)}

When

\begin{equation}
\tag{4.1}
a_1 = a_2 = \cdots = a_n = a,
\end{equation}

we say that \(H_n\) is isotropic. The geometry and the analysis of the heat kernel in the isotropic case is much simpler and more precise than in the anisotropic case, i.e. when (4.1) does not hold. In this section we shall discuss the isotropic case separately.

Once again we obtain classical Hamiltonian mechanics by completing the boundary conditions (1.11) with the condition \(t(0) = 0,\) and setting \(\tau = 1.\) Then necessarily we have:

\begin{equation}
\tag{4.2}
t = a \mu (2a) \|x\|^2, \quad \|x\|^2 = \sum_{j=1}^{2n} x_j^2.
\end{equation}

This agrees with (1.32), and the results of Section 1 carry over to the general isotropic case with no change in the proofs. In particular, Theorem 1.36 and Theorem 1.56 are valid as stated. Theorem 1.41 is also valid when modified: each family of geodesics of length \(d_m\) from \((0, 0)\) to \((0, t)\) is parametrized by the \((2n - 1)\)-sphere.

The heat kernel in the general isotropic case is

\begin{equation}
\tag{4.3}
P(x, t; u) = \frac{1}{(2\pi u)^{n+1}} \int_{-\infty}^{\infty} e^{-f(x, t, \tau)/u} V(\tau) \, d\tau,
\end{equation}

where again

\begin{equation}
\tag{4.4}
f(x, t, \tau) = \tau g(x, t, \tau) = -i\tau t + a \tau \coth(2a \tau) \|x\|^2,
\end{equation}

and

\begin{equation}
\tag{4.5}
V(\tau) = \left( \frac{2a \tau}{\sinh(2a \tau)} \right)^n.
\end{equation}

The derivation sketched in Section 2 carries over to (4.3). However the statements and proofs of the estimates and asymptotics must be modified, as we now show.
THEOREM 4.6. – The heat kernel \( P(x, t; u) \) for the isotropic Heisenberg group \( H_n \) satisfies the estimate
\[
P(x, t; u) \leq C e^{-d(x, t)^2/2u} \left( \min \left\{ 1 + \frac{d(x, t)^2}{u}, \frac{d(x, t)}{|x|} \right\} \right)^{n-1} \times \min \left\{ 1, \left( \frac{u}{|x|d(x, t)} \right)^{1/2} \right\}, \quad (x, t, u) \in H_n \times \mathbb{R}_+.
\]
(4.7)

As in the proof of Theorem 2.17 we may assume that \(|x| \neq 0\) and \(t > 0\). We begin with the analogue of Lemma 2.20.

LEMMA 4.8. – For any \( \varepsilon_0 > 0 \), the estimate (4.7) is valid for \((x, t)\) such that \(2a\theta_c(x, t) \leq \pi - \varepsilon_0\), with a constant that depends only on \(\varepsilon_0\).

Proof. – As in (2.22), we move the contour of integration:
\[
P(x, t; u) = \frac{1}{(2\pi u)^{n+1}} \int_{\text{Im} \tau = \theta_c} e^{-f(x, t, \tau)/u} V(\tau) \, d\tau.
\]
(4.9)

In the region under consideration, \(d(x, t)/|x|\) is bounded away from 0 and \(\infty\), so the estimate (4.7) is equivalent to:
\[
P(x, t; u) \leq C e^{-d(x, t)^2/4u} \min \left\{ 1, \frac{\sqrt{|x|}}{\sqrt{|x|d(x, t)}} \right\}. \quad (4.10)
\]

The proof of (4.10) is the same as the proof of Lemma 2.20. □

Proof of Theorem 4.6. – As in the proof of Theorem 2.17, to control the estimates as \(2a\theta_c \to \pi\), we replace the contour in (4.9) by a circle around \(\pi i/2a\) of radius \(\pi/2a - \theta_c\), together with the line \(\text{Im}(2a\tau) = 3\pi/2\):
\[
P(x, t; u) = \frac{1}{(2\pi u)^{n+1}} \int_{|\tau - 2a\tau| = \pi - 2a\theta_c} e^{-f(x, t, \tau)/u} V(\tau) \, d\tau
\]
(4.11)
\[
+ \frac{1}{(2\pi u)^{n+1}} \int_{\text{Im}(2a\tau) = 3\pi/2} e^{-f(x, t, \tau)/u} V(\tau) \, d\tau
\]
\[
= P_0(x, t; u) + P_1(x, t; u).
\]

We estimate \(P_1\), defined by the second integral in (4.11), exactly as we estimated \(P_1\) in the proof of Theorem 2.17. The result is:
\[
|P_1(x, t; u)| \leq \frac{e^{-3\pi/4a}}{(2\pi u)^{n+1}} \int_{\mathbb{R}} V\left( s + i \frac{3\pi}{4a} \right) \, ds \leq C e^{-d(x, t)^2/2u} \frac{e^{-d(x, t)^2/4u}}{(2\pi u)^{n+1}}.
\]
(4.12)

The argument following (2.31) implies that the right side of (4.12) is dominated by the right side of (4.7).

We turn now to estimating \(P_0\), defined by integration over a circle around \(\pi i/2a\). Again we take
\[
2a\tau = i\pi - i\zeta, \quad F = \frac{\pi |x|^2}{2}.
\]
(4.13)
and write
\begin{equation}
\begin{aligned}
f &= -\frac{F}{\zeta} + G(\zeta) + \frac{t}{2a}(\pi - \zeta), \\
V(t) &= \frac{2a W(\zeta)}{\zeta^n},
\end{aligned}
\end{equation}
with \(G\) and \(W\) holomorphic on the circle \(\zeta = e^{i\varphi}, \varphi = \pi - 2a\theta_\varepsilon\); thus
\begin{equation}
P_0 = \frac{1}{(2\pi u)^{n+1}} \int_{|\zeta|=\varepsilon} e^{-f/\mu} W(\zeta) \frac{d\zeta}{\zeta^n}.
\end{equation}
As in (2.40), the integrand is dominated by \(\exp(-d^2/2\mu)/\varepsilon^n\), so we obtain immediately the following estimate:
\begin{equation}
|P_0| \leq \frac{C}{u^{n+1}} e^{-d^2/2\mu} \frac{1}{\varepsilon^{n-1}} \sim \frac{C}{u^{n+1}} e^{-d^2/2\mu} \left( \frac{d(x,t)}{\|x\|^2} \right)^{n-1},
\end{equation}
since we are in the range where \(\varepsilon^2 \sim \|x\|^2/t \sim \|x\|^2/d(x,t)^2\). As in the proof of Theorem 2.17, if \(u/\|x\|d(x,t) < 1\) then this estimate and the one to follow can be improved by a factor \((u/\|x\|d(x,t))^{1/2}\). Therefore we only need to prove the estimate
\begin{equation}
|P_0| \leq \frac{C}{u^{n+1}} \left( 1 + \frac{d^2}{u} \right)^{n-1} e^{-d^2/2\mu}.
\end{equation}
Note that, for small \(x\), (2.19) yields
\begin{equation}
F \sim \frac{\varepsilon^2 t}{2a} \sim \frac{\varepsilon^2 d(x,t)^2}{2\pi},
\end{equation}
see (1.38). We recall (2.38) and (2.39):
\begin{equation}
f - f_c \sim \left\{ \frac{\varepsilon}{a\pi} + O(\varepsilon^4) \right\} t(1 - \cos \varphi) \geq 0
\end{equation}
if \(|\zeta| = \varepsilon \leq \varepsilon_0\), for some \(\varepsilon_0 > 0\). Next we integrate by parts:
\begin{equation}
P_0 = \frac{1}{(2\pi u)^{n+1}} \int_{|z|^2-2\pi} \frac{\partial}{\partial \zeta} \left\{ e^{-f/\mu} W(\zeta) \right\} \frac{d\zeta}{\zeta^{n-1}}.
\end{equation}
Derivatives of \(W\) are \(O(1)\). The first derivative of \(f/\mu\) is
\begin{equation}
\frac{F}{\varepsilon^2 u} + \frac{G'(\zeta)}{u} - \frac{t}{2au} = O\left( \frac{d^2}{u} \right)
\end{equation}
on the circle.
If \(n = 2\), then it follows from (4.19) and (4.20) that \(P_0\) is the product of a factor that is \(O(1+d^2/u)\) and an integral that has the same form as the integral for \(P_0\) in the \(H_1\) case. Therefore
the estimate (4.17) is immediate. In the general case we proceed recursively. We write $I_m$ for any integral of the form

\begin{equation}
I_m = \int_{|\zeta|=\epsilon} e^{-f/\epsilon} W_m(\zeta) \frac{d\zeta}{\zeta^m},
\end{equation}

where $W_m$ is holomorphic in a disc that contains the circle of integration. We claim that $P_0$ has the form

\begin{equation}
P_0 = \frac{1}{u^{n+1}} O \left( \left( 1 + \frac{d^2}{u} \right)^2 \epsilon^2 I_{n+1} + I_{n-1} \right).
\end{equation}

In fact all terms that come from differentiation on the right side of (4.21), with $m = n$, can be grouped into the $I_{n-1}$ portion of (4.22), with the exception of

\begin{equation}
\frac{F}{\epsilon^2 u} = O \left( \frac{d^2}{u} \right) \frac{\epsilon^2}{\epsilon^2},
\end{equation}

which accounts for the $\epsilon^2 I_{n+1}$ term in (4.22).

If $n - 1 > 1$ we can integrate by parts once in each of the integrals in (4.22) and regroup, to obtain

\begin{equation}
P_0 = \frac{1}{u^{n+1}} O \left( \left( 1 + \frac{d^2}{u} \right)^2 \epsilon^4 I_{n+2} + \epsilon^2 I_n + I_{n-2} \right).
\end{equation}

Continuing in this fashion, we obtain

\begin{equation}
P_0 = \frac{1}{u^{n+1}} O \left( \left( 1 + \frac{d^2}{u} \right)^{n-1} \sum_{j=0}^{n-1} \epsilon^{2j} I_{1+2j} \right).
\end{equation}

The estimate (4.17) follows immediately from (4.23), the form (4.21) of $I_m$, and our previous estimate $\text{Re } f \geq f_\epsilon$ on the circle. This completes the proof of Theorem 4.6.

We turn now to the small time behavior in the general isotropic case.

**Theorem 4.24.** – Suppose that $x \neq 0$, and we let $\theta_\epsilon$ denote the solution of (4.2) in the interval $[0, \pi/2a]$, where we replace $t$ by $|t|$. Then the heat kernel for isotropic $H_n$ has the following small time behaviour:

\begin{equation}
P(x, t; u) = \frac{1}{(2\pi u)^{n+1}} e^{-d(x, t)^2/2u} \left\{ \Theta(x, t) \sqrt{2\pi u} + O(u) \right\}, \quad u \to 0+,
\end{equation}

where

\begin{equation}
\Theta(x, t) = \left( \frac{1}{f''(\theta_\epsilon)} \right)^{1/2} V(\theta_\epsilon) = \frac{\theta_\epsilon}{\left( 1 - 2a\theta_\epsilon \cot(2a\theta_\epsilon) \right)^{1/2}} \left\{ \frac{2a\theta_\epsilon}{\sin(2a\theta_\epsilon)} \right\}^{n-1}.
\end{equation}

Moreover we have (2.45).

**Theorem 4.27.** – At points $(0, t), t \neq 0$,

\begin{equation}
P(x, t; u) = \frac{e^{-d(0, t)^2/2u}}{(n-1)! (2\sqrt{au})^{2n}} \left\{ 1 + O(u) \right\}, \quad u \to 0+.
\end{equation}
Proof of Theorem 4.24. – Is exactly the same as that of Theorem 2.42. □

Proof of Theorem 4.27. – We may assume that \( t \) is positive. Up to a term that is \( O(e^{-d/u}) \) as \( u \to 0^+ \), \( P \) is given by the first residue:

\[
2\pi i \text{Res} \left\{ \frac{e^{i \tau/u} V(\tau)}{(2\pi u)^{n+1}} ; 2a\tau = i\pi \right\},
\]

see (2.58). We can write

\[
V(\tau) = \frac{W(\tau)}{(2a\tau - i\pi)^n}, \quad W \left( \frac{i\pi}{2a} \right) = (-i\pi)^n,
\]

\( W(\tau) \) analytic around \( i\pi/2a \), so the residue is:

\[
\frac{1}{(n-1)!(2a)^n} \left. \left\{ \frac{\partial}{\partial \tau} \right\}^{n-1} \left\{ \frac{e^{i \tau/u} W(\tau)}{(2\pi u)^{n+1}} \right\} \right|_{2a\tau = i\pi}.
\]

This residue is the product of \( \exp(-i\tau/2au) \) and a polynomial of degree \( 2n \) in \( 1/u \). We are interested only in the principal term as \( u \to 0^+ \), which is obtained when each derivative falls on the exponential. Thus the principal term in (4.29) is:

\[
2\pi i e^{-i\pi/2au} \frac{1}{(n-1)!(2a)^n (2\pi u)^{n+1}} \left( \frac{\partial}{\partial \tau} \right)^{n-1} \left( \frac{-i\pi}{n} \right) = \frac{e^{-i\pi/2au}a^n}{(n-1)!(4au^2)^{n}}.
\]

Again, \( d(0,t)^2 = \pi t/a \), so we deduce (4.28) from (4.31). □

5. The heat kernel on general \( H_n \)

With \( a_1, \ldots, a_n \) arbitrary,

\[
0 < a_1 \leq \cdots \leq a_n,
\]

the heat kernel is given by

\[
P(x, t; u) = \frac{1}{(2\pi u)^{n+1}} \int_{-\infty}^{\infty} e^{-f(x, t, \tau)/u} V(\tau) \, d\tau,
\]

where

\[
f(x, t, \tau) = \tau g(x, t, \tau) = -i\pi t + \sum_{j=1}^{n} a_j \tau \coth(2a_j \tau) \tau_j^2
\]

and

\[
V(\tau) = \prod_{j=1}^{n} \frac{2a_j \tau}{\sinh(2a_j \tau)}.
\]

The argument which gives (2.4) also yields (5.2).
A uniform estimate for the heat kernel

If we regard the defining constants (5.1) as being fixed once and for all, we can obtain estimates like (4.7), with the power $n - 1$ replaced by $m - 1$, where $m$ is the number of indices $a_j$ that are equal to the maximal index $a_n$; see Corollary 5.60. For later applications we need more precise estimates that are uniform with respect to the constants $a_j$.

In the previous estimates, when the first critical point was close to the first singularity, it was convenient to replace the original contour of integration by a circle around the first singularity and a horizontal line above the singularity. In the general case, the configuration of singularities may make it impossible to carry out the same argument in an effective way. Therefore we estimate the integral over the line through the first critical point, by direct examination.

The following estimate can be improved by taking advantage of stationary phase arguments, as we did in (4.7), but the sharpest estimates of this kind are complicated to state and are very far from uniform.

**THEOREM 5.5.** – The following estimate holds for the heat kernel (5.2):

$$P(x, t; u) \leq e^{-d(x, t)^2/2u} \prod_{j=1}^{n-1} \min \left\{ \frac{a_n}{a_n - a_j}, 1 + \frac{d(x, t)^2}{u} \right\}.$$  

The constant $C$ depends only on $n$ and on an upper bound for $a_n$.

As before, we may assume that $x'' \neq 0$ and $t > 0$, and that the first critical point $i\theta_c$ lies in the interval $(0, \pi/2a_n)$. So long as the critical point is bounded away from $i\pi/2a_n$, it is easy to obtain an estimate that is stronger than (5.6).

**LEMMA 5.7.** – For any $\varepsilon > 0$, if $2a_n\theta_c(x, |t|) \leq \pi - \varepsilon_0$, then

$$P(x, t; u) \leq e^{-d(x, t)^2/2u} \prod_{j=1}^{n-1} \min \left\{ \frac{a_n}{a_n - a_j}, 1 + \frac{d(x, t)^2}{u} \right\}.$$  

Here and below the constants depend only on $n$, and $\varepsilon_0$.

**Proof.** We move the path of integration in (5.2) to the line $2a_n \tau = i2a_n \theta_c + s$. On this line $e^{-f''/u}$ is dominated by $e^{-f''/u}$, see (1.72), while $V(\tau)$ is dominated by $|2a_n \tau / \sinh(2a_n \tau)|$, with a constant that depends only on $n$ and $\varepsilon_0$. Integration with respect to $s = 2a_n \Re \tau$ gives (5.8). \qed

The proof of Theorem 5.5 takes up the rest of this section. We continue to assume that $x'' \neq 0$ and $t$ is positive and that the critical point nearest the origin occurs at $i\theta_c$, $0 < 2a_n\theta_c < \pi$. We begin with an analysis of $f$ on the line $\Im \tau = \theta_c$.

**LEMMA 5.9.** – There are positive constants $s_0$ and $\varepsilon_0$ such that

$$\Re \left\{ \sigma(s) \coth \sigma(s) - \sigma(0) \coth \sigma(0) \right\} \geq \frac{\pi s^2}{\varepsilon(s^2 + \varepsilon^2)} + \frac{t^2}{4}$$

for real $s$, where $\sigma(s) = i(\pi + s) + s$, $0 < \varepsilon \leq \varepsilon_0$, $|s| < s_0$.

**Proof.** Write $\zeta = i\varepsilon - s$, so that $\sigma(s) = i\pi - \zeta$ and

$$\sigma(s) \coth \sigma(s) = (i\pi - \zeta) \coth(-\zeta) = (\zeta - i\pi) \frac{\cosh \zeta}{\sinh \zeta}$$
\[ (5.11) \]
\[
\frac{\zeta - i\pi}{\zeta} \left\{ 1 + \frac{\zeta^2}{3} + O(\zeta^4) \right\} = -\frac{\pi}{\varepsilon + 1s} + \frac{i\pi s}{3} - \frac{2i\varepsilon s}{3} + \frac{s^2}{3} + 1 + \frac{\varepsilon^2}{3} - \frac{\varepsilon^2}{3} + O(\zeta^3). 
\]

Therefore

\[ (5.12) \]
\[
\text{Re}\left[ \sigma(s) \coth \sigma(s) - \sigma(0) \coth \sigma(0) \right] = \frac{\pi s^2}{\varepsilon(s^2 + \varepsilon^2)} - \frac{s^2}{3} + O(s^4 + \varepsilon s^2),
\]

since \(\sigma(s) \coth \sigma(s)\) is a symmetric function of \(s\). The inequality (5.10) follows if \(\varepsilon_0\) and \(s_0\) are small enough. \(\square\)

**Lemma 5.13.** Let \(\varepsilon_0\) and \(s_0\) be as in Lemma 5.9. Suppose that

\[ (5.14) \]
\[
\pi - 2a_j \theta_c \leq \varepsilon_0 \quad \text{if and only if} \quad j = m + 1, \ldots, n.
\]

Let

\[ (5.15) \]
\[
\sigma(s) = 2a_n \tau(s) = i2a_n \theta_c + s, \quad s \in \mathbb{R}.
\]

\[ \varepsilon_j = \frac{\alpha_n}{a_j} (\pi - 2a_j \theta_c), \quad j = m + 1, \ldots, n; \]

\[ F_j = \frac{\alpha_n \pi r_j^2}{a_j}, \quad j = m + 1, \ldots, n, \]

and set

\[ (5.16) \]
\[
\Phi(s) = f(x, t, \tau(s)) - f(x, t, \tau(0)).
\]

Then, for \(|s| < s_0\), one has the following estimates:

\[ (5.17) \]
\[
\sum_{j=m+1}^{n} \frac{F_j}{\varepsilon_j} \frac{s^2}{s^2 + \varepsilon_j^2} + C^{-1}r^2 s^2 \lesssim \text{Re} \Phi(s) \lesssim \sum_{j=m+1}^{n} \frac{F_j}{\varepsilon_j} \frac{s^2}{s^2 + \varepsilon_j^2} + C r^2 s^2,
\]

\[ (5.18) \]
\[
\left| \text{Im} \Phi + \sum_{j=m+1}^{n} \frac{F_j}{\varepsilon_j^2} \frac{s^3}{s^2 + \varepsilon_j^2} \right| \lesssim C r^2 s^3, \quad r^2 = \sum_{j=1}^{n} r_j^2.
\]

**Proof.** It follows from the calculation in (5.11) and the form of \(f\) that:

\[ f(x, t, \tau(s)) = -\sum_{j=m+1}^{n} \frac{F_j}{\varepsilon_j + is} + G(s) + t\theta_c - \frac{its}{2a_n}, \]

where \(G\) is holomorphic in a strip containing the real line, while \(G\) and its derivatives are \(O(r^2)\), with estimates that depend only on \(n\) and \(\varepsilon_0\). At the critical point \(\tau(0)\) we have:

\[ 0 = \Phi'(0) = \sum_{j=m+1}^{n} \frac{iF_j}{\varepsilon_j} + \left\{ G'(0) - \frac{it}{2a_n} \right\} \]

so
\[ \Phi(s) = - \sum_{j=m+1}^{n} \left\{ \frac{F_j}{\varepsilon_j + is} - \frac{F_j}{\varepsilon_j} \right\} + \left\{ G'(0) - \frac{it}{2a_n} \right\} s + O(r^2s^2) \]

(5.19)

\[ = \sum_{j=m+1}^{n} F_j \left\{ \frac{1}{\varepsilon_j} - \frac{1}{\varepsilon_j + is} - \frac{is}{\varepsilon_j^2} \right\} + O(r^2s^2). \]

The upper bound for Re \( \Phi \) in (5.17) follows immediately. The lower bound is obtained by writing \( \Phi \) as a sum of \( n \) differences. We note that the real part of \( f \) – and of each of its various pieces – is an even function of \( s \), so the first derivative vanishes at \( s = 0 \). For \( j \leq m \) the difference dominates \( r_j^2 s^2 \) by (2.24), and for \( j > m \) the difference dominates the right hand side of (5.10) multiplied by \( r_j^2/2 \).

The estimate (5.18) is an immediate consequence of (5.19) and the fact that Im \( \Phi \) is odd with respect to \( s \), so the remainder term is \( O(r^2s^3) \) rather than \( O(r^2s^2) \).

**Remark 5.20.** – It will be useful to note that under the assumption (5.14), the \( \alpha_j, j > m \), cannot differ much from \( a_n \). Indeed \( j > m \) implies

\[ \varepsilon_0 \geq \pi - 2\alpha_j \theta_c = \pi - \frac{\alpha_j}{a_n} \cdot 2a_n \theta_c \geq \pi - \frac{\alpha_j}{a_n} \pi. \]

By assumption, \( \alpha_j \leq a_n \), all \( j \), so

(5.21)

\[ 1 - \frac{\varepsilon_0}{\pi} \leq \frac{\alpha_j}{a_n} \leq 1, \quad j = m + 1, \ldots, n. \]

**Lemma 5.22.** – Under the assumptions of Lemma 5.13, the derivative of the function \( \Phi \) of (5.16) is dominated by \( f_c \).

**Proof.** – According to (3.19), \( f_c = S(x, t, 1; \theta_c) \) can be written:

(5.23)

\[ f_c = \sum_{j=1}^{n} \frac{(2\alpha_j \theta_c)^2}{\sin^2(2\alpha_j \theta_c)} \frac{r_j^2}{2} \geq C^{-1} \left\{ \sum_{j=m+1}^{n} \frac{r_j^2}{2} \right\}. \]

This clearly dominates \( \Phi'(s) = d f(x, t, \tau(s))/ds \); see (5.19).

**Lemma 5.24.** – Under the assumptions of Lemma 5.13, the derivatives of \( \Phi \) may be estimated as follows:

(5.25) \[ C^{-1} \left| \frac{d}{ds} \Phi'(s) \right| \leq \left| \Phi'(s) \right| \leq C \left| \frac{d}{ds} \Phi'(s) \right|; \]

(5.26) \[ |s|^2 \left| \frac{d}{ds} \Phi''(s) \right| \leq C \left| \Phi'(s) \right|; \]

(5.27) \[ |s| \left| \frac{d}{ds} \Phi'(s) \right| \leq C \left| \Phi'(s) \right|. \]

**Proof.** – We may assume \( s \geq 0 \). The estimates (5.25) follow from (5.18) and the trivial estimate

\[ \frac{s^3}{s^2 + \varepsilon_j^2} \leq s \left| \frac{d}{ds} \left( \frac{s^3}{s^2 + \varepsilon_j^2} \right) \right| \leq \frac{3s^3}{s^2 + \varepsilon_j^2}. \]

The estimate (5.26) is proved by differentiating a second time. The estimate (5.27) is proved by differentiating \( s^2/(s^2 + \varepsilon_j^2) \) and using (5.17).
We write the volume function \( V(\tau) \) of (5.4) as:

\[
(5.28) \quad V(\tau) = \prod_{j > m} \left\{ \frac{a_n \tau}{a_j (\tau) + i \varepsilon} + W_j (2a_j \tau) \right\} = \sum_{J \subset \{m+1, \ldots, n\}} \left\{ W_J (\tau) \prod_{j \in J} \frac{1}{s - 1 \varepsilon j} \right\},
\]

where the \( W_j \) and \( W_J \) are holomorphic in a strip containing the line \( \text{Im}(\tau) = \pi \). These functions and their derivatives decay exponentially and satisfy estimates that depend only on \( n \) and \( \varepsilon_0 \).

There is an immediate estimate

\[
(5.29) \quad P(x, t; u) \leq C e^{-\varepsilon_0 f(u)} \prod_{j > m} \frac{1}{\varepsilon_j}.
\]

(The factor \( 1/a_n \) comes, as before, from the relation \( d(\tau) = ds/2a_n \).) To improve this estimate when possible, we integrate by parts. For this purpose we introduce some notation. If \( h(\tau) \) is any function that is holomorphic in a neighborhood of the closed half plane \( \text{Im} \xi \leq 0 \) and uniformly \( O(\varepsilon) \) at infinity, we define operations:

\[
(5.30) \quad D_0^{-1} h(s) = \int_{-\infty}^{s} h(s_1) \, ds_1
\]

\[
D_j^{-1} h(s) = \frac{\varepsilon_j^2}{(s - 1 \varepsilon j)^2} D_0^{-1} h(s), \quad j > m.
\]

The integral is independent of the path in the closed half plane. We shall denote by \( Q_k \) any function of the form

\[
(5.31) \quad Q_k = D_k^{-1} D_{k-1}^{-1} \cdots D_1^{-1} \prod_{j \in J - I} \frac{1}{s - 1 \varepsilon j}, \quad k = |J| - 1 \geq 1, \quad 0 \leq I \leq |J| - 1,
\]

where \( J \subset \{m + 1, \ldots, n\} \) and \( |J| \) is the cardinality. We denote by \( I_k \) any integral of the form

\[
(5.32) \quad I_k = \frac{1}{a_n u^m + 1} \int_{-\infty}^{\infty} e^{-f/u} W(s) Q_k(s) \, ds,
\]

where \( W \) and its derivatives are exponentially decreasing, with estimates that depend only on \( n \) and \( \varepsilon_0 \). Thus the integral that defines \( P \) is itself a sum of terms \( I_k \), \( 1 \leq k \leq n - m \).

**Lemma 5.33.** – If \( I_k, \ k > 1, \) has the form (5.32) then it is equal to a sum of terms of the form

\[
(5.34) \quad h(x, t, u) I_k - 1, \quad h(x, t, u) = O(1 + f_c/u).
\]

**Proof.** – In fact (5.32) is equal to

\[
(5.35) \quad -\frac{1}{a_n u^m + 1} \int_{-\infty}^{\infty} \frac{d}{ds} \left\{ e^{-\phi/u} W(s) \right\} Q_{k-1}(s) \, ds,
\]

where \( Q_{k-1} = D_0^{-1} Q_k \). Now:
\[
\frac{d\Phi}{ds} = -i \sum_{j=m+1}^{n} \frac{E_j}{s_j^2 (s - i\varepsilon_j)^2} + h_0(x, t) + O(r^2|s|)
\]
(5.36)
\[
= \sum_{j=m+1}^{n} \frac{E_j}{s_j^2 (s - i\varepsilon_j)^2} + h_0(x, t) + O(r^2|s|), \quad h_v = O(f_c),
\]
see (5.19) and (5.23). The \(O(r^2|s|)\) term yields the \(O(1)\) term in (5.34). The conclusion follows immediately.

**Corollary 5.37.** – The heat kernel \(P(x, t; u)\) is a sum of terms of the form \(hI_1\), \(h = O((1 + f_c/u)^{n-1})\).

We need next to estimate the functions \(Q_1\) associated to the integrals \(I_1\).

**Lemma 5.38.** – For any function \(Q_1\),

\[
|\text{Re} \ Q_1(s)| \leq C \frac{|s|}{s^2 + \varepsilon_n^2}, \quad \int_{-\infty}^{\infty} |\text{Im} \ Q_1(s)| \, ds < C.
\]
(5.39)

**Proof.** – Given a term \(P_J = \prod_{j \in J} (s - i\varepsilon_j)^{-1}\), one has the estimate

\[
|P_J(s)| \leq \frac{1}{(|s| + \varepsilon_n)^{|J|}}.
\]
(5.40)

The factors \(\varepsilon_j^2/(s - i\varepsilon_j)^2\) have modulus \(\leq 1\) in the lower half plane, so it follows inductively that the terms \(Q_k\) obtained by successive integrations satisfy

\[
|Q_k(s)| \leq \frac{C(|J|)}{(|s| + \varepsilon_n)^{|J|/2}}.
\]
(5.41)

This can be improved as follows. Decompose any one of the factors of \(P_J\) as

\[
\frac{1}{s - i\varepsilon_j} = \frac{s}{s^2 + \varepsilon_j^2} + \frac{i\varepsilon_j}{s^2 + \varepsilon_j^2},
\]
so that \(P_J\) itself decomposes as \(P'_J + P''_J\), with

\[
|P'_J(s)| \leq \frac{\varepsilon_j}{s^2 + \varepsilon_j^2} \left(\frac{1}{(|s| + \varepsilon_n)^{|J|/2}}\right).\]

If \(|J| = 1\), then \(P''_J\) is integrable, with a bound which is independent of \(\varepsilon_j\). Suppose that \(|J| = k + 1 > 1\) and suppose that \(Q_1\) is obtained from \(P_J\) by \(k\) integrations \(D_0^{-1}\). We decompose \(Q_1 = Q'_1 + Q''_1\), where \(Q'_1\) is the \(k\)-fold integral of \(P''_J\). Then for \(s \geq 0\),

\[
|Q''_1(s)| \leq \int_{s_1}^{\infty} \cdots \int_{s_k}^{\infty} \left|P''_J(s_1)\right| \, ds_1 \, ds_2 \cdots ds_k
\]
(5.42)
Integrating once more gives
\[ \int_0^\infty |Q_1'(s)| \, ds \leq \frac{1}{(k-1)!} \int_0^\infty \frac{\varepsilon_j}{s^2 + \varepsilon_j^2 + s + \varepsilon_n} \, ds \leq \frac{1}{(k-1)!} \pi.
\]

The same argument applies to the integral for \( s < 0 \), if we integrate from \(-\infty\) to \( s \); recall that the integral is independent of the path in the closed lower half-plane.

This argument proves that when we decompose the factors of \( P_J \) into their real and imaginary parts, each imaginary part leads to a term that is uniformly integrable. Multiplication by \( \varepsilon_j^2(s - i\varepsilon_j)^{-2} \) does not change this. Therefore we only need to consider \( k \)-fold integrals, in the sense of (5.30), of
\[
\prod_{j=1}^{k+1} \frac{s}{s^2 + \varepsilon_j^2} \leq \frac{1}{(s + \varepsilon_n)^{k+1}}.
\]

This yields an imaginary term in \( Q_1 \) only if there is a multiplier
\[
\operatorname{Im} \frac{\varepsilon_j^2}{(s - i\varepsilon_j)^2}.
\]

A \( k \)-fold integral of (5.43) multiplied by (5.44) is dominated by
\[
\frac{\varepsilon_j^3 s}{(s^2 + \varepsilon_j^2)^2 + s + \varepsilon_n} \leq \frac{\varepsilon_j^3}{(s^2 + \varepsilon_j^2)^2},
\]

which is uniformly integrable on \((-\infty, \infty)\). Thus all imaginary terms in \( Q_1 \) satisfy the uniform integrability condition of (5.39).

The real part of \( P_J \) is even as a function of \( s \) when \( |J| \) is even, and odd as a function of \( s \) when \( |J| \) is odd. The opposite correspondence holds for the imaginary part of \( P_J \). The parity of the real part is changed by each integration and unchanged by multiplication by \( \varepsilon_j^2(s - i\varepsilon_j)^{-2} \). Thus the real part of each \( Q_1 \) is odd. To obtain the first estimate in (5.39) it is enough to consider \( s > 0 \). For \( s \geq \varepsilon_n \) the estimate (5.39) is equivalent to the case \( k = 1 \) of (5.41). The estimate is immediate in the case \( |J| = 1 \). If \( |J| > 1 \) we note that the estimate (5.40) for \( Q_1(s) \) yields \( |dQ_1(s)/ds| \leq C(s + \varepsilon_n)^{-2} \), and taking advantage of the fact that \( \operatorname{Re} Q_1(s) \) is odd, which implies that \( \operatorname{Re} Q_1(0) = 0 \), we obtain
\[
|\operatorname{Re} Q_1(s)| \leq C \int_0^s \frac{dx}{(s + \varepsilon)^2} = C \frac{s}{\varepsilon(s + \varepsilon)},
\]

which is equivalent to the first estimate in (5.39) when \( 0 < s < \varepsilon \).

**Lemma 5.45.** – Under the assumptions of Lemma 5.13, suppose that \( \varepsilon_0 \) and \( s_0 \) are sufficiently small. Then for any positive \( u \) the function \( \zeta(s) = -\operatorname{Im} \Phi(s)/u \) is strictly increasing on the
interval \([-s_0, s_0]\). On its domain of definition, the function

\[ h(\zeta(s)) = e^{-\text{Re} \Phi(s)/u} \text{Re} Q_1(s) \frac{1}{\zeta'(s)} \]

satisfies estimates

\[ |h(\zeta)| \leq \frac{C}{|\zeta|}, \quad \left| \frac{dh}{d\zeta}(\zeta) \right| \leq \frac{C}{|\zeta|^2}. \]

Proof. – The estimate (5.18) implies that \( \zeta(s) \) is strictly increasing over a small enough interval. (5.25) implies that \( \frac{d}{ds} = \frac{ds}{d\zeta} \) is comparable to \( \frac{s}{s(1 + f_c)} \), so for \( s \neq 0 \),

\[ C^{-1} \frac{ds}{d\zeta} \leq \frac{s}{s(1 + f_c)} \leq C \frac{ds}{d\zeta}. \]

Since \( \text{Re} \Phi \geq 0 \) and \( \text{Re} Q_1 \) is dominated by \((|s| + \epsilon)^{-1}\), the estimate for \( h \) is immediate.

By (5.27) and (5.48),

\[ \frac{d}{d\zeta} \left( e^{-\text{Re} \Phi/u} \right) = -\frac{\text{Re} \Phi'}{u} e^{-\text{Re} \Phi/u} \frac{ds}{d\zeta} = O \left( \frac{\text{Re} \Phi}{u|s|} e^{-\text{Re} \Phi/u} \frac{s}{\zeta} \right) = O \left( \frac{1}{|\zeta|} \right). \]

Similarly

\[ \frac{dQ_1}{d\zeta} = Q_1' \frac{ds}{d\zeta} = O \left( \frac{1}{s^2 + \epsilon^2} \right) = O \left( \frac{1}{s \zeta^2} \right). \]

By (5.26), \( s^2 \zeta'' = O(\zeta) \), so

\[ \frac{d}{d\zeta} \left( \frac{ds}{d\zeta} \right) = \frac{d}{d\zeta} \left( \frac{1}{\zeta} \right) = -\frac{1}{(\zeta')^3} \zeta'' \frac{ds}{d\zeta} = -\frac{\zeta''}{(\zeta')^3} = O \left( \frac{|s|^3}{|\zeta|} \right) = O \left( \frac{|s|}{\zeta^2} \right). \]

These three estimates yield the estimate for \( \frac{dh}{d\zeta} \). \( \square \)

Proof of Theorem 5.5. – We begin by proving the estimate

\[ P(x, t; u) \leq C \frac{e^{-f_c/u}}{a_n u^{n+1}} \left( 1 + \frac{f_c}{u} \right)^{n-m-1}. \]

Recall that \( 2f_c = d(x, t)^2 \). Corollary 5.37 implies that it is enough to prove the following estimate for integrals of type \( I_1 \):

\[ |I_1| \leq C \frac{e^{-f_c/u}}{a_n u^{n+1}}. \]

Any integral \( I_1 \) has the form

\[ I_1 = \frac{e^{-f_c/u}}{a_n u^{n+1}} \int_{-\infty}^{\infty} e^{-\Phi/u} Q_1 W(s) \, ds. \]

By Lemma 5.38, \( \text{Im} Q_1 \) is integrable, while the rest of the integrand is bounded. Therefore we only need to prove (5.50) with \( Q_1 \) replaced by \( \text{Re} Q_1 \). Because of the exponential decay of \( W \)
and the estimates on $Q_1$, it is enough to consider the integral over the interval $|s| \leq s_0$, where $s_0$ is the constant of Lemma 5.9. Note also that $Q_1(s)[W(s) - W(0)]$ is bounded, so we may replace $W$ by a constant. Proving (5.49) has been reduced to proving:

$$\left| \int_{-s_0}^{s_0} e^{-\Phi/\pi} \text{Re} \ Q_1(s) \, ds \right| \leq C. \quad (5.51)$$

As noted in the proof of Lemma 5.38, $\text{Re} \ Q_1$ is odd. Since $\text{Re} \ \Phi(s)$ is an even function of $s$ and $\text{Im} \ \Phi(s)$ is an odd function of $s$, the integral to be estimated is reduced to:

$$\int_{-s_0}^{s_0} e^{-\text{Re} \ \Phi/\pi} \sin(- \text{Im} \ \Phi/\pi) \, \text{Re} \ Q_1 \, ds. \quad (5.52)$$

In the notation of Lemma 5.45, this integral is

$$\int_{-\xi(v_0)}^{\xi(v_0)} h(\zeta) \sin \zeta \, d\zeta. \quad (5.52)$$

The estimate on $h$ in (5.47) gives a uniform bound for integration over $|\zeta| < 1$. Integration by parts and the estimate on $h'$ in (5.47) gives a uniform bound for the rest of the integral (5.52). This completes the proof of (5.49).

To complete the proof of Theorem 5.5, we need to consider the effect of taking fewer singularities into account. First, since $\varepsilon_n = \pi - 2a_n \theta_c$, (5.14) implies that

$$\frac{a_j}{a_n} < \frac{\pi - \varepsilon_0}{\pi - \varepsilon_n}, \quad j = 1, \ldots, m. \quad (5.53)$$

We assume that $\varepsilon_0 > 0$ has been fixed, and that $\varepsilon_n < \varepsilon_0$ (otherwise we have an estimate (5.8)). Then (5.53) implies that the quotients $a_n/(a_n - a_j)$, $j = 1, \ldots, m$, in (5.6) are bounded from above, so that (5.6) is equivalent to an estimate involving only $j > m$:

$$P(x, t; u) \leq C \frac{e^{-f_c/u}}{a_n u^{n+1}} \prod_{j=m+1}^{n-1} \min \left\{ \frac{a_n}{a_n - a_j}, 1 + \frac{f_c}{u} \right\}. \quad (5.54)$$

We note that (5.15) yields the following estimates:

$$\varepsilon_j \geq \pi - 2a_j \theta_c = \pi - \frac{a_j}{a_n} (2a_n \theta_c) > \pi - \frac{a_j}{a_n} \pi = \pi \frac{a_n - a_j}{a_n}. \quad (5.55)$$

Therefore $1/\varepsilon_j < a_n/(a_n - a_j)$, and (5.54) follows from

$$P(x, t; u) \leq C \frac{e^{-f_c/u}}{a_n u^{n+1}} \prod_{j=m+1}^{n-1} \min \left\{ \frac{1}{\varepsilon_j}, 1 + \frac{f_c}{u} \right\}. \quad (5.56)$$

We shall prove (5.56). We already have (5.49), so we are interested in the case when $1/\varepsilon_j < 1 + f_c/u$ for some $j = m + 1, \ldots, n$. Let us start by carrying out the previous analysis
using only the indices \( j > m + 1 \), i.e. \( j = m + 1 \) has been omitted from the sum in the representation (5.19) of \( \Phi \), and in the products (5.31) for the \( Q_j \). Then the term \( W \) in integrals of the type (5.32) has an extra factor \( 2a_{m+1}\tau/\sinh(2a_{m+1}\tau) \). To reach a term \( Q_1 \) we have at most \( n - m - 2 \) integrations by part, but the estimates for \( W \) may be worse by a factor that is \( \text{O}(1/\varepsilon_{m+1}) \), and each derivative of \( W \) may give another such factor. If all derivatives fall on the exponential we obtain a term dominated by

\[
\frac{e^{-f_c/u}}{a_n u^{n+1}} \frac{1}{\varepsilon_{m+1}}^{n-m-2}.
\]

If all derivatives fall on \( W \) we obtain a term dominated by

\[
\frac{e^{-f_c/u}}{a_n u^{n+1}} \frac{1}{\varepsilon_{m+1}}^{n-m-1}.
\]

We may also obtain all the other terms in between (5.57) and (5.58). If \( 1/\varepsilon_{m+1} < 1 + f_c/u \), then \( P(x, t; u) \) is dominated by (5.57), and this, with (5.49), implies that

\[
P(x, t; u) \leq C \frac{e^{-f_c/u}}{a_n u^{n+1}} \min \left\{ \frac{1}{\varepsilon_{m+1}}, 1 + \frac{f_c}{u} \right\} \left( 1 + \frac{f_c}{u} \right)^{n-m-2}.
\]

Next we carry out the previous analysis only for \( j > m + 2 \). If \( 1/\varepsilon_{m+2} < 1 + f_c/u \), we obtain

\[
P(x, t; u) \leq C \frac{e^{-f_c/u}}{a_n u^{n+1}} \frac{1}{\varepsilon_{m+1} \varepsilon_{m+2}} \left( 1 + \frac{f_c}{u} \right)^{n-m-3}.
\]

As long as \( 1/\varepsilon_{m+1} < 1 + f_c/u \), (5.57) and (5.59) imply

\[
P(x, t; u) \leq C \frac{e^{-f_c/u}}{a_n u^{n+1}} \frac{1}{\varepsilon_{m+1}} \min \left\{ \frac{1}{\varepsilon_{m+2}}, 1 + \frac{f_c}{u} \right\} \left( 1 + \frac{f_c}{u} \right)^{n-m-3}.
\]

If \( 1/\varepsilon_{m+1} > 1 + f_c/u \), one has (5.49); note that \( 1/\varepsilon_{k+1} > 1 + f_c/u \) implies \( 1/\varepsilon_{j} > 1 + f_c/u \), \( j > k \). Consequently we have derived the following estimate:

\[
P(x, t; u) \leq C \frac{e^{-f_c/u}}{a_n u^{n+1}} \prod_{j=m+1}^{m+2} \min \left\{ \frac{1}{\varepsilon_{j}}, 1 + \frac{f_c}{u} \right\} \left( 1 + \frac{f_c}{u} \right)^{n-m-3}.
\]

Continuing in this manner we derive (5.56). The completes the proof of Theorem 5.5. \( \square \)

**Corollary 5.60.** – Assuming (3.2), one has:

\[
P(x, t; u) \leq C \frac{e^{-d(x, t)^2/2u}}{a_n u^{n+1}} \left( 1 + \frac{d(x, t)^2}{u} \right)^{n-p-1},
\]

where \( C \) depends only on \( n \) and on an upper bound for \( a_n \).
Small time behaviour of the heat kernel

**Theorem 5.62.** Given \((x, t) \in H_n\), assume that (3.25) has a solution \(c \in \mathbb{T}_0 = \mathbb{R} / 2 \pi \mathbb{N}\), i.e. \((x, t)\) is generic in \(H_n\); then

\[
P(x, t; u) = \frac{1}{(2\pi u)^{n+1}} e^{-d(x, t)^2/2u} \left( \Theta(x, t) \sqrt{2\pi u} + O(u) \right),
\]

as \(u \to 0^+\), where

\[
\Theta(x, t) = \left( \frac{1}{f''(x, t, i\theta_c(x, t))} \right)^{1/2} \prod_{j=1}^n \frac{2a_j \theta_c}{\sin(2a_j \theta_c)},
\]

and \(f'' = d^2 f / dx^2\).

**Proof.** We argue as in Theorem 2.42. We set \(s = f(x, t, s + i\theta_c) - f(x, t, i\theta_c)\), \(s \in \mathbb{R}\), and note that (1.72) yields

\[
\Re \Phi(s) \geq cs^2 \|x\|^2, \quad |s| \leq 1,
\]

with a positive constant \(c\). Again we write

\[
I = \int_{-\infty}^{\infty} e^{-\Phi(s)/u} V(s + i\theta_c) \, ds
\]

\[
= \left\{ \int_{-\delta}^{\delta} + \int_{|s|>\delta} \right\} e^{-\Phi(s)/u} V(s + i\theta_c) \, ds
\]

\[
= I_0 + I_0',
\]

where \(\delta\) is still to be chosen, \(\delta \in (0, 1)\). First

\[
|I_0'| \leq e^{-\Re \Phi(\delta)/u} \int_{-\infty}^{\infty} \left| V(s + i\theta_c) \right| \, ds \leq C e^{-\Re \Phi(\delta)/u}, \quad C = C(x, t) > 0.
\]

As for \(\lim_{u \to 0^+} I_0\), we use the method of stationary phase. Note that \(\Phi(0) = 0\), \(\Phi'(0) = f'(x, t, i\theta_c(x, t)) = 0\), so (2.24) implies

\[
\Phi''(0) = f''(x, t, i\theta_c(x, t)) \geq \frac{2a_1 \|x\|^2}{3}.
\]

Consequently one has

\[
\Phi(s) = \Phi''(0) \frac{s^2}{2} (1 + O(|s|)).
\]
This allows us to choose a $\delta$, $0 < \delta < \min(1, \pi / 2a_n - \theta_c)$, and introduce a new variable $z$, such that

$$\Phi(s) = \Phi''(0) \frac{z^2}{2}, \quad |s| < \delta.$$  

Then we write $I_k$ in the form (2.53), and note that the argument following (2.53) which proves Theorem 2.42, also completes the proof of Theorem 5.62. \hfill \Box

Remark 5.71. – $\tau = 0 \iff \theta_c = 0$, and (5.63), (5.64) yield

$$P(x, 0; u) \sim \frac{1}{(2\pi u)^{n+1/2}} \frac{e^{-|x|^2/2u}}{u^{2n}} \quad u \to 0+.$$  

Next we assume that

$$|t| \geq \sum_{j=1}^{n} a_j \mu(2a_j \pi / 2a_n) r_j^2,$$

which implies

$$r_{p+1} = \cdots = r_n = 0.$$  

Again the asymptotics can be computed explicitly, but we state a complete result only when strict inequality holds in (5.73).

Theorem 5.75. – Suppose that, for a given $(x, t) \in H_n$, we have

$$|t| > \sum_{j=1}^{n} a_j \mu(2a_j \pi / 2a_n) r_j^2,$$

or, that (5.74) holds and $E_p(x, t) > 0$, see (3.105); then

$$P(x, t; u) = \frac{1}{(2\pi)^n} \frac{1}{(n - p - 1)!} \frac{e^{-d(x, t)^2/2u}}{u^{2n-p}} \left[ \Theta_p(x, t) + O(u) \right],$$

where

$$\Theta_p(x, t) = \left( \frac{\pi}{2a_n} \right)^{n-p} \left| t - \sum_{j=1}^{p} a_j \mu \left(2a_j \pi / 2a_n\right) r_j^2 \right|^{n-p-1} \prod_{j=1}^{p} \frac{2a_j \pi / 2a_n}{\sin(2a_j \pi / 2a_n)}.$$  

Proof. – Suppose first that $x \neq 0$ and let $q$ be the largest index such that $r_q \neq 0$. By assumption $q \leq p$, so $a_q \leq a_p < a_n$. The modified action $f$ can be considered as being taken with respect to the subgroup $H_q$. The point $(x, t)$ is generic on $H_q$, so there is a critical value $\theta_c$ with $|2a_q \theta_c| < \pi$. We may assume $t > 0$, so $\theta_c > 0$. We use again the fact that $\text{Re} \ f$ is bounded below and $V$ decreases exponentially in the strip to displace the contour of integration (2.4) to the line $\text{Im} \ \tau = \theta_c$, picking up residues that correspond to the poles of $V$ at the points $i\tau / 2a_j$, $a_j > a_q$. Thus we have

$$\int_{-\infty}^{\infty} e^{-f(x, t, \tau)/u} V(\tau) \, d\tau = \int_{-\infty}^{\infty} e^{-f(x, t, \tau+i\theta_c)/u} V(\tau) \, d\tau$$
\[ (5.79) \quad + \sum_{a_j > a_q} 2\pi i \text{Res}(e^{-f(x,t,\tau)/u} V(\tau); \ \tau = i\pi/2a_j). \]

The integral along the contour $\text{Im} \ \tau = \theta_c$ is dominated by

\[ (5.80) \quad e^{-d_q(x,t)^2/2u}, \]

where $d_q$ denotes the Carnot–Caratheodory distance in $H_q$. According to the proof of Lemma 3.91, $d_q(x,t) > d(x,t)$. If $2a_j \theta_c = \pi$ for some $a_j > a_q$, then we slightly reduce $\theta_c$; Lemma 3.91 works and we still have (5.80). Therefore the contribution of this integral to (5.77) can be absorbed in the remainder term.

To evaluate the residue at $i\pi/2a_n$ we note that

\[ (\sinh(2a_n \tau))^{-(n-p)} = \left( \tau - i \frac{\pi}{2a_n} \right)^{-(n-p)} \left( -\frac{1}{2a_n} \right)^{n-p} + O(\tau). \]

where $h(\tau) = O(\tau - i\pi/2a_n)$. Thus we need the $\tau$-derivative of order $n-p-1$ of the function:

\[ (5.81) \quad e^{-f(x,t,\tau)/u} \left( \prod_{j=1}^{p} \frac{2a_j \tau}{\sinh(2a_j \tau)} \right)^{n-p} \left( \frac{-1}{2a_n} \right)^{n-p} + O(\tau) \]

evaluated at $\tau = i\pi/2a_n$. The leading term in $u$ as $u \to 0+$ is the one with the largest negative power of $u$, and this is obtained by applying all the $n-p-1$ $\tau$-derivatives to $e^{-f/u}$. Thus

\[ (5.82) \quad \text{Res}(e^{-f(x,t,\tau)/u} V(\tau); \ \tau = i\pi/2a_n) \]
\[ = \left\{ -\frac{f_x(x,t,i\pi/2a_n)}{u} \right\}^{n-p-1} e^{-f(x,t,i\pi/2a_n)/u} \]
\[ \times \frac{2\pi i}{(n-p-1)!} \prod_{j=1}^{p} \frac{2a_j \pi/2a_n}{\sin(2a_j \pi/2a_n)} \left( \frac{-i\pi}{2a_n} \right)^{n-p} \left[ 1 + O(u) \right], \]

where $f_x$ denotes the derivative

\[ (5.83) \quad \frac{\partial f}{\partial \tau}(x, t, i \frac{\pi}{2a_n}) = -i \left[ t - \sum_{j=1}^{p} a_j \mu \left( 2a_j \frac{\pi}{2a_n} \right) \right]. \]

We note that (5.82) accounts for the main term of (5.77). The remaining residues contribute terms of the same form but with exponentials that involve the Carnot–Caratheodory distance with respect to $H_p$ or smaller subgroups and thus are strictly larger than $d(x,t)$; see the proof of Lemma 3.91. Consequently the remaining residues can be absorbed in the remainder term.

When $x = 0$ we may assume $t > 0$; then

\[ (5.84) \quad \frac{\partial f}{\partial \tau}(x, t, \tau) = -it. \]

We displace the contour of integration for the heat kernel to a line $\text{Im} \ \tau = \theta > \pi/2a_n$:  

The integral has a factor $e^{-\theta/\eta}$, and is thus exponentially smaller than the residue as $u \to 0^+$. The computation of the residue proceeds as above, and we use (3.103). This completes the proof.

Remark 5.86. – When $(x, t)$ is a caustic for which equality holds in (5.73), the same procedure can be followed. However $\frac{\partial f}{\partial \tau} = 0$ at $\tau = i \text{sgn} t / 2a_n$, so the determination of the principal term of the residue is more complicated.

Remark 5.87. – We always have

$$-\log P_{\mu}(x, t; u) \sim \frac{1}{2} d(x, t)^2.$$ 

On the other hand the power of $u$ in the coefficient of the exponential is $-n - \frac{1}{2}$ at the generic points in Theorem 5.62 and $p - 2n$ at the non-generic points $(x, t)$ of Theorem 5.75; here $n - p$ is the number of $a_j$’s equal to the largest, $a_n$. Even in the isotropic case, all $a_j$ equal, the points $(0, t), t \neq 0$ come under Theorem 5.75. It follows that asymptotics like (2.43) cannot be uniform in any neighborhood of the origin.

REFERENCES


