Advanced Studies in Pure Mathematics 31, 2001
Taniguchi Conference on Mathematics Nara '98
pp. 1–49

Hardy Spaces, Carleson Measures and a Gradient Estimate for Harmonic Functions on Negatively Curved Manifolds

Hitoshi Arai

Abstract.

In this paper we study Hardy spaces, BMO, Carleson measures, Green potential and Bloch functions on a Cartan-Hadamard manifold M of pinched negative curvature. Further, using our results on Carleson measure and BMO, we give a gradient estimate for harmonic functions on M. It is different from Yau's gradient estimates, and is applied to the existence problem of harmonic Bloch functions described in §10. We deal also with boundary behavior of harmonic Bloch functions on M.

§1. Introduction

In their paper [22], C. Fefferman and E. Stein developed the theory of Hardy spaces of harmonic functions on the upper-half spaces endowed with the Euclidean metric. As is well known, their theory have played crucial roles in the classical harmonic analysis. In 1981 and 1982, D. Jerison, E. Fabes, C. Kenig and U. Neri extended some important parts of the theory of Hardy spaces on the upper-half spaces to more general Euclidean domains with non-smooth boundaries ([21], [20] and [26]).

On the other hand, after the work of A. Korányi on the boundary behavior of harmonic functions on symmetric spaces ([28]), Hardy spaces have been investigated also for symmetric spaces. In particular, D. Geller ([23]) and A. Debiard ([16], [17]) studied Hardy spaces on Siegel upper-half spaces of type II, and in somewhat later, P. Cifuentes extended the classical theorems on the probabilistic characterization and area integral characterization of Hardy spaces to rank one symmetric spaces of noncompact type ([12] and [13]).

Now in this paper we study Hardy spaces, BMO and Carleson measures on a complete, simply connected n-dimensional Riemannian manifold (M, g) such that the sectional curvatures K_M of M satisfy $-\kappa_2^2 \leq$

Received July 3, 1999.

$K_M \leq -\kappa_1^2$, for some constants κ_1 and κ_2 with $0 < \kappa_1 \leq \kappa_2 < +\infty$. A typical example of such a manifold is a rank one symmetric space of noncompact type, but also many other examples are known.

For the manifold M, the boundary behavior of harmonic functions has been studied by several authors: For instance, the solvability of the Dirichlet problem for the manifold M was proved by Anderson [2] and D. Sullivan [46], and moreover in Anderson and Schoen [3] it was proved that the Eberlein-O'Neill compactification of M is homeomorphic to the Martin compactification (see also Kifer [27], Ancona [1]). Then Anderson and Schoen [3], Ancona [1], Arai [5], Mouton [38], and Cifuentes and Korányi [14] studied boundary behavior of harmonic functions on M.

This paper consists of two parts. First part is concerned with the Hardy spaces of harmonic functions on M, and the second part with Carleson measures. Then we will give an application to Bloch functions on M.

We will begin in §2 with a quick review of some preliminaries about harmonic functions on M. In §3 we will define analogues to the manifold M of the classical Stoltz domain and of the classical Hardy spaces of harmonic functions. Section 4 contains a review real analysis on the sphere at infinity. In §5 we prove some elementary properties of Hardy spaces H^p, and then in §6 we prove that the Hardy space H^1, atomic Hardy space H^1_{atom} and probabilistic Hardy space H^1_{prob} are mutually equivalent. Some results in this section were announced already in our paper [5], but in the announcement we assumed an additional geometric condition in order to show that every $(1,\infty)$-atom is in H^1. In the present paper we call it the condition (β). As pointed out in [5], there are some examples of manifolds having the condition (β). However, recently Cifuentes and Korányi [14] proved that M possesses always the condition. Therefore combining theorems announced in [5] with their result, we gain the equivalence of the three different definitions of Hardy spaces.

In the second part of this paper we study Carleson measure and its application to Bloch functions on M. In §7 we are concerned with relationship between Carleson measures and L^p boundedness of the Martin integral, and in §8 we give a characterization of Carleson measure in terms of a certain Green potential. Using it, in §9 we prove Carleson measure characterization of BMO functions. In the classical Euclidean case, this characterization was found by C. Fefferman and E. Stein ([22]), and in the case of the complex unit ball endowed with the Bergman metric, the characterization was proved by Jevtic [25]. However, his proof is based on the nature of the ball. Our proof is different from it. In §10 we will study harmonic Bloch functions defined on M. From our

Carleson measure characterization of BMO functions, we give a gradient estimate for harmonic functions on M (see (45)), and prove existence of unbounded harmonic Bloch function on M. Moreover in §11 we study boundary behavior of unbounded harmonic Bloch functions, and prove a generalization of Lyons' theorem on the law of iterated logarithm.

Notation: In this paper we fix a point o in M as a reference point. The constants depending only on g, n, κ_1, κ_2 and o will usually be denoted by C, C' or C_j ($j = 1, 2, \ldots$). But C and C' may change in value from one occurrence to the next, while constants C_j ($j = 1, 2, \ldots$) retain a fixed value. For two nonnegative functions f and g defined on a set U, the notation $f \lessapprox g$ indicate that $f(x) \leq Cg(x)$ for all $x \in U$, and $f \approx g$ means that $f \lessapprox g$ and $g \lessapprox f$.

Acknowledgement. The author would like to thank to Professor Adam Korányi for his suggestion and encouragement. The author is honored to be invited to the last Taniguchi conference on Mathematics, and thanks to the organizing comittee of the conference.

§2. Preliminaries

In this section, we review Harnack type inequalities for positive harmonic functions and some facts about the Martin compactification. Both are very important for us.

A C^2 function f in an open set U of M is called harmonic in U if $\Delta_M f = 0$ in U, where Δ_M is the Laplace-Beltrami operator of (M, g). For $x \in M$ and $r > 0$, let $B(x, r) := \{y \in M : d(x, y) < r\}$, where $d(\ ,\ )$ is the distance function with respect to the Riemannian metric g. Then Moser's Harnack inequality implies

Theorem H (Interior Harnack inequality). *Let $R > 0$. Then for every positive harmonic function u on a ball $B(x, 2R)$,*

$$C_{1,R}^{-1} u(y) \leq u(x) \leq C_{1,R} u(y), \tag{1}$$

for all $y \in B(x, R)$, where $C_{1,R}$ is a positive constant depending only on M and R.

In this paper we will use the so-called boundary Harnack inequalities. They were proved firstly by Anderson and Schoen [3] and then also by Ancona [1]. To describe them we need some notation. Denote by $S(\infty)$ the sphere at infinity of M and by $\overline{M}$ the Eberlein and O'Neill compactification $M \cup S(\infty)$ of M (see [19] for definitions). For $x \in M$ and $y \in \overline{M}$ ($x \neq y$), let γ_{xy} be the unit speed geodesic such that

$\gamma_{xy}(0) = x$ and $\gamma_{xy}(t) = y$ for some $t \in (0, +\infty]$. Since such a number t is unique, we denote it by t_{xy}.

For $p \in M$, $v \in T_pM$ and $\delta > 0$, let $C(p, v, \delta)$ be the cone about the tangent vector v of angle δ defined by

$$C(p, v, \delta) := \{x \in \overline{M} \setminus \{p\} : \angle_p(v, \dot{\gamma}_{px}(0)) < \delta\},$$

where $\angle_p$ denotes the angle in T_pM and $\dot{\gamma}_{px}(t)$ is its tangent vector at t.

The following is called boundary Harnack inequality:

Theorem BH1 (Anderson and Schoen [3]: see also [1], [27]). *Let $p \in M$ and $v \in T_pM$ with $g_p(v, v) = 1$. Denote $C = C(p, v, \pi/4)$ and $T = C(p, v, \pi/8) \setminus B(p, 1)$. Let u and h be positive harmonic functions on $C \cap M$, continuous up to the closure $\overline{C}$ of C in $\overline{M}$ and vanishing on $\overline{C} \cap S(\infty)$. Then*

$$C_1 \exp\{-C_2 d(p, x)\} \leq \frac{u(x)}{u(p_0)} \leq C_3 \exp\{-C_4 d(p, x)\}, \tag{2}$$

$$C_5^{-1} \frac{u(p_0)}{h(p_0)} \leq \frac{u(x)}{h(x)} \leq C_6 \frac{u(p_0)}{h(p_0)}, \tag{3}$$

for $x \in T$, where $p_0 = \exp_p(v)$, and $C_1, \ldots, C_6$ are constants depending only on M.

For $z \in M \setminus \{o\}$ and $t \in \mathbf{R}$, we denote

$$C(z, t) = C(\gamma_{oz}(t_{oz} + t), \dot{\gamma}_{oz}(t_{oz} + t), \pi/4), \quad \text{and} \quad z(t) = \gamma_{oz}(t_{oz} + t).$$

In this paper we will use the following variation of Theorem BH1:

Theorem BH2 (Ancona [1]). (1) *Let u and h be positive harmonic functions on a cone $C(z, t) \cap M$ and vanishing continuously on $C(z, t) \cap S(\infty)$. Then*

$$C_7^{-1} \frac{u(z(t+1))}{h(z(t+1))} \leq \frac{u(x)}{h(x)} \leq C_7 \frac{u(z(t+1))}{h(z(t+1))}, \quad \text{for all } x \in C(z, t+1) \cap M,$$

where C_7 is a positive constant depending only on M.

(2) *Let u and h be positive harmonic functions on $M \setminus C(z, t+1)$, and vanishing continuously on $(\overline{M} \setminus C(z, t+1)) \cap S(\infty)$. Then*

$$C_8^{-1} \frac{u(z(t))}{h(z(t))} \leq \frac{u(x)}{h(x)} \leq C_8 \frac{u(z(t))}{h(z(t))}, \quad \text{for all } x \in M \setminus C(z, t),$$

where C_8 is a positive constant depending only on M.

The second statement in Theorem BH2 seems to be a little different from (1), but it is actually a special case of Theorem 5′ in [1]. For reader's convenience, we give a proof of Theorem BH2 (2) in Appendix 1.

The following theorem is an important consequence of boundary Harnack inequalities:

Theorem AS (Anderson and Schoen [3]; [1], [27]). (1) *The Martin compactification of M with respect to the Laplacian Δ_M is homeomorphic to Eberlein and O'Neill's compactification $\overline{M}$, and the Martin boundary consists only of minimal points.*

(2) *For every $z \in M$, there exists a unique function $K_z(x, Q)$ ($Q \in S(\infty)$, $x \in \overline{M} \setminus \{Q\}$) such that for every $Q \in S(\infty)$,*

(4) $\quad K_z(\,\cdot\,, Q)$ is positive harmonic on M,

(5) $\quad K_z(\,\cdot\,, Q)$ is continuous on $\overline{M} \setminus \{Q\}$,

(6) $\quad K_z(Q', Q) = 0$ for all $Q' \in S(\infty) \setminus \{Q\}$, and

(7) $\quad K_z(z, Q) = 1$.

(*This function is called the Poisson kernel normalized at z.*)

(3) *For every $z \in M$ and for every positive harmonic function u on M, there exists a unique Borel measure m_u^z on $S(\infty)$ such that*

$$u(x) = \int_{S(\infty)} K_z(x, Q)\, dm_u^z(Q), \quad x \in M. \tag{8}$$

(*The measure m_u^z is called a Martin representing measure relative to u and z.*)

Throughout this paper, we denote $K(x, Q) = K_o(x, Q)$, and write simply ω^x the Martin representing measure relative to the constant function 1 and $x \in M$. It is called the harmonic measure relative to x. In particular, let $\omega = \omega^o$. Note that $\omega^x(S(\infty)) = 1$ and $d\omega^x(Q) = K(x, Q)\, d\omega(Q)$, for all $x \in M$.

Theorem H yields that for any compact sets $E \subset M$, there exists a positive constant c_E satisfying

$$c_E^{-1}\omega^x(A) \le \omega^y(A) \le c_E \omega^x(A) \tag{9}$$

for all $x, y \in E$ and for all Borel sets $A \subset S(\infty)$.

§3. Approach regions and Hardy spaces

In order to study Hardy space H^1 and BMO on M, we begin with recalling two analogues for M of the classical Stoltz region. First is the following:

Definition 3.1 (Anderson and Schoen [3]). For $Q \in S(\infty)$ and $d > 0$, let

$$T_d(Q) = \bigcup_{t>0} B(\gamma_{oQ}(t), d). \tag{10}$$

Following [3], we call such a set the nontangential region at $Q \in S(\infty)$.

In this paper we will be mainly concerned with another analogue of the Stoltz region in some technical reasons: For $x \in M$ and $t \in \mathbf{R}$, let

$$\Delta(x,t) = C(x,t) \cap S(\infty). \tag{11}$$

Definition 3.2 ([5]). For $Q \in S(\infty)$ and $\alpha \in \mathbf{R}$, let

$$\Gamma_\alpha(Q) = \{z \in M : Q \in \Delta(z,\alpha)\}, \tag{12}$$

and we call this set an admissible region at Q.

This definition is motivated by the following consideration about classical Stoltz domains in the upper half-plane $\mathbf{R}^2_+$: Recall that the Stoltz domain $S_\alpha(x)$ at $x \in \mathbf{R}$ with angle $\alpha > 0$ is the set $\{(y,t) \in \mathbf{R} \times (0,+\infty) : |y-x| < \alpha t\}$. For $z\,(= (z_0,t)) \in \mathbf{R}^2_+$, let $\mathcal{D}(z,\alpha) := \{y \in \mathbf{R}$: the angle at z between the segment $\overline{zy}$ and $\overline{zz_0}$ is less than $\alpha\}$, where $\overline{pq}$ is the segment joining p and q for $p, q \in \overline{\mathbf{R}}^2_+$. Denote $S'_\alpha(x) = \{z \in \mathbf{R}^2_+ : x \in \mathcal{D}(z,\alpha)\}$. Then $S_\alpha(x) = S'_\alpha(x)$. Now we can easily see that our set $\Delta(z,\alpha)$ corresponds to the domain $\mathcal{D}(z,\alpha)$, and $\Gamma_\alpha(x)$ to the set $S'_\alpha(x)$.

We note that if we define Hardy spaces by using our admissible domains, we may apply "tent" method to our case as we will show later. By this reason, in this paper, we use the admissible regions instead of nontangential regions in the sense of [3]. However, it should be noted that both are closely related to each other:

Theorem CK1 (Cifuentes and Korányi [14]). *Two families of approach regions $\{T_d(Q)\}$ and $\{\Gamma_\alpha(Q)\}$ are equivalent in the sense of* [14], *that is, for all $\alpha \in \mathbf{R}$, there exists $d > 0$ and $R > 0$ such that for all $Q \in S(\infty)$,*

$$\Gamma_\alpha(Q) \cap (M \setminus B(o,R)) \subset T_d(Q) \cap (M \setminus B(o,R)), \tag{13}$$

and vice versa.

For the notational convenience, for every $f \in L^1(\omega)$ $(= L^1(S(\infty), \omega))$, let

$$\tilde{f}(x) = \begin{cases} \int_{S(\infty)} K(x,Q) f(Q)\, d\omega(Q), & x \in M, \\ f(x), & x \in S(\infty). \end{cases}$$

Then $\tilde{f}$ is harmonic on M. Now Hardy spaces $H^p(\omega, \alpha)$ $(1 \le p \le \infty$, $\alpha \in \mathbf{R})$ are defined as follows: For a function u on M, let

$$N_\alpha(u)(Q) := \sup_{x \in \Gamma_\alpha(Q)} |u(x)|, \quad Q \in S(\infty),$$

and let

$$H^p(\omega, \alpha) = \{f \in L^p(\omega) : N_\alpha(\tilde{f}) \in L^p(\omega)\}.$$

Denote $\|f\|_{H^1(\omega,\alpha)} := \|N_\alpha(\tilde{f})\|_{L^p(\omega)}$.

Here we should note that when u is a continuous function on M, then $N_\alpha(u)$ is lower semicontinuous on $S(\infty)$. Indeed for every $\lambda > 0$, the set $E := \{Q \in S(\infty) : N_\alpha(u) > \lambda\}$ is open. For if $Q \in E$, then there exist $\varepsilon > 0$ and $z \in \Gamma_\alpha(Q)$ such that $|u(z)| > \lambda + \varepsilon$. By the definition of $\Gamma_\alpha(Q)$, we have $Q \in \Delta(z, \alpha)$. We can take an open subset U of $S(\infty)$ such that $Q \in U \subset \Delta(z, \alpha)$. Since $z \in \Gamma_\alpha(Q')$ for all $Q' \in U$, we have $U \subset E$. This implies that E is open.

As we will see in §5, these spaces $H^p(\omega, \alpha)$ and $H^p(\omega, \beta)$ are equivalent for every $\alpha, \beta \in \mathbf{R}$. We denote

$$H^p(\omega) = H^p(\omega, 0), \quad \text{and} \quad \|f\|_{H^p(\omega)} = \|f\|_{H^p(\omega,0)}.$$

Remark 1. As in the classical case, another Hardy spaces of harmonic functions on M are defined by

$$H^p(M) := \{u : u \text{ is harmonic on } M \text{ and } N_0(u) \in L^p(\omega)\}, \quad 1 \le p \le \infty.$$

See Appendix 3 for these Hardy spaces.

§4. Real analysis at infinity – Quick Review –

Before going to the main body of this paper, we set down the basic facts about real analysis on the sphere at infinity $S(\infty)$ of M. All theorems stated in this section follow immediately from results in [3] and, in particular, from the abstract theory of real analysis due to Coifman and Weiss [15]. For any $Q \in S(\infty)$ we define $\Delta_t(Q)$ to be the "ball" in $S(\infty)$ centered at Q,

$$\Delta_t(Q) := \Delta(\gamma_{oQ}(t), 0) \ (= C(\gamma_{oQ}(t), \dot{\gamma}_{oQ}(t), \pi/4) \cap S(\infty)),$$

when $t \geq 0$, and let $\Delta_t(Q) = S(\infty)$ when t is negative. Then we can see that the family of the sets $\{\Delta_t(Q)\}$ defines a quasi-distance ρ on $S(\infty)$ which makes the triple $(S(\infty), \rho, \omega)$ is a space of homogeneous type in the sense of Coifman and Weiss [15] as follows: By [3] the family of "balls" $\{\Delta_t(Q)\}$ satisfies the following properties:

(H1) For all $s > 0$ and $r > 0$

$$S(\infty) = \lim_{t\to-\infty} \Delta_t(Q) \supset \Delta_r(Q) \supset\supset \Delta_{r+s}(Q) \supset \lim_{t\to\infty} \Delta_t(Q) = \{Q\},$$

where $A \supset\supset B$ means that A contains the closure of B. Furthermore, $\{\Delta_r(Q) : r \in \mathbf{R}\}$ is a fundamental system of neighborhoods of Q.

(H2) Let $Q_1, Q_2 \in S(\infty)$ and $r \in \mathbf{R}$. If $\Delta_r(Q_1) \cap \Delta_r(Q_2) \neq \emptyset$, then $\Delta_{r-k}(Q_1) \supset \Delta_r(Q_2)$, where k is a positive integer depending only on the curvature bounds κ_1 and κ_2.

(H3) $0 < \omega(\Delta_r(Q)) \leq 1$ for every $Q \in S(\infty)$ and $r \in \mathbf{R}$.

(H4) For every $\Delta_r(Q)$ and $l > 0$, $\omega(\Delta_{r-l}(Q)) \leq C(l)\omega(\Delta_r(Q))$, where $C(l)$ is a positive constant depending only on M, o and l.

Without loss of generality, we may assume $k \geq 2$. Note that the function

$$\rho_0(Q, Q') := \inf\{e^{-t} : Q' \in \Delta_t(Q)\}, \quad Q, Q' \in S(\infty)$$

satisfies that

(D1) $\rho_0(Q, Q') = 0$ implies $Q = Q'$,

(D2) $\rho_0(Q, Q') \leq e^k \rho_0(Q', Q)$, where k is the constant in (H2),

(D3) $\rho_0(Q, Q'') \leq e^{2k}(\rho_0(Q, Q') + \rho(Q', Q''))$,

(D4) $\{Q' : \rho_0(Q, Q') < r\} = \Delta_{\log(1/r)}(Q)$ $(r > 0)$ and

$$\omega(\{Q' : \rho_0(Q, Q') < 2r\}) \leq C\omega(\{Q' : \rho_0(Q, Q') < r\}). \tag{14}$$

Consequently, the symmetrization

$$\rho(Q, Q') = \frac{\rho_0(Q, Q') + \rho_0(Q', Q)}{2}$$

is a quasi-distance in the sense of [15] such that $(S(\infty), \rho, \omega)$ is a space of homogeneous type, because there exists positive constants k_1 and k_2 depending only on M such that

$$\Delta_{\log(1/r)+k_1}(Q) \subset \{Q' : \rho(Q, Q') < r\} \subset \Delta_{\log(1/r)-k_2}(Q). \tag{15}$$

Therefore the abstract theory in [15] can be transplanted to our case. For instance, some covering lemmas, theorems on atomic Hardy spaces and BMO on spaces of homogeneous type hold true for $(S(\infty), \omega, \rho)$. We will sketch statements of some of them. As first, we deal with a covering lemma of Vitali type. Since the family of balls defined by the quasi-distance ρ and the family of sets $\{\Delta_t(Q)\}$ are equivalent (see (15)), we can state Vitali's covering lemma in terms of $\{\Delta_t(Q)\}$:

Lemma V (Vitali type covering lemma: see [3], [15]). *Let $E \subset S(\infty)$. Suppose $\{\Delta_{r(Q)}(Q) : Q \in E\}$ is a covering of E. Then there exist $Q_1, Q_2, \ldots$ in E such that*
(16) $\Delta_{r(Q_i)}(Q_i) \cap \Delta_{r(Q_j)}(Q_j) = \emptyset$, *$i \neq j$, and*
(17) *for every $Q \in S(\infty)$, there exists i with $\Delta_{r(Q)}(Q) \subset \Delta_{r(Q_i)-k'}(Q_i)$, where k' is a positive constant depending only on M and o.*

As known, from this lemma it follows the Hardy-Littlewood maximal theorem. To mention the theorem, we need the uncentered Hardy-Littlewood maximal function of $f \in L^1(\omega)$ defined as

$$(18)\quad \mathfrak{M}(f)(Q) := \sup_{\Delta_t(Q') : Q \in \Delta_t(Q')} \frac{1}{\omega(\Delta_t(Q'))} \int_{\Delta_t(Q')} |f|\, d\omega, \quad Q \in S(\infty).$$

Then we have

Theorem HL (see [15]). (1) *There exists a positive constant C_9 such that*

$$\omega(\{Q \in S(\infty) : \mathfrak{M}(f)(Q) > \lambda\}) \leq C_9 \lambda^{-1} \|f\|_{L^1(\omega)}$$

for all $f \in L^1(\omega)$ and for all $\lambda > 0$.

(2) *For every $1 < p \leq \infty$, there exists a positive constant $C_{10,p}$ such that*

$$\|\mathfrak{M}(f)\|_{L^p(\omega)} \leq C_{10,p} \|f\|_{L^p(\omega)}$$

for every $f \in L^p(\omega)$.

Now let us mention the definition of atomic Hardy spaces on $S(\infty)$. In [15], atomic Hardy spaces and BMO on a space of homogeneous type are defined in terms of its quasi-distance. However in our case, as we have seen, the family of balls defined by ρ is equivalent to $\{\Delta_t(Q)\}$. For this reason, one can define atomic Hardy spaces and BMO in terms of $\{\Delta_t(Q)\}$ which are equivalent to those defined by the quasi-distace ρ: Let $0 < p < q$ and $p \leq 1 \leq q \leq \infty$. A function a on $S(\infty)$ is called (p, q)-atom if the support of a is contained in a "ball" $\Delta_r(Q)$, $\int_{S(\infty)} a\, d\omega = 0$, and $\|a\|_{L^q(\omega)} \leq \omega(\Delta_r(Q))^{1/q-1/p}$. Since $\omega(S(\infty)) = 1$, we regard also the constant function 1 as a (p, q)-atom.

For a continuous function f on $S(\infty)$, let

$$|f|_\alpha = \sup\left\{\frac{|f(Q)-f(Q')|}{\omega(\Delta_r(Q''))^\alpha} : r \in \mathbf{R},\ Q'' \in S(\infty) \text{ with } Q, Q' \in \Delta_r(Q'')\right\}.$$

Let Λ^α be the set of all continuous functions f with $|f|_\alpha < \infty$. The atomic Hardy spaces $H^{p,q}(\omega)$ $(= H^{p,q}(S(\infty),\omega))$ are defined as follows:

(i) If $0 < p < 1 \le q \le \infty$, then $H^{p,q}(\omega)$ is the subspace of the dual of $\Lambda^{1/p-1}$ consisting of those linear functionals admitting an atomic decomposition

$$h = \sum_{j=1}^{\infty} \lambda_j a_j, \tag{19}$$

where $\lambda_j \in \mathbf{R}$, and a_j's are (p,q)-atoms and $\sum_{j=1}^{\infty} |\lambda_j|^p < \infty$.

(ii) If $p = 1 < q \le \infty$, then $H^{1,q}(\omega)$ is defined as the set of all functions h in $L^1(S(\infty),\omega)$ such that h has an atomic decomposition (19), where a_j's are $(1,q)$-atom and $\sum_{j=1}^{\infty} |\lambda_j| < \infty$.

In any case we set

$$|h|_{p,q}^p = \inf\left\{\sum_{j=1}^{\infty} |\lambda_j|^p : h = \sum_{j=1}^{\infty} \lambda_j a_j,\ a_j\text{'s are } (p,q)\text{-atoms}\right\}$$

for $h \in H^{p,q}(\omega)$. Then the function $\phi(h,f) = |h-f|_{p,q}$ is defines a complete metric on $H^{p,q}(\omega)$.

By Coifman and Weiss [15], we obtain that $H^{p,q}(\omega)$ is isomorphic to $H^{p,\infty}(\omega)$, for $1 < q < \infty$.

Let $\mathrm{BMO}(\omega)$ be the set of all functions $f \in L^1(S(\infty),\omega)$ such that

$$|f|_{\mathrm{BMO}(\omega)} = \sup_{Q \in S(\infty),\, r \in \mathbf{R}} \frac{1}{\omega(\Delta_r(Q))} \int_{\Delta_r(Q)} |f - m_{\Delta_r(Q)} f| \, d\omega + \|f\|_{L^1(\omega)} < \infty,$$

where

$$m_{\Delta_r(Q)} f = \frac{1}{\omega(\Delta_r(Q))} \int_{\Delta_r(Q)} f \, d\omega.$$

Moreover, since the definitions of $H^{p,q}(\omega)$ and $\mathrm{BMO}(\omega)$ are equivalent to those by [15], we have the following:

Theorem CW ([15]). (1) *$H^{p,q}(\omega) = H^{p,\infty}(\omega)$, and moreover the (quasi-) norms $|\cdot|_{p,q}$ and $|\cdot|_{p,\infty}$ are equivalent* ($0 < p \leq 1$, $1 < q < \infty$).

(2) *If $p < 1$, $\alpha = 1/p - 1$ and $1 < q \leq \infty$, then Λ^α is isomorphic to the dual of $H^{p,q}(\omega)$.*

(3) *For every $1 < q \leq \infty$, the dual of $H^{1,q}(\omega)$ is regarded as the space* BMO(ω) *in the following sense: If $h = \sum \lambda_j a_j \in H^{1,q}(\omega)$, then for each $\ell \in$* BMO(ω)

$$\langle h, \ell \rangle := \lim_{m \to \infty} \lambda_j \int_X \ell a_j \, d\omega$$

is a well defined continuous linear functional and its norm is equivalent to $|\ell|_{\mathrm{BMO}}$. Moreover, every linear continuous functional on $H^{1,q}(\omega)$ has this form.

In this paper we write

$$H^1_{\mathrm{atom}}(\omega) = H^{1,\infty}(\omega), \quad \text{and} \quad \|\cdot\|_{1,\mathrm{atom}} = |\cdot|_{1,\infty}.$$

§5. Some basic properties of Hardy spaces

This section is concerned with an elementary properties of Hardy spaces. As first we prove the equivalence of $H^p(\omega, \alpha)$ ($\alpha \in \mathbf{R}$). From now on, for $z \in M \setminus \{o\}$, we denote $z_\infty = \gamma_{oz}(+\infty)$ and $t(z) = t_{oz}$, where t_{oz} is a unique positive number such that $\gamma_{oz}(t_{oz}) = z$.

Proposition 5.1. *Suppose $-\infty < \alpha < \beta < \infty$. For every $1 \leq p \leq \infty$, $H^p(\omega, \alpha) = H^p(\omega, \beta)$. Moreover, the norms $\|f\|_{H^p(\omega,\beta)}$ and $\|f\|_{H^p(\omega,\alpha)}$ are equivalent.*

Proof. This proposition is a direct consequence of the following lemma:

Lemma 5.2. *For every continuous function u on M,*

$$\begin{aligned} &\omega(\{Q \in S(\infty) : N_\beta(u)(Q) > \lambda\}) \\ &\quad \leq \omega(\{Q \in S(\infty) : N_\alpha(u)(Q) > \lambda\}) \\ &\quad \leq C_{\alpha,\beta}\omega(\{Q \in S(\infty) : N_\beta(u)(Q) > \lambda\}), \end{aligned}$$

for all $\lambda > 0$, where $C_{\alpha,\beta}$ is a positive constant depending only on α, β, o and M.

Proof of Lemma 5.2. We adapt a standard argument (cf. [10]) to our case. Since by definition, $\Gamma_\beta(Q) \subset \Gamma_\alpha(Q)$, it is sufficient to prove the second inequality. Let f be the characteristic function of the set $\{N_\beta(u) > \lambda\}$. For every $Q \in \{N_\alpha(u) > \lambda\}$, there is a point $z \in M$ such that $|u(z)| > \lambda$ and $Q \in \Delta(z,\alpha)$. Then since $Q \in \Delta(z,\alpha) = \Delta_{t(z)+\alpha}(z_\infty)$, we have

$$\mathfrak{M}(f)(Q) \geq \frac{1}{\omega(\Delta(z,\alpha))} \int_{\Delta(z,\alpha)} |f|\, d\omega. \tag{20}$$

On the other hand, (H2) implies that

$$\begin{aligned}\Delta(z,\beta) \subset \Delta(z,\alpha) &= \Delta_{t(z)+\alpha}(z_\infty) \subset \Delta_{t(z)+\alpha-k}(Q)\\ &\subset \Delta_{t(z)+\alpha-2k}(z_\infty) = \Delta_{t(z)+\beta-(\beta+2k-\alpha)}(z_\infty).\end{aligned}$$

Hence by (20),

$$\begin{aligned}\mathfrak{M}(f)(Q) &\geq \frac{1}{\omega(\Delta_{t(z)+\beta-(\beta+2k-\alpha)}(z_\infty))} \int_{\Delta(z,\beta)} |f|\, d\omega\\ &\geq C(\beta+2k-\alpha)^{-1} \frac{1}{\omega(\Delta_{t(z)+\beta}(z_\infty))} \int_{\Delta(z,\beta)} |f|\, d\omega \quad (:= (\mathrm{I}), \text{ say}).\end{aligned}$$

Further, since for $Q' \in \Delta(z,\beta)$, $N_\beta(u)(Q') > \lambda$, so we have that $|f(Q')| \geq 1$, and therefore $(\mathrm{I}) \geq C(\beta+2k-\alpha)^{-1} > 0$. Consequently, $\{N_\alpha(u) > \lambda\} \subset \{\mathfrak{M}(f) > C(\beta+2k-\alpha)^{-1}\}$. Accordingly by Theorem HL (1) we obtain that

$$\begin{aligned}\omega(\{N_\alpha(u) > \lambda\}) &\leq C_9 C(\beta+2k-\alpha)\|f\|_{L^1(\omega)}\\ &= C_9 C(\beta+2k-\alpha)\omega(\{N_\beta(u) > \lambda\}).\end{aligned}$$

Q.E.D.

Next we prove some estimates for the Poisson kernel which will be used in this paper:

Lemma 5.3. (i) *Suppose $r > 0$. Then there exists a positive constant $C_{0,r}$ such that*

$$C_{0,r}^{-1} \leq K(x,Q) \leq C_{0,r}, \quad \text{for all } x \in B(o,r) \text{ and } Q \in S(\infty).$$

(ii) *There exist positive constants C_{11}, C_{12} and C_{13} satisfying the following* (a) *and* (b):

(a) *For every $x \in M \setminus B(o,3)$ and for every positive integer j with $d(o,x) > j+1$,*

$$\sup\{K(x,Q) : Q \in \Delta(x,-j-1) \setminus \Delta(x,-j)\} \leq C_{11} \frac{\exp(-jC_4)}{\omega(\Delta(x,-j-1))}.$$

Let N be the largest positive integer N with $d(o,x) > N+1$. Then

$$\sup\{K(x,Q) : Q \in S(\infty) \setminus \Delta(x,-N-1)\} \leq C_{12} \exp(-C_4 d(o,x)).$$

(b) *For every $x \in M \setminus \{o\}$,*

$$\sup\{K(x,Q) : Q \in \Delta(x,0)\} \leq C_{13} \frac{1}{\omega(\Delta(x,0))}.$$

(iii) *Suppose $r > 0$ and $Q \in S(\infty)$. There exists a positive constants C_{14} depending only on M such that for all $t > r$ and $Q' \in S(\infty) \setminus \Delta_r(Q)$,*

$$K(\gamma_{oQ}(t), Q') \leq C_{14} C_{0,r} \exp(-C_4(t-r)).$$

Proof. (i) This is proved easily by Harnack inequality: Let $x \in B(o,r)$. Then Theorem H implies that for all $y \in M \setminus B(o,r+1)$, $G(x,y) \leq C_r G(o,y)$, where C_r is a positive constant depending only on M and r. Hence the construction of the Martin kernel (see [3]),

$$K(x,Q) \leq \sup_{y \in M \setminus B(o,r+1)} \frac{G(x,y)}{G(o,y)} \leq C_r,$$

and $1 = K(o,Q) \lessapprox K(x,Q)$. Consequently we have (i).

(ii) Suppose $x \in M \setminus B(o,3)$, and let j be a positive integer such that $d(o,x) > j+1$. For simplicity, let $x(-j) = \gamma_{ox}(t(x)-j)$. Let F be an arbitrary Borel subset of $\Delta(x,-j-1) \setminus \Delta(x,-j)$. Then by Theorem BH2, we have that

$$\frac{\omega^x(F)}{G(x,o)} \approx \frac{\omega^{x(-j+1)}(F)}{G(x(-j+1),o)} \approx \frac{\omega^{x(-j)}(F)}{G(x(-j),o)}.$$

On the other hand, Theorems BH1 and BH2 imply that

$$\begin{aligned} \frac{G(o,x)}{G(o,x(-j))} &\approx \frac{G(x(-j-1),x)}{G(x(-j-1),x(-j))} \approx G(x(-j-1),x) \\ &\lessapprox \exp(-C_4 d(x(-j-1),x)). \end{aligned}$$

Combining these inequalities we have

$$\omega^x(F) \lessapprox \exp(-C_4 j)\omega^{x(-j-1)}(F) \approx \exp(-C_4 j)\omega^{x(-j-1)}(F).$$

Since there exists a positive constant c such that for every $z \in M \setminus \{o\}$,

(21) $\qquad \omega^z(\Delta(z,0)) \geq c, \quad$ (see the proof of [3, Lemma 7.4]),

we have by Theorem BH2 (2) that

$$\omega^{x(-j-1)}(F) \lessapprox \frac{\omega^{x(-j-1)}(F)}{\omega^{x(-j-1)}(\Delta(x,-j-1))} \approx \frac{\omega(F)}{\omega(\Delta(x,-j-1))}.$$

Hence

$$\frac{\omega^x(F)}{\omega(F)} \lessapprox \exp(-C_4 j)\omega(\Delta(-j-1))^{-1}.$$

Therefore for $Q \in \Delta(x,-j-1) \setminus \Delta(x,-j)$, we have the first inequality in (a):

$$K(x,Q) = \lim_{F \to \{Q\}} \frac{\omega^x(F)}{\omega(F)} \le C \exp(-C_4 j)\omega(\Delta(-j-1))^{-1}.$$

The second inequality in (a) is a direct consequence of Theorems BH2 and BH1: From (i) and Theorem BH2 it follows that for $Q \in S(\infty) \setminus \Delta(x,-N-1)$,

$$\frac{K(x,Q)}{G(x,o)} \approx \frac{K(x(-N),Q)}{G(x(-N),o)} \approx 1,$$

and consequently $K(x,Q) \approx G(x,o) \lessapprox \exp(-C_4 d(o,x))$.

To prove (b) we use (21). For $Q \in \Delta(x,0)$,

$$K(x,Q) \approx \frac{K(x,Q)}{\omega^x(\Delta(x,0))} \approx \frac{K(o,Q)}{\omega(\Delta(x,0))} = \frac{1}{\omega(\Delta(x,0))}.$$

This yields (b).

We prove (iii). Let $t > r+1$. Then by Theorem BH2 we have

$$\begin{aligned}\frac{K(\gamma_{oQ}(t),Q')}{G(\gamma_{oQ}(t),\gamma_{oQ}(r))} &\approx \frac{K(\gamma_{oQ}(r+1),Q')}{G(\gamma_{oQ}(r+1),\gamma_{oQ}(r))} \approx K(\gamma_{oQ}(r+1),Q') \\ &\approx K(\gamma_{oQ}(r),Q') \le C_{0,r}.\end{aligned}$$

Hence $K(\gamma_{oQ}(t),Q') \le C_{0,r} G(\gamma_{oQ}(t),\gamma_{oQ}(r))$, and this implies (iii).

Q.E.D.

Using Lemma 5.3, we have

Lemma 5.4. *There exists a positive constant C_{15} such that for every $f \in L^1(\omega)$,*

$$N_0(\tilde{f})(Q) \le C_{15}\mathfrak{M}(f)(Q), \quad Q \in S(\infty).$$

We can prove this lemma by combining [3, Theorem 7.3] with Theorem CK1. However here we give a direct proof using Lemma 5.3:

Proof. Let x be an arbitrary point in $\Gamma_0(Q) \setminus B(o,3)$. Then $Q \in \Delta(x,0)$. Let N be the largest positive integer with $d(o,x) > N+1$.

$$\begin{aligned} |\tilde{f}(x)| &\leq \int_{S(\infty)\setminus\Delta(x,-N-1)} + \sum_{j=0}^{N} \int_{\Delta(x,-j-1)\setminus\Delta(x,-j)} \\ &\quad + \int_{\Delta(x,0)} K(x,Q)|f(Q)|\, d\omega(Q) \\ &(= \text{(I)} + \text{(II)} + \text{(III)}, \quad \text{say}). \end{aligned}$$

Then by Lemma 5.3,

$$\text{(I)} \leq C_{12} \int_{S(\infty)} |f|\, d\omega \leq C_{12}\mathfrak{M}(f)(Q), \quad \text{and}$$

$$\begin{aligned} \text{(II)} &\leq C_{11} \sum_{j=0}^{N} \frac{\exp(-jC_4)}{\omega(\Delta(x,-j-1))} \int_{\Delta(x,-j-1)} |f|\, d\omega \\ &\leq C_{11} \sum_{j=0}^{N} \exp(-jC_4)\mathfrak{M}(f)(Q) \lessapprox \mathfrak{M}(f)(Q). \end{aligned}$$

Further,

$$\text{(III)} \leq \frac{C_{13}}{\omega(\Delta(x,0))} \int_{\Delta(x,0)} |f|\, d\omega \leq C_{13}\mathfrak{M}(f)(Q).$$

Consequently, we have $|\tilde{f}(x)| \lessapprox \mathfrak{M}(f)(Q)$.

Now we consider the case of $x \in B(o,3)$. By Lemma 5.3 (i),

$$\begin{aligned} |\tilde{f}(x)| &\leq \int_{S(\infty)} K(x,\zeta)|f(\zeta)|\, d\omega(\zeta) \\ &\leq C_{0,3} \int_{S(\infty)} |f(\zeta)|\, d\omega(\zeta) \leq C_{0,3}\mathfrak{M}(f)(Q). \end{aligned}$$

Thus we obtain the desired inequality. Q.E.D.

Now we have the following theorem as known in the classical case.

Theorem 5.5. (i) *For* $1 \leq p \leq \infty$, $\|f\|_{L^p(\omega)} \leq \|f\|_{H^p(\omega)}$, *for all* $f \in H^1(\omega)$.

(ii) *Suppose* $1 < p \leq \infty$. *Then* $H^p(\omega) = L^p(\omega)$, *and there exists a positive constant* $C_{16}(p)$ *such that* $\|f\|_{H^p(\omega)} \leq C_{16}(p)\|f\|_{L^p(\omega)}$ *for all* $f \in H^p(\omega)$.

Proof. Note that for a given function $f \in L^1(\omega)$ and for a Lebesgue point Q of f,

$$\lim_{t\to\infty} \tilde{f}(\gamma_{oQ}(t)) = f(Q).$$

Indeed by Lemma 5.4 and Lemma 5.3 (iii), we can prove this assertion by a similar way as the classical case (see [44, p.244]). Accordingly (i) is obvious.

Theorem HL and Lemma 5.4 guarantee (ii). Q.E.D.

We close this section with making some remarks on truncated maximal functions. For $r > 0$, let

$$N_{0,r}(f)(Q) := \sup\{|\tilde{f}(z)| : z \in \Gamma_0(Q) \cap (M \setminus B(o,r))\}.$$

By the same way as in the case of $N_0(f)$, we have that the function $N_{0,r}(f)$ is lower semicontinuous on $S(\infty)$. Using Lemma 5.3 (ii) we obtain that

$$N_{0,r}(f)(Q) \leq N_0(f)(Q) \leq C_r\|f\|_{L^1(\omega)} + N_{0,r}(f)(Q), \quad Q \in S(\infty). \tag{22}$$

Therefore for every $r > 0$,

$$\|f\|_{H^p} \approx \|N_{0,r}(f)\|_{L^p(\omega)}. \tag{23}$$

§6. Hardy spaces, atoms and Brownian motion

In this section we prove the equivalence of the spaces $H^1(\omega)$, H^1_{atom} and probabilistic analogues of Hardy spaces which will be mentioned later. To describe the analogues we recall some facts and notions in probability theory:

Let W be the set of all continuous maps from $[0, \infty)$ to M, and let $Z_t(w) = w(t)$, $w \in W$. Since by Yau [50] the life time of Brownian motion on M is equal to $+\infty$, so there exists a system of probability measures $\{P_x\}_{x\in M}$ on W such that (P_x, Z_t) is a Brownian motion starting at x. From Sullivan [46] or Kifer [27] it follows the following (A) and (B):

(A) There exists a limit $Z_\infty(w) := \lim_{t\to\infty} Z_t(w)$ for almost sure $w \in W$ with respect to P_x, $x \in M$. Moreover, $Z_\infty(w) \in S(\infty)$ for P_x-a.s. $w \in W$.

(B) For every $x \in M$ and for every Borel subset F of $S(\infty)$,

$$\omega^x(F) = P_x(\{w \in W : Z_\infty(w) \in F\}).$$

Since for every $f \in L^1(\omega)$, we have $f \in L^1(\omega^x)$ for all $x \in M$ by (9). Therefore for every $f \in L^1(\omega)$, $\tilde{f}(x) = E_x[f(Z_\infty)]$ for all $x \in M$ and $\lim_{t\to\infty} \tilde{f}(Z_t) = f(Z_\infty)$ P_x-a.s., where $E_x[\ ]$ denotes the expectation with respect to P_x $(x \in M)$.

We denote $P = P_o$ and $E[\ ] = E_o[\ ]$. First we describe a probabilistic analogue of Hardy spaces:

$$H^p_{\text{prob}} := \left\{ f \in L^p(\omega) : \|f\|_{H^p_{\text{prob}}} = E\left[\sup_{0\le t<\infty} \left|\tilde{f}(Z_t)\right|^p\right]^{1/p} < \infty \right\},$$
$$1 \le p < \infty.$$

Next we will deal with another probabilistic analogue of Hardy spaces. To define it, we recall some facts on Markov properties of $\{P_x\}_{x\in M}$: Let $\mathcal{B}$ (resp. $\mathcal{B}_t$) be the smallest σ-field for which all random variables Z_s, $s \ge 0$ (resp. Z_s, $0 \le s \le t$) are measurable. For a probability Borel measure μ on M, let $P_\mu(A) = \int_{S(\infty)} P_x(A)\, d\mu(x)$, $A \subset W$. We denote by $(W, \mathcal{F}^\mu, \mathcal{F}^\mu_t, P_\mu)$ the usual P_μ augmentation of $(W, \mathcal{B}, \mathcal{B}_t, P_\mu)$ in the sense of [43, III 9]. In particular, $(W, \mathcal{F}^x, \mathcal{F}^x_t, P_x)$ denotes the P_x-augmentation of $(W, \mathcal{B}, \mathcal{B}_t, P_\mu)$. Put $\tilde{\mathcal{F}} := \bigcap \mathcal{F}^\mu$ and $\tilde{F}_t := \bigcap \mathcal{F}^\mu_t$, where the intersection is taken over all probability Borel measures μ on M. Then $(Z_t, W, \tilde{\mathcal{F}}, \tilde{\mathcal{F}}_t, P_x : x \in M)$ is a strong Markov process. If fact, considering that M is diffeomorphic to $\mathbf{R}^n$, it is a honest FD diffusion in the sense of [43, III 3, III 13].

It is known that the usual P_x-augmentation $(W, \mathcal{F}^x, \mathcal{F}^x_t, P_x)$ satisfies the so-called usual condition (see [43, III 9]). Moreover, for every harmonic function u on M, the process $u(Z_t)$ is a continuous local $(P_x, \mathcal{F}^x_t)$-martingale. Denote by $(W, \mathcal{F}, \mathcal{F}_t, P)$ the usual P_o-augmentation $(W, \mathcal{F}^o, \mathcal{F}^o_t, P_o)$. As usual, Hardy spaces of martingales are defined as follows:

$$\mathcal{M}^p := \left\{ X \in L^1(W, \mathcal{F}, P) : \|X\|_{\mathcal{M}^p} := E\left[\sup_{0\le t<\infty} |E[X|\mathcal{F}_t]|^p\right]^{1/p} < \infty \right\},$$

$(1 \le p < \infty)$, where and always $E[\,\cdot\,|\mathcal{C}]$ denotes the conditional expectation with respect to P and a sub σ-field $\mathcal{C}$ of $\mathcal{F}$. Note that Meyer's previsibility theorem ([43, VI 15, Theorem 15.4]) implies that for every $X \in L^1(W, P)$, the process $(E[X/\mathcal{F}_t])_{t\ge0}$ is an $(\mathcal{F}_t)$-continuous martingale.

For $X \in L^1(W, \mathcal{F}, P)$, let $\mathcal{N}'(X) := E[X|\sigma(Z_\infty)]$, where $\sigma(Z_\infty)$ is the sub σ-field of $\mathcal{F}$ generated by the random variable Z_∞. Then by (A) there exists a unique element $f \in L^1(\omega)$ such that $\mathcal{N}'(X) = f(Z_\infty)$, P-a.s. Denote the function f by $\mathcal{N}X$. Now we can mention the second probabilistic analogue of Hardy spaces:

$$H^p_{\mathrm{mart}} := \{\mathcal{N}(X) : X \in \mathcal{M}^p\}, \quad 1 \leq p < \infty,$$

and $\|\mathcal{N}(X)\|_{H^p_{\mathrm{mart}}} := \inf\{\|Y\|_{\mathcal{M}^p} : \mathcal{N}(Y) = \mathcal{N}(X),\, Y \in \mathcal{M}^p\}$.

To describe our results, we use the following notation: For two normed spaces $(A, \|\ \|_A)$ and $(B, \|\ \|_B)$, we denote by $A \preceq B$ that $A \subset B$ and $\|x\|_B \leq C\|x\|_A$ for every $x \in A$, where C is a constant independent of x. Further we set $A \simeq B$ if $A \preceq B$ and $B \preceq A$.

Theorem 6.1.

$$H^1(\omega) \simeq H^1_{\mathrm{atom}}(\omega) \simeq H^1_{\mathrm{prob}} \simeq H^1_{\mathrm{mart}}.$$

Before proving this theorem, we would like to refer to both a work of Cifuentes and Korányi ([14]) and our previous announcement [5]. As pointed out briefly in Introduction, we announced in [5] the following two theorems:

Theorem 6.2.

$$H^1(\omega) \preceq H^1_{\mathrm{prob}} \preceq H^1_{\mathrm{mart}} \preceq H^1_{\mathrm{atom}}(\omega).$$

Theorem 6.3. *Consider the following geometric condition:*
(β) *For every $Q \in S(\infty)$, $t > 0$ and $z \in C(\gamma_{oQ}(t), 0)$,*

$$\Delta_t(\gamma_{oz}(+\infty)) \cap \Delta_t(Q) \neq \emptyset.$$

If our manifold M satisfies the condition (β), we have $H^1_{\mathrm{atom}}(\omega) \preceq H^1(\omega)$.

It is easy to see that when M is rotationally symmetric at o or the dimension of M is two, the condition (β) is satisfied. However recently, Cifuentes and Korányi proved the following

Theorem CK2 ([14]). *The manifold M satisfies always the condition (β).*

Therefore combining our Theorems 6.2 and 6.3 with Theorem CK2, we gain finally Theorem 6.1. For this reason, in order to get Theorem 6.1, we prove in this section Theorems 6.2 and 6.3.

First we prove the following

Proposition 6.4. *For every continuous function u on M, and for every $\lambda > 0$,*

$$P\left(\left\{w \in W : \sup_{0 \le t < \infty} |u(Z_t)| > \lambda\right\}\right) \lessapprox \omega(N_0(u) > \lambda).$$

In particular, we have $H^1(\omega) \preceq H^1_{\mathrm{prob}}$.

Proof. We adapt to our case of an idea of Burkholder, Gundy and Silverstein [10]. Let $F = \{N_0(u) > \lambda\}$. By the definition of admissible regions, when $|u(z)| > \lambda$, then $\Delta(z, 0) \subset F$. Hence by (21) $\omega^z(F) \ge \omega^z(\Delta(z, 0)) \ge c > 0$. Denote by χ_F the characteristic function of F. From Doob's maximal theorem it follows that

$$\begin{aligned} P(\{\sup_t |u(Z_t)| > \lambda\}) &\le P(\{\sup_t \omega^{Z_t}(F) > c\}) = P(\{\sup_t E[\chi_F/\mathcal{F}_t] > c\}) \\ &\le C\|\sup_t E[\chi_F/\mathcal{F}_t]\|^2_{L^2(W,P)} \le CE[\chi_F(Z_\infty)] = C\omega(F). \end{aligned}$$

Q.E.D.

For $f \in H^1_{\mathrm{prob}}$, we have that $\mathcal{N}f(Z_\infty) = f$ and $E[f(Z_\infty)/\mathcal{F}_t] = \tilde{f}(Z_t)$. Accordingly $H^1_{\mathrm{prob}} \preceq H^1_{\mathrm{mart}}$.

Next we prove $H^1_{\mathrm{mart}} \preceq H^1_{\mathrm{atom}}$. For this aim, we need to recall a probabilistic version of BMO: For $f \in L^1(\omega)$, let

$$\|f\|_{\mathrm{BMO_{prob}}} := \sup_{0 \le t < \infty} \left\| E\left[|\tilde{f}(Z_\infty) - \tilde{f}(Z_t)| / \mathcal{F}_t \right] \right\|_{L^\infty(W,P)} + \|f\|_{L^1(\omega)},$$

and let $\mathrm{BMO_{prob}} := \{f \in L^1(\omega) : \|f\|_{\mathrm{BMO_{prob}}} < \infty\}$.

As in the classical case, one can consider the following version of BMO norm called "Garsia norm":

$$\|f\|_G := \sup_{x \in M} \int_M |f(Q) - \tilde{f}(x)|\, d\omega^x(Q) + \|f\|_{L^1(\omega)} \, (\le \infty),$$

for $f \in L^1(\omega)$.

Before proving $H^1_{\mathrm{mart}} \preceq H^1_{\mathrm{atom}}(\omega)$, we show the following relation among these variants of BMO norms by using ideas in [48]:

Proposition 6.5. *Let $f \in L^1(\omega)$. Then*

$$\|f\|_{\mathrm{BMO_{prob}}} \lessapprox \|f\|_G \lessapprox \|f\|_{\mathrm{BMO}}. \tag{24}$$

Proof. Let $F_s(w) := |\tilde{f}(w(s)) - \tilde{f}(w(0))|$, $w \in W$, and let θ_t be the shift operator, i.e., $\theta_t(w)(s) := w(s+t)$. Then the Markov property of Brownian motion on M, we have that $E[F_s \circ \theta_t / \mathcal{F}_t] = E_{X_t}[F_s]$. Hence $E[|f(X_{t+s}) - f(X_t)| / \mathcal{F}_t] = E_{X_t}[|f(X_s) - f(X_0)|]$ P-a.s. Letting $s \to \infty$, we have

$$\begin{aligned} E[|f(X_\infty) - f(X_t)| / \mathcal{F}_t] &= E_{X_t}[|f(X_\infty) - f(X_0)|] \\ &= \int_M |f(Q) - f(X_t)| \, d\omega^{X_t}(Q), \end{aligned}$$

P-a.s. Consequently, we obtain the first inequality of (24).

For $x \in M \setminus B(o,3)$, we set $\Delta(x) = \Delta(x,0)$. Then

$$\begin{aligned} &\int_{S(\infty)} |f(Q) - \tilde{f}(x)| \, d\omega^x(Q) \\ &\leq \int_{S(\infty)} |f(Q) - m_{\Delta(x)} f| \, d\omega^x(Q) + \int_{S(\infty)} |m_{\Delta(x)} f - \tilde{f}(x)| \, d\omega^x \\ &= \int_{S(\infty)} |f(Q) - m_{\Delta(x)} f| \, d\omega^x(Q) + |\tilde{f}(x) - m_{\Delta(x)} f| \\ &\leq 2 \int_{S(\infty)} |f - m_{\Delta(x)} f| \, d\omega^x \\ &\leq 2 \int_{S(\infty) \setminus \Delta(x)} |f - m_{\Delta(x)} f| \, d\omega^x + 2 \int_{\Delta(x)} |f - m_{\Delta(x)} f| \, d\omega^x \\ &(:= (\mathrm{I}) + (\mathrm{II})) \end{aligned}$$

By Lemma 5.3 (ii) (b), we have

$$(\mathrm{II}) \leq C \frac{1}{\omega(\Delta(x))} \int_{\Delta(x)} |f - m_{\Delta(x)} f| \, d\omega \leq C \|f\|_{\mathrm{BMO}}.$$

To estimate (I), we use Lemma 5.3 (ii) (a): Let N be the largest positive integer with $d(o,x) > N+1$, and let $\Delta(j) = \Delta(x,-j)$. Then

$$\begin{aligned} (\mathrm{I}) &= \left(\int_{S(\infty) \setminus \Delta(N+1)} + \sum_{j=1}^{N} \int_{\Delta(j+1) \setminus \Delta(j)} \right) |f - m_{\Delta(x)}| \, d\omega^x \\ &=: (\mathrm{III}) + (\mathrm{IV}). \end{aligned}$$

From Lemma 5.3 it follows that

$$\begin{aligned}
(\mathrm{IV}) &\lessapprox \sum_{j=1}^{N} \frac{\exp(-jC_4)}{\omega(\Delta(j+1))} \int_{\Delta(j+1)} |f - m_{\Delta(x)} f| \, d\omega \\
&\lessapprox \sum_{j=1}^{N} \frac{\exp(-jC_4)}{\omega(\Delta(j+1))} \left\{ \int_{\Delta(j+1)} |f - m_{\Delta(j+1)} f| \, d\omega \right. \\
&\qquad + \int_{\Delta(j+1)} |m_{\Delta(j+1)} f - m_{\Delta(j)} f| \, d\omega \\
&\qquad \left. + \cdots + \int_{\Delta(j+1)} |m_{\Delta(1)} f - m_{\Delta(0)} f| \, d\omega \right\}.
\end{aligned}$$

Since (H4) implies $|m_{\Delta(j+1)} f - m_{\Delta(j)} f| \lessapprox \|f\|_{\mathrm{BMO}}$, we have

$$(\mathrm{IV}) \lessapprox \sum_{j=1}^{N} \exp(-jC_4) \{ \|f\|_{\mathrm{BMO}} + j \|f\|_{\mathrm{BMO}} \} \lessapprox \|f\|_{\mathrm{BMO}}.$$

Moreover, by Lemma 5.3 (ii) (a),

$$\begin{aligned}
(\mathrm{III}) &\lessapprox \exp(-C_4 d(o,x)) \int_{S(\infty)\setminus\Delta(N+1)} |f - m_{\Delta(x)}| \, d\omega \\
&\leq 2 \exp(-C_4 d(o,x)) \|f\|_{L^1(\omega)}.
\end{aligned}$$

When $x \in B(o,3)$, applying Theorem H we have

$$\int_M |f - \tilde{f}| \, d\omega^x \lessapprox \int_M |f - \tilde{f}| \, d\omega \leq 2 \|f\|_{L^1(\omega)}.$$

Consequently we gain $\|f\|_G \lessapprox \|f\|_{\mathrm{BMO}}$. Q.E.D.

As a consequence of this proposition and of a probabilistic version of Fefferman's inequality ([24]), we have

Proposition 6.6. *For $X \in \mathcal{M}^1$,*

$$\|\mathcal{N} X\|_{1,\mathrm{atom}} \leq C \|X\|_{\mathcal{M}^1}.$$

Therefore $H^1_{\mathrm{mart}} \preceq H^1_{\mathrm{atom}}$.

Proof. Suppose $X \in L^2(W,P)$. Then from H^1_{atom}-BMO duality theorem in [15] it follows that

$$\begin{aligned}\|\mathcal{N}X\|_{1,\text{atom}} &= \sup\left\{|\Psi(\mathcal{N}X)| : \Psi \in (H^1_{\text{atom}})^*, \|\Psi\|_{(H^1_{\text{atom}})^*} \le 1\right\} \\ &\lessapprox \sup\left\{\left|\int_{S(\infty)} \psi\mathcal{N}X\, d\omega\right| : \psi \in \mathrm{BMO}(\omega),\ \|\psi\|_{\mathrm{BMO}} \le 1\right\} \\ &=: (\mathrm{I}).\end{aligned}$$

Now using a martingale version of Fefferman's inequality ([24]), we have that

$$\begin{aligned}(\mathrm{I}) &= \sup\{|E[\psi(Z_\infty)\mathcal{N}X(Z_\infty)]| : \psi \in \mathrm{BMO}(\omega), \|\psi\|_{\mathrm{BMO}} \le 1\} \\ &= \sup\{|E[\psi(Z_\infty)X]| : \psi \in \mathrm{BMO}(\omega), \|\psi\|_{\mathrm{BMO}} \le 1\} \\ &\lessapprox \|X\|_{\mathcal{M}^1} \sup\left\{\|\psi\|_{\mathrm{BMO}_{\text{prob}}} : \psi \in \mathrm{BMO}(\omega), \|\psi\|_{\mathrm{BMO}} \le 1\right\} \\ &\lessapprox \|X\|_{\mathcal{M}^1},\end{aligned}$$

where the last inequality is proved by Proposition 6.5.

Now suppose $X \in \mathcal{M}^1$. It is well known that for $X \in \mathcal{M}^1$, there exist $X^k \in \mathcal{M}^2$ $(k = 1, 2, \dots)$ such that $\|X^k - X\|_{\mathcal{M}^1} \to 0$ as $k \to \infty$. From what we have proved it follows that $\|\mathcal{N}X^k - \mathcal{N}X^l\|_{H^1_{\text{atom}}} \lessapprox \|X^k - X^l\|_{\mathcal{M}^1} \to 0$ $(k, l \to \infty)$. Since H^1_{atom} is complete ([15]), there exists $h \in H^1_{\text{atom}}$ such that $\|h - \mathcal{N}X^k\|_{1,\text{atom}} \to 0$ as $k \to \infty$. Therefore

$$\begin{aligned}\|\mathcal{N}X - h\|_{L^1(\omega)} &\le \|\mathcal{N}X - \mathcal{N}X^k\|_{L^1(\omega)} + \|\mathcal{N}X^k - h\|_{L^1(\omega)} \\ &\le \|X - X^k\|_{L^1(W,P)} + \|\mathcal{N}X^k - h\|_{1,\text{atom}} \\ &\le \|X - X^k\|_{\mathcal{M}^1} + \|\mathcal{N}X^k - h\|_{1,\text{atom}} \to 0, \quad (k \to \infty).\end{aligned}$$

Hence $\mathcal{N}X = h$ ω-a.e. Therefore

$$\begin{aligned}\|\mathcal{N}X\|_{1,\text{atom}} = \|h\|_{1,\text{atom}} &= \lim_{k\to\infty} \|\mathcal{N}X^k\|_{1,\text{atom}} \lessapprox \lim_{k\to\infty} \|X^k\|_{\mathcal{M}^1} \\ &= \|X\|_{\mathcal{M}^1}.\end{aligned}$$

Q.E.D.

When M is the open unit disc, this proposition for BMO was proved in [48]. See also [51] for balls.

What we have proved implies Theorem 6.2.

In order to prove Theorem 6.3, we need the following estimate:

Lemma 6.7. *Suppose $Q_0 \in S(\infty)$, $r > 3k$ and $Q \in \Delta_r(Q_0)$. Let N be the biggest positive integer such that $r > (N+2)k$. Let $C(j) = C(\gamma_{oQ_0}(r-jk), 0)$ $(j = 0, 1, \ldots, N)$. Then for every $j \in \{0, 1, \ldots, N\}$ and for every $x \in M \setminus C(j)$,*

$$|K(x,Q) - K(x,Q_0)| \lessapprox c^j K(x,Q_0),$$

where c is a positive constant such that $c < 1$ and is depending only on M.

Proof. The following proof is based on an idea in Anderson and Schoen [3, p.449]. Let $D(j) = M \setminus C(j)$, and let

$$\overline{\varphi}_j = \sup_{z \in D(j)} \frac{K(z,Q)}{K(z,Q_0)}, \quad \underline{\varphi}_j = \inf_{z \in D(j)} \frac{K(z,Q)}{K(z,Q_0)}.$$

Let $u_j(z) = K(z,Q) - \underline{\varphi}_{j-1} K(z,Q_0)$. Then u_j is harmonic on M and positive on $D(j-1)$. Hence by Theorem BH2 (2),

$$\sup_{z \in D(j)} \frac{u_j(z)}{K(z,Q_0)} \leq C_8 \inf_{z \in D(j)} \frac{u_j(z)}{K(z,Q_0)}. \tag{25}$$

(Note that $C_8 > 1$.) By (25), we have

$$\overline{\varphi}_j - \underline{\varphi}_{j-1} \leq C_8 \big(\underline{\varphi}_j - \underline{\varphi}_{j-1}\big). \tag{26}$$

On the other hand, if we consider the function $v_i(z) = \overline{\varphi}_{j-1} K(z,Q_0) - K(z,Q)$ in stead of u_j, we obtain the following estimate:

$$\overline{\varphi}_{j-1} - \underline{\varphi}_j \leq C_8 \big(\overline{\varphi}_{j-1} - \overline{\varphi}_j\big). \tag{27}$$

Let $\operatorname{osc}(j) = \overline{\varphi}_j - \underline{\varphi}_j$. Then by (26) and (27), we have

$$\operatorname{osc}(j) \leq \frac{C_8 - 1}{C_8 + 1} \operatorname{osc}(j-1).$$

Hence when $x \in M \setminus C(j)$, then

$$\begin{aligned} \left| \frac{K(x,Q)}{K(x,Q_0)} - 1 \right| &= \left| \frac{K(x,Q)}{K(x,Q_0)} - \frac{K(o,Q)}{K(o,Q_0)} \right| \leq \operatorname{osc}(j) \\ &\leq \left(\frac{C_8 - 1}{C_8 + 1} \right)^{j-1} \operatorname{osc}(1). \end{aligned}$$

Moreover, since
$$\frac{K(z,Q)}{K(z,Q_0)} \approx \frac{K(o,Q)}{K(o,Q_0)} = 1,$$
for all $z \in M \setminus C(j-1)$ by Theorem BH2 (2), we have that osc(1) is bounded by a positive constant depending only on M. Now what we have obtained implies the desired inequality. Q.E.D.

To prove the following lemma, we need the condition (β) stated in Theorem 6.3:

Lemma 6.8. *For every $Q \in S(\infty)$, $t > k$ and $Q' \in S(\infty) \setminus \Delta_{t-k}(Q)$, we have $\Gamma_0(Q') \cap C(\gamma_{oQ}(t), 0) = \emptyset$.*

Proof. Suppose that there exists a point $z \in \Gamma_0(Q') \cap C(\gamma_{oQ}(t), 0)$. Then $Q' \in \Delta(z, 0)$ and $d(o,z) > t$. Further by the condition (β), we have $\Delta_t(\gamma_{oz}(+\infty)) \cap \Delta_t(Q) \neq \emptyset$. Therefore from (H3) it follows that
$$\Delta_{t-k}(Q) \supset \Delta_t(\gamma_{oz}(+\infty)) \supset \Delta_{d(o,z)}(\gamma_{oz}(+\infty)) = \Delta(z,0) \ni Q'.$$
This contradicts to that $Q' \in S(\infty) \setminus \Delta_{t-k}(Q)$. Q.E.D.

Now we are ready to prove Theorem 6.3:

Proof of Theorem 6.3. It is sufficient to prove that for every $(1,\infty)$-atom a, $\|a\|_{H^1} \leq C$, where C is a positive constant depending only on M. Let a be a $(1,\infty)$-atom, that is, there exists a ball $\Delta_r(Q_0)$ such that
$$\operatorname{supp} a \subset \Delta_r(Q_0), \quad \int_M a\, d\omega = 0, \quad \|a\|_{L^\infty} \leq 1/\omega(\Delta_r(Q_0)).$$
Suppose $r > 3k$. For simplicity, let $\Delta = \Delta_r(Q_0)$, $\Delta(j) = C(\gamma_{oQ_0}(r - jk), 0) \cap S(\infty)\,(= \Delta_{r-jk}(Q_0))$, and $A(j) = \gamma_{oQ_0}(r-jk)$. Then
$$\|N_0(a)\|_{L^1(\omega)} = \int_{S(\infty)\setminus\Delta(N+2)} N_0(a)\, d\omega + \sum_{j=1}^{N} \int_{\Delta(j+2)\setminus\Delta(j+1)} N_0(a)\, d\omega$$
$$+ \int_{\Delta(2)} N_0(a)\, d\omega \;(:= \text{(I)} + \text{(II)} + \text{(III)}),$$
where N is the biggest integer with $r > (N+2)k$. First we estimate (II). Let $p \in \Delta(j+2) \setminus \Delta(j+1)$. Then by Lemma 6.8, $\Gamma_0(p) \cap C(\gamma_{oQ_0}(r - jk), 0) = \emptyset$. For $x \in M \setminus C(\gamma_{oQ_0}(r-jk), 0)$,
$$K(x,Q_0) \leq \frac{K(x,Q_0)}{\omega^x(\Delta(j))} \approx \frac{K(o,Q_0)}{\omega^o(\Delta(j))} = \frac{1}{\omega(\Delta(j))}.$$

Hence by Lemma 6.7, for every $x \in \Gamma_0(p)$,

$$\begin{aligned}
|\tilde{a}(x)| &= \left| \int_\Delta a(Q)\{K(x,Q) - K(x,Q_0)\}\, d\omega \right| \\
&\le \|a\|_{L^\infty} \int_\Delta |K(x,Q) - K(x,Q_0)|\, d\omega \\
&\lessapprox c^j K(x,Q_0) \lessapprox c^j \frac{1}{\omega(\Delta(j))}.
\end{aligned}$$

Using this estimate, we have that

$$\int_{\Delta(j+2)\setminus\Delta(j+1)} N_0(a)\, d\omega \lessapprox c^j \int_{\Delta(j+2)} \omega(\Delta(j))^{-1}\, d\omega \lessapprox c^j.$$

Accordingly (II) $\le C$. To estimate (I), let $p \in S(\infty) \setminus \Delta(N+2)$. Then $\Gamma_0(p) \cap C(\gamma_{oQ_0}(r - Nk), 0) = \emptyset$. Note that $\operatorname{dist}\big(\partial C(\gamma_{oQ_0}(r - Nk), 0), \partial C(\gamma_{oQ_0}(r - (N-1)k), 0)\big) \ge c$ for some positive constant depending only on the curvature bounds κ_1 and κ_2 (see [6, pp.310–311] for a related estimate). Therefore for every $x \in \Gamma_0(p)$ and $Q \in \Delta_r(Q_0)$,

$$\begin{aligned}
K(x,Q) &\le C \frac{K(x,Q)}{G(x, A(N-1))} \approx \frac{K(A(N),Q)}{G(A(N), A(N-1))} \\
&\approx K(A(N),Q) \le C,
\end{aligned}$$

where the last inequality is proved by Lemma 5.3 (ii). Hence $|\tilde{a}(x)| \le C$, and therefore $N_0(a)(p) \le C$. Consequently we have (I) $\le C$.

Next we estimate (III). For $x \in M$,

$$\begin{aligned}
|\tilde{a}(x)| &\le \int_{\Delta_r(Q_0)} K(x,Q)|a(Q)|\, d\omega(Q) \\
&\le \|a\|_{L^\infty} \int_{S(\infty)} d\omega^x \le \frac{1}{\omega(\Delta_r(Q_0))}.
\end{aligned} \tag{28}$$

This implies that (III) $\le C$.

Lastly we consider the case $r < 3k$. By Theorem H and (21),

$$\omega(\Delta_r(Q_0)) \ge \omega(\Delta_{3k}(Q_0)) \approx \omega^{A(3)}(\Delta_{3k}(Q_0)) \ge c$$

for some positive constant c depending only on M. Therefore (28) yields that $N_0(a) \le c^{-1}$, and that $\|a\|_{H^1(\omega)} \le c^{-1}$. Q.E.D.

Now we have proved Theorems 6.2 and 6.3. Therefore by combining these theorems with Theorem CK2, we gain Theorem 6.1.

As an immediate consequence of Theorem 6.1, we have the equivalence of BMO norm, Garsia norm and probabilistic BMO norm.

Corollary 6.9. *Let $f \in L^1(\omega)$. Then*

$$\|f\|_{\mathrm{BMO}} \approx \|f\|_G \approx \|f\|_{\mathrm{BMO}_{\mathrm{prob}}}.$$

Proof. By Proposition 6.5, it is sufficient to prove $\|f\|_{\mathrm{BMO}} \lessapprox \|f\|_{\mathrm{BMO}_{\mathrm{prob}}}$. Note that if $f \in \mathrm{BMO}_{\mathrm{prob}}$, then $f \in L^2(\omega)$. Moreover, $L^2(\omega)$ is dense in H^1_{atom}. Therefore by H^1_{atom}-BMO duality theorem (cf. [15]) and Theorem 6.1, we have that

$$\begin{aligned}
\|f\|_{\mathrm{BMO}} &\approx \sup\left\{\left|\int_{S(\infty)} fh\, d\omega\right| : h \in L^2(\omega),\ \|h\|_{1,\mathrm{atom}} = 1\right\} \\
&\approx \sup\left\{|E[f(Z_\infty)h(Z_\infty)]| : h \in L^2(\omega),\ \|h\|_{1,\mathrm{prob}} = 1\right\} \\
&\lessapprox \|f\|_{\mathrm{BMO}_{\mathrm{prob}}},
\end{aligned}$$

where for the last inequality, a probabilistic version of Fefferman's inequality is used. Q.E.D.

Remark 2. For $f \in L^1(\omega)$, let

$$\|f\|_{\mathrm{BMO},p} := \sup_{t\in\mathbf{R},\, Q\in S(\infty)} m_{\Delta_t(Q)}\left(|f - m_{\Delta_t(Q)}(f)|^p\right)^{1/p} + \|f\|_{L^p(\omega)},$$

$$\|f\|_{G,p} := \sup_{z\in M}\left(\int_M |f - \tilde{f}(z)|^p\, d\omega^z\right)^{1/p} + \|f\|_{L^p(\omega)},$$

$$\|f\|_{\mathrm{BMO}_{\mathrm{prob}},p} := \sup_{0\le t<\infty}\left\|E\left[|\tilde{f}(Z_\infty) - \tilde{f}(Z_t)|^p\right]^{1/p}\right\|_{L^\infty(W,P)} + \|f\|_{L^p(\omega)}.$$

It is known that if $f \in \mathrm{BMO}(\omega)$, then $f \in L^p(\omega)$ and $\|f\|_{\mathrm{BMO},p} \approx \|f\|_{\mathrm{BMO}}$ (cf. [15]). The same is true for probabilistic BMO, that is, if $f \in \mathrm{BMO}_{\mathrm{prob}}$, then $\tilde{f}(Z_\infty) \in L^p(W,P)$ and $\|f\|_{\mathrm{BMO}_{\mathrm{prob}},p} \approx \|f\|_{\mathrm{BMO}_{\mathrm{prob}}}$ (cf. [36]). Now by a similar way as the proof of Proposition 6.5 we can prove that $\|f\|_{\mathrm{BMO}_{\mathrm{prob}},p} \le C\|f\|_{G,p} \le C\|f\|_{\mathrm{BMO},p}$, for every $f \in L^p(\omega)$. Therefore what we have noted guarantees that

$$\|f\|_{G,p} \approx \|f\|_G. \tag{29}$$

§7. Carleson measures and Martin integrals

In this section we study a condition on a measure μ on M in order that the Martin integral operator,

$$K[f](z) = \int_{S(\infty)} K(z,Q)f(Q)\,d\omega(Q)\,(=\tilde{f}(z)), \quad z \in M,$$

is bounded from $L^p(\omega)$ to $L^p(M,\mu)$. This problem was studied by L. Carleson in the classical Euclidean case, and he found a necessary and sufficient condition called now "Carleson condition". We will study a version to M of "Carleson condition":

Definition 7.1. For a set $A \subset S(\infty)$ and $r > 0$, let

$$S_r[A] := \{z \in M \setminus B(o,r) : \Delta(z,0) \subset A\}.$$

A given complex Borel measure μ on M is said to be a Carleson measure on M if for every $r > 0$,

$$\|\mu\|_{c,r} := \sup_{Q\in S(\infty),\, t>1} \frac{|\mu|(S_r[\Delta_t(Q)])}{\omega(\Delta_t(Q))} + |\mu|(M) < \infty,$$

where $|\mu|$ is the total variation of μ. We wirte $\|\mu\|_c = \|\mu\|_{c,1}$.

As an analogue of the classical Carleson-Hörmander's theorem, we will prove the following by using Stein's idea:

Theorem 7.1. *Let μ be a complex Borel measure on M. Then the following are equivalent*:

(i) *μ is a Carleson measure on M.*

(ii) *$\|\mu\|_{c,r} < \infty$ for some $r > 0$.*

(iii) *For every $1 \le p < \infty$, the Martin integral operator K is bounded from $H^p(\omega)$ to $L^p(M,|\mu|)$.*

(iv) *For every $1 < p < \infty$, the operator K is bounded from $L^p(\omega)$ to $L^p(M,|\mu|)$.*

(v) *For some $1 < p < \infty$, the operator K is bounded from $L^p(\omega)$ to $L^p(M,|\mu|)$.*

Furthermore, for every $r > 0$, there is a constant C'_r depending only on M, o and r such that

$$C_r'^{-1}\|\mu\|_{c,r} \le \|\mu\|_c \le C'_r\|\mu\|_{c,r}.$$

Proof. We begin with proving "(ii) $\Rightarrow$ (iii)". We may assume that $r > k' + 1$, where k' is the positive constant in Lemma V. Suppose

$f \in H^p(\omega)$ and $\lambda > 0$. Let $E := \{Q \in S(\infty) : N_{0,r}(f) > \lambda\}$, and $G := \{z \in M \setminus B(o,r) : |\tilde{f}(z)| > \lambda\}$. Since $\{\Delta(z,0) : z \in S_r(E)\}$ is a covering of the bounded open set E, Vitali type covering lemma (Lemma V) guarantees that there exist $z_1, z_2, \ldots \in S_r(E)$ satisfying that $\Delta(z_i,0) \cap \Delta(z_j,0) = \emptyset$ $(i \neq j)$, and that for every $z \in S_r(E)$, $\Delta(z,0) \subset \Delta(z_i,-k')$ for some i.

Now let $z \in G$, and $Q \in \Delta(z,0)$. Then $z \in \Gamma_0(Q) \cap (M \setminus B(o,r))$, and therefore $N_{0,r}(f)(Q) > \lambda$. Hence $\Delta(z,0) \subset E$. Hence there exists i such that $\Delta(z,0) \subset \Delta(z_i,-k')$. This implies that $z \in S_r[\Delta(z,0)] \subset S_r[\Delta(z_i,-k')]$. Consequently, $G \subset \bigcup_j S_r[\Delta(z_j,-k')]$. Using this we have

$$\begin{aligned}|\mu|(G) &\leq \sum_j |\mu|(S_r[\Delta(z_j,-k')]) \leq \|\mu\|_{c,r} \sum_j \omega(\Delta(z_j,-k')) \\ &\leq C(k')\|\mu\|_{c,r} \sum_j \omega(\Delta(z_j,0)) \leq C(k')\|\mu\|_{c,r}\omega(E).\end{aligned}$$

Accordingly by these estimates, Lemma 5.3 (1) and (23) we have that

$$\begin{aligned}\int_M |\tilde{f}|^p \, d|\mu| &= \int_{B(o,r)} |\tilde{f}|^p \, d|\mu| + \int_{M\setminus B(o,r)} |\tilde{f}|^p \, d|\mu| \\ &\lessapprox \sup_{z\in B(o,r)} |\tilde{f}(z)|^p |\mu|(M) + \|\mu\|_{c,r} \int_{S(\infty)} N_{0,r}(f)^p \, d\omega \\ &\leq C_r |\mu|(M) \|f\|_{L^1(\omega)} + \|\mu\|_{c,r} \|N_{0,r}(f)\|_{L^p(\omega)} \\ &\leq (C_r + 1)\|\mu\|_{c,r}\|f\|_{H^p(\omega)}.\end{aligned}$$

The part "(iii) $\Rightarrow$ (iv)" and "(iv) $\Rightarrow$ (v)" are obvious. We prove "(v) $\Rightarrow$ (i)". Let f be the characteristic function of $\Delta_t(Q)$ $(t > 1,\ Q \in S(\infty))$. Suppose $r > 0$. If $z \in S_r[\Delta_t(Q)]$, then $\Delta(z,0) \subset \Delta_t(Q)$. Hence for every $z \in S_r[\Delta_t(Q)]$, $\tilde{f}(z) \geq \omega^z(\Delta(z,0)) > c/2$, where c is the positive constant in (21). Denote by C_p the operator norm of K from $L^p(\omega)$ to $L^p(|\mu|)$. Therefore,

$$\begin{aligned}|\mu|(S_r[\Delta_t(Q)]) &\leq |\mu|(\{z \in M : \tilde{f}(z) > c/2\}) \leq (c/2)^{-p} \int_M \tilde{f}^p \, d\mu \\ &\leq (c/2)^{-p} C_p^p \|f\|_{L^p(\omega)}^p = (c/2)^{-p} C_p^p \omega(\Delta_t(Q)).\end{aligned}$$

Moreover, $|\mu|(M) \leq C_p^p \|1\|_{L^p(\omega)} \leq C_p^p$. Thus we have $\|\mu\|_{c,r} \leq ((c/2)^{-p} + 1)C_p^p$. Q.E.D.

§8. Carleson measures and Green potentials

For a Borel measure μ on M, the function

$$G[\mu](x) = \int_M G(x,y)\, d\mu(y), \quad x \in M$$

is called the Green potential of μ. In this section we study boundary behavior of the Green potentials of the following weighted measures: for a nonnegative Borel measure μ on M, let

$$\mu_0(A) = \int_A \frac{1}{G(o,w)}\, d\mu(w), \quad A \subset M.$$

A nonnegative function f on M is said to be asymptotically bounded if there exists a positive constant $R > 0$ such that $\sup_{x \in M \setminus B(o,R)} f(x) < \infty$. Then we have the following

Theorem 8.1. *Let μ be a nonnegative Borel measure on M. Suppose that $\mu(H) < \infty$ for every compact set H in M. Then the following are equivalent*:

(i) $G[\mu_0]$ *is asymptotically bounded on M.*

(ii) μ *is a Carleson measure and satisfies the following condition* (F):

(F) *There exist positive constants r and C such that*

$$(30) \qquad \int_{B(z,1)} G(z,w)\, d\mu(w) \le CG(o,z) \quad \text{for every } z \in M \setminus B(o,r).$$

We denote by $C_{r,\mu}$ the infimum of constants C in the condition (F).

This theorem is used in order to prove a Carleson measure characterization of BMO stated in the next section.

Proof. First we prove "(ii) $\Rightarrow$ (i)". In order to prove this we need the following lemma:

Lemma 8.2. *For $Q \in S(\infty)$ and $t > 0$, let $C(Q,t) = C(\gamma_{oQ}(t), 0)$. Then for $0 < r' < t + k$,*

$$C(Q, t+k) \subset S_{r'}[\Delta_t(Q)], \quad (Q \in S(\infty)).$$

Proof of Lemma 8.2. Let $w \in C(Q, t+k)$. By the condition (β) in Theorem 6.3 (see also Theorem CK2), we have that $\Delta_{t+k}(w(+\infty)) \cap$

$\Delta_{t+k}(Q) \neq \emptyset$. Hence (H3) implies that $\Delta_{t+k}(w(+\infty)) \subset \Delta_t(Q)$. Since $\Delta(w,0) \subset \Delta_{t+k}(w(+\infty))$, we have $w \in S_{r'}[\Delta_t(Q)]$.

End of the proof of Lemma 8.2.

We proceed to prove Theorem 8.1 "(ii) $\Rightarrow$ (i)". By Lemma 8.2, we have

$$\frac{\mu(C(Q,t+k))}{\omega(\Delta_{t+k}(Q))} \lessapprox \frac{\mu(C(Q,t+k))}{\omega(\Delta_t(Q))} \leq \frac{\mu(S_{r'}[\Delta_t(Q)])}{\omega(\Delta_t(Q))} \lessapprox \|\mu\|_c, \tag{31}$$

where r' is a positive number with $r' < t+k$.

We may assume that the number r in the condition (F) is greater than $100k$. Now let $z \in M \setminus B(o, r+2)$. Denote $t(z) = d(o,z)$ and $z(t) = \gamma_{oz}(t(z)+t)$. Let N be the biggest positive integer with $t(z) > N+2$. Let

$$\begin{aligned} E_0 &= C(z(+\infty), t(z)) \quad \text{and} \\ E_j &= C(z(+\infty), t(z)-j) \setminus C(z(+\infty), t(z)-j+1), \end{aligned}$$

$(j = 1, 2, \ldots, N)$. Then $\mu(E_j) \lessapprox \|\mu\|_c \omega(\Delta_{t(z)-j}(z(+\infty)))$.

We estimate each integral $\int_{E_j} G(z,w)\, d\mu(w)$. Let $w \in E_j$. Note that when $w \in M \setminus B(z(-j), 3)$, then $G(z(-j+2), w) \approx G(z(-j-1), w)$. Suppose $N \geq j \geq 4$. By [1, Theorem 1] we have that

$$\begin{aligned} G(w,z) &\lessapprox G(z(-j+1), z) G(w, z(-j+2)) \\ &\approx G(z(-j+1), z) G(w, z(-j-1)). \end{aligned} \tag{32}$$

Hence for $w \in E_j \setminus B(z(-j), 3)$

$$\begin{aligned} &\frac{\omega(\Delta_{t(z)-j}(z(+\infty)))}{G(o,w)} \approx \frac{\omega^{z(-j-1)}(\Delta_{t(z)-j}(z(+\infty)))}{G(z(-j-1),w)} \\ &\lessapprox \frac{\omega^{z(-j-1)}(\Delta_{t(z)-j}(z(+\infty)))G(z(-j+1),z)}{G(w,z)} \approx \frac{G(z, z(-j+1))}{G(w,z)} \\ &\approx \frac{G(z, z(-j+2))}{G(w,z)} \lessapprox \exp(-C_4 d(z, z(-j+2))) \frac{1}{G(w,z)} \\ &\lessapprox \exp(-C_4 j) \frac{1}{G(w,z)}. \end{aligned}$$

Accordingly, for $w \in E_j \setminus B(z(-j), 3)$,

$$\frac{G(z,w)}{G(o,w)} \leq C \exp(-C_4 j) \frac{1}{\omega(\Delta_{t(z)-j}(z(+\infty)))}. \tag{33}$$

Suppose $w \in E_j \cap B(z(-j), 3)$. In this case, we have $G(z,w) \approx G(z, z(-j))$ and $G(o,w) \approx G(o, z(-j))$. Moreover,

$$\frac{\omega(\Delta_{t(z)-j}(z(+\infty)))}{G(o,w)} \approx \frac{\omega(\Delta_{t(z)-j}(z(+\infty)))}{G(o,z(-j))}$$
$$\approx \frac{\omega^{z(-j-1)}(\Delta_{t(z)-1}(z(+\infty)))}{G(z(-j-1), z(-j))} \approx 1.$$

Therefore

$$\frac{G(z,w)}{G(o,w)} \approx \frac{G(z,z(-j))}{G(o,w)} \approx \frac{G(z,z(-j))}{G(o,w)} \frac{G(o,w)}{\omega(\Delta_{t(z)-j}(z(+\infty)))}$$
$$\approx \exp(-C_4 j) \frac{1}{\omega(\Delta_{t(z)-j}(z(+\infty)))}.$$

Consequently for $4 \le j \le N$ and for $w \in E_j$ we have

$$\frac{G(z,w)}{G(o,w)} \le C \exp(-C_4 j) \frac{1}{\omega(\Delta_{t(z)-j}(z(+\infty)))}. \tag{34}$$

Using (34),

$$\int_{E_j} G(z,w)\, d\mu_0(w) \lessapprox \exp(-C_4 j) \frac{\mu(E_j)}{\omega(\Delta_{t(z)-j}(z(+\infty)))} \tag{35}$$
$$\lessapprox \exp(-C_4 j) \|\mu\|_c.$$

Next we consider the case of $w \in E_0' = E_0 \cup \cdots \cup E_3$. Suppose $w \in E_0' \setminus B(z,r)$, where r is a constant in the condition (F). Then by Theorem BH2 and (21),

$$\frac{\omega(E_0')}{G(o,w)} \approx \frac{\omega^{z(-1)}(E_0')}{G(z(-1),w)} \approx \frac{1}{G(z,w)}.$$

Hence for $0 < t < t(z) - 3$

$$\int_{E_0' \setminus B(z,r)} G(z,w)\, d\mu_0(w) \lessapprox \frac{\mu(C(z(+\infty), t(z)-3))}{\omega(\Delta_{t(z)-3}(z(+\infty)))}$$
$$\lessapprox \frac{\mu(S_t[\Delta_{t(z)-3-k}(z(+\infty))])}{\omega(\Delta_{t(z)-3-k}(z(+\infty)))} \le C\|\mu\|_c.$$

From (F) it follows that

$$\int_{B(z,r)} G(z,w)\, d\mu_0(w) \lessapprox \frac{C_r}{G(o,z)} \int_{B(z,r)} G(z,w)\, d\mu(w) \lessapprox C_r C_{r,\mu}.$$

We consider the case of $w \in M \setminus C(z(+\infty), t(z) - N)$. By Theorem BH2 we have that

$$\frac{G(w,z)}{G(o,z)} \approx \frac{G(w, z(-N+1))}{G(o, z(-N+1))} \approx G(w, z(-N+1)) \leq C.$$

Therefore

$$\int_{M \setminus C(z(+\infty), t(z)-N)} G(z,w)\, d\mu_0(w) \leq C \int_M d\mu \leq C \|\mu\|_c.$$

Summing up, we have that

$$G[\mu_0](z) \leq C(\|\mu\|_c + C_{r,\mu}), \quad \text{whenever } z \in M \setminus B(o, r+2). \tag{36}$$

Next we prove (i) $\Rightarrow$ (ii). Suppose $H := \sup_{z \in M \setminus B(o,R)} G[\mu_0](z) < \infty$ for a positive number R. It is easy to see that $S_r[\Delta_t(Q)] \subset C(Q,t)$ for all $r > 0$. Suppose $t > R + 2$. For $w \in S_R[\Delta_t(Q)]$, $z = \gamma_{oQ}(t - 1/4)$, and $z' = \gamma_{oQ}(t - 1/2)$,

$$\frac{G(z,w)}{G(o,w)} \approx \frac{G(z',w)}{G(o,w)} \approx \frac{G(z',z)}{G(o,z)} \approx \frac{1}{G(o,z)}.$$

On the other hand, for every $x \in M \setminus B(o, 1/2)$,

$$G(o,x) \approx \omega(\Delta_{d(o,x)}(x(+\infty))), \tag{37}$$

because for $x' = \gamma_{ox}(d(o,x) - 1/2)$, Theorem BH2 implies that

$$\frac{G(o,x)}{\omega^o(\Delta_{d(o,x)}(x(+\infty)))} \approx \frac{G(x',x)}{\omega^{x'}(\Delta_{d(o,x)}(x(+\infty)))} \approx 1.$$

Therefore $G(o,z) \approx \omega(\Delta_t(Q))$. Accordingly,

$$\frac{G(z,w)}{G(o,w)} \approx \frac{1}{\omega(\Delta_t(Q))}.$$

From this estimates it follows that

$$H \geq \int_{S_R[\Delta_t(Q)]} \frac{G(z,w)}{G(o,w)}\, d\mu(w) \approx \frac{\mu(S_R[\Delta_t(Q)])}{\omega(\Delta_t(Q))}.$$

To prove $\mu(M) < \infty$, let z be a point in M with $d(o,z) = R + 1/2$. Then

$$H \geq \int_{M \setminus B(o,R+1)} \frac{G(z,w)}{G(o,w)}\, d\mu(w) + \int_{B(o,R+1) \setminus B(o,R)} \frac{G(z,w)}{G(o,w)}\, d\mu(w).$$

Note that

$$\frac{G(z,w)}{G(o,w)} \geq c_1, \quad w \in M \setminus B(o, R+1),$$

$$\frac{G(z,w)}{G(o,w)} \geq c_2 G(z,w) \geq c_3, \quad w \in B(o, R+1) \setminus B(o,R),$$

where c_i $(i = 1, 2, 3)$ are positive constants depending only on M and R. Using this we have that $\mu(M \setminus B(o,R)) \lessapprox H$, and consequently by the assumption, $\mu(M) \lessapprox H + \mu(B(o,R)) < \infty$. Therefore $\|\mu\|_{c,R} < \infty$.

For $z \in M \setminus B(o, R+2)$,

$$\begin{aligned} H &\geq \int_M \frac{G(z,w)}{G(o,w)}\, d\mu(w) \geq \int_{B(z,1)} \frac{G(z,w)}{G(o,w)}\, d\mu(w) \\ &\approx \frac{1}{G(o,z)} \int_{B(z,1)} G(z,w)\, d\mu(w). \end{aligned}$$

Thus μ satisfies the condition (F). Q.E.D.

§9. Littlewood-Paley measures and BMO

In this section we prove a Carleson measure characterization of BMO functions. To state our theorem, we consider Littlewood-Paly type measure on M: for $f \in L^1(\omega)$, let

$$d\mu_f(w) = G(o,w)|\nabla \tilde{f}(w)|^2\, dV(w),$$

where dV is the volume measure with respect to the metric g, and $|\nabla \tilde{f}(w)|$ is the norm of the gradient of $\tilde{f}$ with respect to g, that is, in a local coordinate neighborhood,

$$|\nabla \tilde{f}(w)|^2 = \sum_{ij} g^{ij}(w) \frac{\partial f(w)}{\partial x_i} \frac{\partial f(w)}{\partial x_j},$$

where $(g^{ij}(w))$ is the inverse matrix of the metric $(g_{ij}(w))$. This is an analogue to M of the classical Littlewood-Paley measure.

First we prove the following theorem

Theorem 9.1. *Let $f \in L^2(\omega)$. Then*

$$\mu_f(M) = \int_{S(\infty)} |f(Q) - \tilde{f}(o)|^2\, d\omega(Q) < \infty.$$

Proof. This is an immediate consequence of Dynkin's formula: By (9), $f \in L^2(\omega^z)$. Let $h = f - \tilde{f}(z)$. Then $M_t^h := \tilde{h}(Z_t)$ is an L^2 bounded martingale with respect to $(W, \mathcal{F}^z, \mathcal{F}_t^z, P^z)$. Note that

$$G(z,w) = \int_0^\infty p(t,z,w)\,dt,$$

where $p(t,z,w)$ is the minimal fundamental solution of the equation $\partial/\partial t - \Delta_g$. Hence by Ito's formula we have that

(38)

$$\begin{aligned}
\int_{S(\infty)} |h(Q)|^2\,d\omega^z(Q) &= E^z[|h(Z_\infty)|^2] = E^z\left[\int_0^\infty d\langle M^h, M^h\rangle\right] \\
&= E^z\left[\int_0^\infty |\nabla\tilde{h}(Z_t)|^2\,dt\right] = \int_0^\infty E^z[|\nabla\tilde{h}(Z_t)|^2]\,dt \\
&= \int_0^\infty \int_M p(t,z,w)|\nabla\tilde{h}(w)|^2\,dV(w)\,dt \\
&= \int_M G(z,w)|\nabla h(w)|^2\,dV(x).
\end{aligned}$$

Taking $z = o$, the theorem was proved. Q.E.D.

We can characterize BMO functions in terms of Carleson measures and Green potentials:

Theorem 9.2. *Let $f \in L^2(\omega)$. Then the following are equivalent:*
(i) $f \in \mathrm{BMO}(\omega)$.
(ii) μ_f *is a Carleson measure on* M.
(iii) *The Green potential*

$$G_f(x) := \int_M G(x,w)|\nabla\tilde{f}(w)|^2\,dV(w)$$

is asymptotically bounded.
(iv) *The potential G_f defined in* (iii) *is bounded on* M.

Proof. First we prove "(i) $\Rightarrow$ (ii)". Let $f \in \mathrm{BMO}(\omega)$. Then by Corollary 6.9, $f \in \mathrm{BMO}_{\mathrm{prob}}$. Therefore $M_t := \tilde{f}(Z_t) - \tilde{f}(Z_0)$ is a BMO-martingale with respect to $(W, \mathcal{F}, \mathcal{F}_t, P)$. Hence by [36, p.333], we have that for every $(\mathcal{F}_t)$-stopping time T,

$$E\left[\int_T^\infty d\langle M, M\rangle\right] \le \|f\|_{\mathrm{BMO}_{\mathrm{prob}}} P(T < \infty). \tag{39}$$

Using this inequality we will prove the desired part. For this aim we need the following variant of (38):

Lemma 9.3. *Let $h \in L^1(\omega)$ with $\tilde{h}(0) = 0$, and let $M_t^h = \tilde{h}(Z_t)$. Then for every nonnegative Borel function F on M,*

$$E\left[\int_0^\infty F(Z_t)\, d\langle M^h, M^h\rangle\right] = \int_M F(x)G(o,x)|\nabla\tilde{h}(x)|^2\, dV(x).$$

Since this is proved by the same way as (38), we omit the proof of this lemma. We proceed to prove Theorem 9.2. We will prove that μ_f satisfies the condition (v) in Theorem 7.1 with $p = 2$. Let $\psi \in L^2(\omega)$. For $\lambda > 0$, let $T_\lambda = \inf\{t > 0 : |\tilde{\psi}(Z_t)| > \lambda\}$. Since $\{w \in W : |\tilde{\psi}(Z_s)| > \lambda\} \subset \{w \in W : T_\lambda(w) \leq s\}$, we have

$$L := \{(s,w) \in [0,\infty) \times W : |\tilde{\psi}(Z_s)| > \lambda\} \subset [T_\lambda, \infty[,$$

where $[T_\lambda, \infty[$ is the usual stochastic interval, i.e., $\{(s,w) : T_\lambda(w) \leq t < \infty\}$. Denote $d\nu(s,w) = d\langle M, M\rangle(dw)$. By (39) we have the following inequalities:

$$\begin{aligned}\nu(L) &\leq \nu([T_\lambda,\infty[) \leq \|f\|_{\mathrm{BMO_{prob}}} P(T_\lambda < \infty)\\ &\leq \|f\|_{\mathrm{BMO_{prob}}} P(\{\sup_t |\tilde{\psi}(Z_t)| > \lambda\}).\end{aligned}$$

Therefore we have by Lemma 9.3 and Doob's maximal inequality that

$$\begin{aligned}\int_M |\tilde{\psi}(w)|^2\, d\mu_f(w) &= E\left[\int_0^\infty |\tilde{\psi}(Z_t)|^2\, dt\right] \leq \|f\|_{\mathrm{BMO_{prob}}} E[\sup_t |\tilde{\psi}(Z_t)|^2]\\ &\leq 4\|f\|_{\mathrm{BMO_{prob}}} E[|\psi(Z_\infty)|^2] = 4\|f\|_{\mathrm{BMO_{prob}}} \|\psi\|^2_{L^2(\omega)}.\end{aligned}$$

By the proof of Theorem 7.1 (v) $\Rightarrow$ (i), we have

$$\|\mu\|_c \lessapprox \|f\|_{\mathrm{BMO}}. \tag{40}$$

Next we prove "(ii) $\Rightarrow$ (iii)". By Theorem 8.1, we need to prove only that μ_f satisfies the condition (F). To prove (F), we use the following inequality due to Mouton (see [38, p.502 and p.501]): there exists a positive constant C such that for every harmonic function u on M, and for every $z \in M$,

$$|\nabla u(z)|^2 \leq C \int_{B(z,1)} |\nabla u(w)|^2\, dV(w). \tag{41}$$

Suppose $z \in M \backslash B(o,r)$ for a sufficient large number r to be chosen later. Then $G(o,z) \approx G(o,y)$ $(y \in B(z,1))$. By (37), $G(o,z) \approx \omega(\Delta(z,0))$. Hence

$$\begin{aligned}|\nabla \tilde{f}(z)|^2 &\leq C \int_{B(z,1)} \frac{G(o,y)}{G(o,z)} |\nabla \tilde{f}(y)|^2 \, dV(y) \\ &\leq \frac{C}{\omega(\Delta(z,0))} \int_{B(z,1)} G(o,y) |\nabla \tilde{f}(y)|^2 \, dV(y) \\ &\leq \frac{C}{\omega(\Delta(z,0))} \int_{C(z(+\infty),d(o,z)-c)} G(o,y) |\nabla \tilde{f}(y)|^2 \, dV(y),\end{aligned}$$

where c is a positive constant depending only on κ_1 and κ_2 such that $B(z,1) \subset C(z(\infty), d(o,z)-c)$. We suppose $r > c+1$. Then that μ_f is a Carleson measure implies that

$$\frac{1}{\omega(\Delta(z,0))} \int_{C(z(+\infty),d(o,z)-c)} G(o,y) |\nabla \tilde{f}(y)|^2 \, dV(y) \leq C \|\mu_f\|_c,$$

that is,

$$\sup_{z \in M \backslash B(o,r)} |\nabla \tilde{f}(z)|^2 \leq C \|\mu_f\|_c. \tag{42}$$

Therefore for $z \in M \setminus B(o, r+1)$,

$$\begin{aligned}\int_{B(z,1)} G(z,w) \, d\mu_f(w) &= \int_{B(z,1)} G(z,w) G(o,w) |\nabla \tilde{f}(w)|^2 \, dV(w) \\ &\leq C \|\mu_f\|_c G(o,z) \int_{B(z,1)} G(z,w) \, dV(w) \leq C' \|\mu_f\|_c G(o,z).\end{aligned}$$

This implies that μ_f satisfies the condition (F) with $r+1$.

Now we prove "(iii) $\Rightarrow$ (i)". By (38) we have

$$\int_{S(\infty)} |f - \tilde{f}(z)|^2 \, d\omega^z = \int_M G(z,w) |\nabla \tilde{f}(w)|^2 \, dV(w). \tag{43}$$

Hence the Hölder inequality and the condition (iii) yield that for some $r > 0$,

$$\sup_{z \in M \backslash B(o,r)} \int_{S(\infty)} |f - \tilde{f}(z)| \, d\omega^z < \infty.$$

However, (9) implies that for every $z \in B(o,r)$,

$$\begin{aligned}\int_{S(\infty)} |f - \tilde{f}(z)| \, d\omega^z &\leq C_r \int_{S(\infty)} |f - \tilde{f}(z)| \, d\omega \\ &\leq C_r \|f\|_{L^1(\omega)} + C_r^2 \|f\|_{L^1(\omega)}.\end{aligned}$$

Consequently $\|f\|_G < \infty$, and by Corollary 6.9 we have $f \in \mathrm{BMO}(\omega)$.

Lastly, we prove (i) $\Rightarrow$ (iv). By (43), Corollary 6.9 and Remark 2 we have

$$\|f\|_{G,2}^2 \geq \sup_{x \in M} G_f(x).$$

The part of "(iv) $\Rightarrow$(iii)" is obvious. Thus the theorem was proved.

Q.E.D.

Remark 3. (1) In the classical Euclidean case, the part "(i) $\Leftrightarrow$ (ii)" was proved by C. Fefferman and E. Stein ([22]), and in the case of the complex unit ball endowed with the Bergman metric, the similar reseult to Theorem 9.2 was proved by Jevtic [25]. However, his proof is based on the nature of the ball, and our proof is different from it. See also [26] and [31] for related results.

(2) The proof of the part "(i) $\Rightarrow$ (ii)" is based on an idea in Arai [4]. However this part can be proved also by using Theorem 8.1. Indeed, we have (i) $\Rightarrow$ (iv) and (iv) $\Rightarrow$ (ii) by Theorem 8.1.

§10. Bloch functions on manifolds and a gradient estimate for harmonic functions

In this section we study Bloch functions on M and give an application of Theorem 9.2.

Bloch functions were defined originally on the open unit disc D in $\mathbf{C}$: a holomorphic function f on D is said to be a Bloch function on D if

$$\sup_{z \in D}(1 - |z|)|f'(z)| < \infty. \tag{44}$$

In other word, f is a Bloch function if and only if the norm of gradient $|\nabla f|$ with respect to the Poincáre metric is bounded. Taking this fact into account, Bloch functions on an n-dimensional Riemannian manifold $(\mathcal{R}, h)$ are defined as follows:

Definition 10.1. Let f be a harmonic function on $\mathcal{R}$. Then f is said to be a harmonic Bloch function on $\mathcal{M}$ if

$$\|f\|_B := \sup_{x \in \mathcal{R}} |\nabla f(x)| < \infty,$$

where $|\nabla f|$ is the norm of gradient of f with respect to the metric h, i.e., $|\nabla f(x)|^2 = \sum_{i,j} h^{ij}(x)(\partial f(x)/\partial x_i)(\partial f(x)/\partial x_j)$, where $(h^{ij}(x))$ is the inverse matrix of the Riemannian metric $(h_{ij}(x))$. Denote by $\mathcal{B}(\mathcal{R}, h)$ the linear space consisting of all harmonic Bloch functions on $\mathcal{R}$.

In particular, if $(\mathcal{R}, h)$ is a Kähler manifold, then a function u is said to be a holomorphic Bloch function on M if u is a harmonic Bloch function and holomorphic on $\mathcal{R}$.

The first question we have to ask is whether there exists a nonconstant harmonic Bloch function. As known, existence problem for nonconstant bounded harmonic functions is a crucial theme in geometric analysis. Indeed, this problem has been a motivation of analysis on negatively curved manifold, and M. T. Anderson ([2]) and D. Sullivan ([46]) proved existence of a lot of nonconstant bounded harmonic functions on M. Fortunately, their result implies also existence of nonconstant Bloch functions on M, because a gradient estimate of harmonic functions due to S.-T. Yau ([45, Corollary 3.1], [49]) tells us the following

Proposition Y. *Suppose $(\mathcal{R}, h)$ is a complete Riemannian manifold such that its Ricci curvature is bounded below by a constant. Then a bounded harmonic function on M is a harmonic Bloch function on M.*

Therefore if the manifold $\mathcal{R}$ has a nonconstant bounded harmonic function, then it possesses a nonconstant harmonic Bloch function. However, the converse is not true because of the following easy fact:

Proposition 10.1. *Suppose that $\mathcal{R} = \mathbf{R}^n$ and h is the Euclidean metric. Then $\mathcal{B}(\mathcal{R}, h)$ is an $(n+1)$-dimensional linear space.*

Proof. Since h is the Euclidean metric, we have for $u \in C^1(\mathbf{R}^n)$, $\|u\|_B^2 = \sup_{x\in\mathbf{R}^n} \sum_{j=1}^n |(\partial u(x)/\partial x_j)|^2$. Therefore the coordinate functions $u_j(x) = x_j$ and the constant function $u_0(x) = 1$ are harmonic Bloch functions on $\mathbf{R}^n$. Now let u be a harmonic Bloch function $\mathbf{R}^n$. Then $\partial u/\partial x_j$ is also harmonic on $\mathbf{R}^n$ $(j = 1, \ldots, n)$, and by definition it is bounded on $\mathbf{R}^n$. By Liouville's theorem $\partial u/\partial x_j$ must be a constant. Consequently u must be an affine function on $\mathbf{R}^n$. Q.E.D.

Since by Liouville's theorem there is no nonconstant bounded harmonic functions on $\mathbf{R}^n$, it is interested to find a geometric condition in order that a unbounded harmonic or holomorphic Bloch function exists, but it is beyond the scope of this paper to study the problem. (See Remark 4 (1) and Li and Tam [32].)

However it might be worthwhile to point out that in the case of our manifold M, Theorem 9.2 guarantees that harmonic extensions of unbounded BMO functions are unbounded harmonic Bloch functions:

Theorem 10.2. *Suppose* $f \in \mathrm{BMO}(\omega)$. *Then* $\tilde{f}$ *is a harmonic Bloch function on* M. *Indeed*

$$\sup_{x \in M} |\nabla \tilde{f}(x)| \leq C \|f\|_{\mathrm{BMO}}, \tag{45}$$

where C *is a positive constant depending only on* M *and* o.

In particular, there exists a unbounded BMO *function* b, *and* $\tilde{b}$ *is a unbounded harmonic Bloch function on* M.

Proof. We begin with proving the first assertion: By Theorem 9.2 we have that the Littlewood-Paley measure μ_f is a Carleson measure on M. Hence from (42) it follows that

$$\|\tilde{f}\|_B^2 \lessapprox \sup_{z \in B(o,r)} |\nabla f(z)|^2 + \|\mu_f\|_c \lessapprox \sup_{z \in B(0,r)} |\nabla \tilde{f}(z)|^2 + \|f\|_{\mathrm{BMO}}^2 < \infty,$$

where r is a positive constant in (42). In addition, by [45, Corollary 3.2] and Lemma 5.3 (i) we have

$$\begin{aligned}\sup_{z \in B(o,r)} |\nabla \tilde{f}(z)| &\lessapprox \sup_{z \in B(o,2r)} |\tilde{f}(z)| \leq \sup_{z \in B(o,2r)} \int_{S(\infty)} K(z,Q)|f(Q)|\, d\omega(Q) \\ &\lessapprox \|f\|_{L^1(\omega)} \leq \|f\|_{\mathrm{BMO}}.\end{aligned}$$

Hence $\|\tilde{f}\|_B \lessapprox \|f\|_{\mathrm{BMO}}$.

The second assertion is an immediate consequence of the first one. For if u is a bounded harmonic function on M, then by Fatou's theorem for M ([3]), we have that there exists $f \in L^\infty(\omega)$ satisfying $u = \tilde{f}$ on M. However, $L^\infty(\omega) \subsetneq \mathrm{BMO}(\omega)$ (see Appendix 2). Therefore, a unbounded harmonic Bloch function exists. Q.E.D.

Suppose u is a bounded harmonic function on M. Then there exists $f \in L^\infty(S(\infty), \omega)$ such that $\tilde{f} = u$. From Yau [45, Corollary 3.1] it follows that

$$\sup_{x \in M} |\nabla u(x)| \lessapprox \|f\|_{L^\infty(\omega)}. \tag{46}$$

On the other hand our inequality (45) implies that

$$\sup_{x \in M} |\nabla u(x)| \lessapprox \|f\|_{\mathrm{BMO}} \left(\leq 3\|f\|_{L^\infty(\omega)} \right),$$

which refines (46).

Remark 4. (1) Suppose $\mathcal{R}$ is a complete manifold with nonnegative Ricci curvature and u is a harmonic function on $\mathcal{R}$. Then it is known that by Yau's estimate u is a linear growth harmonic function if and only if $|\nabla u|$ is bounded (see [32]).

(2) Let $\mathbf{T} = \{z \in \mathbf{C} : |z| = 1\}$, and BMOA($\mathbf{T}$) the set of all functions f in BMO($\mathbf{T}$) such that the Poisson integral of f is holomorphic in D. Then it is known that if $f \in$ BMOA($\mathbf{T}$), then its Poisson integral is a holomorphic Bloch function on $D = \{z \in \mathbf{C} : |z| < 1\}$ (cf. [41]). This was extended to bounded, strongly pseudoconvex domain with smooth boundary by Krantz and Ma [30]. Our proof of Theorem 10.2 is different from their proofs. We note that the inequality (45) is an analogue to M of Jerison and Kenig [26, Lemma 9.9].

§11. Boundary behavior of harmonic Bloch functions

In the classical case of the unit disc in $\mathbf{C}$, a lot of unbounded holomorphic Bloch functions are known. For instance, $u(z) = \sum_{k=m}^{\infty} z^{15^k}$ ($z \in D$) is a holomorphic Bloch function, and it is known that for large m,

$$\limsup_{r\to 1} \frac{|u(re^{i\theta})|}{\sqrt{\log(1-r)^{-1}\log\log\log(1-r)^{-1}}} > 0.685\|u\|_B \text{ a.e. } \theta \in [0, 2\pi)$$

(see [41, p.194]). This means that u is not only unbounded, but also it has no boundary limits at almost every boundary point. On the other hand, Makarov proved the following

Theorem M (Makarov [34]; see also Pommerenke [41, p.186]). *Let u be a holomorphic Bloch function on D. Then for almost every $\theta \in [0, 2\pi)$,*

$$\limsup_{r\to 1} \frac{|u(re^{i\theta})|}{\sqrt{\log(1-r)^{-1}\log\log\log(1-r)^{-1}}} \le \|u\|_B.$$

Somewhat later a probabilistic version of Theorem M was proved in [33]:

Theorem L (Lyons [33]). *Let u be a holomorphic Bloch function on D. Let X_t be hyperbolic Brownian motion on D. Then*

$$\limsup_{t\to\infty} \frac{|u(X_t)|}{\sqrt{\log(1-|X_t|)^{-1}\log\log\log(1-|X_t|)^{-1}}} \le \|u\|_B.$$

In the higher dimensional case, little is known about unbounded Bloch functions. Recently, D. Ullrich constructed a holomorphic Bloch function on the open unit ball in $\mathbf{C}^n$ which has no finite radial limits ([47]).

In the rest of this section, we study boundary behavior of harmonic Bloch functions on M, and generalize Theorem L to the manifold M. As first we characterize Bloch functions in terms of Brownian motion:

Theorem 11.1. *For a harmonic function u on M, the following* (i) *and* (ii) *are equivalent:*

(i) *u is a harmonic Bloch function on M.*

(ii) *The stochastic process $\{u(Z_t)\}_t$ satisfies that*

$$\|u\|_{B,\mathrm{prob}}^2 := \sup_{x\in M}\left\{\frac{E_x[|u(Z_T)-u(Z_0)|^2]}{E_x[T]} : T\in\mathcal{T}_x,\ E_x[T]>0\right\}<\infty,$$

where $\mathcal{T}_x$ is the set of all $(\mathcal{F}_t^x)$-stopping times. Furthermore, $\|u\|_B \leq \|u\|_{B,\mathrm{prob}} \leq \sqrt{2}\,\|u\|_B$.

In the case of the open unit disc in $\mathbf{C}$, a martingale characterization of holomorphic Bloch functions was given in Muramoto [39]. We will prove Theorem 11.1 by simplifying and exploiting the method in [39] by combining an idea in Lyons [33].

Proof. (i) $\Rightarrow$ (ii). Let u be a harmonic Bloch function on M. By Ito's formula we have that

$$E_x[|u(Z_T)-u(Z_0)|^2] = 2E_x\left[\int_0^T |\nabla u(Z_s)|^2\,ds\right] \leq 2\|u\|_B^2 E_x[T],$$

for every $x\in M$ and $T\in\mathcal{T}_x$. Therefore $\|u\|_{B,\mathrm{prob}} \leq \sqrt{2}\,\|u\|_B$.

(ii) $\Rightarrow$ (i). Suppose $\|u\|_{B,\mathrm{prob}}<\infty$. Let α be an arbitrary number with $0<\alpha<\|u\|_B$. Then there exists a geodesic ball $B(z,\varepsilon)$ such that $\alpha \leq |\nabla u(x)|$ for all $x\in B(z,\varepsilon)$. Let $x\in B(z,\varepsilon)$ and $T=\inf\{t>0 : Z_t\notin B(z,\varepsilon)\}$. By the definition of $\|u\|_{B,\mathrm{prob}}$ we have

$$\begin{aligned}\alpha^2 E_x[T] &= E_x\left[\int_0^T \alpha^2\,ds\right] \leq E_x\left[\int_0^T |\nabla u(Z_t)|^2\right]\\ &= E_x[|u(Z_T)-u(Z_0)|^2] \leq \|f\|_{B,\mathrm{prob}}^2 E_x[T].\end{aligned}$$

Therefore $\alpha \leq \|u\|_{B,\mathrm{prob}}$. Thus $\|u\|_B \leq \|f\|_{B,\mathrm{prob}}$. Q.E.D.

Now we discuss on boundary behavior along Brownian paths of Bloch functions u on M. Since $M^u := \{u(Z_t) - u(Z_0)\}_t$ is a continuous local $(\mathcal{F}_t, P)$-martingale, it is known that the sets $\{\langle M^u, M^u\rangle_\infty < \infty\}$ and $\{\lim_{t\to\infty} M_t^u \text{ exists}\}$ are almost surely equal. Therefore we are interested to the behavior of M^u on the set $\{\langle M^u, M^u\rangle_\infty = \infty\}$:

Theorem 11.2. *Let u be a harmonic Bloch functions on M. Then*

$$\limsup_{t\to\infty} \frac{|u(Z_t)|}{\sqrt{d(o, Z_t) \log\log d(o, Z_t)}} \le C\|u\|_B \quad P\text{-a.s.}$$

Proof. By virtue of Theorem 11.1 we can apply an idea in [33] to our setting: Let $M_t = u(Z_t) - u(Z_0)$ and $T(t) := \inf\{s : \langle M, M\rangle_s > t\}$. By [42, Theorem 1.7, p.182], there exists an enlargement $(\tilde{W}, \tilde{\mathcal{F}}_t, \tilde{P})$ and a Brownian motion $\tilde{\beta}$ on $\tilde{W}$ independent of M^u such that the process

$$B_t := \begin{cases} M_{T(t)}, & t < \langle M, M\rangle_\infty, \\ M_\infty + \tilde{\beta}_{t-\langle M,M\rangle_\infty}, & t \ge \langle M, M\rangle_\infty, \end{cases}$$

is a standard linear Brownian motion. Therefore by the classical law of the iterated logarithm we have

$$\limsup_{t\to\infty} \frac{|B_t|}{\sqrt{t \log\log t}} = 1 \quad \tilde{P}\text{-a.s.} \tag{47}$$

Now on the set $A = \{\langle M, M\rangle_\infty = \infty\}$, we have $M_t = B_{\langle M,M\rangle_t}$, and

$$\langle M, M\rangle_s = 2\int_0^s |\nabla u(Z_r)|^2\, dr \le 2\|u\|_B^2 s.$$

Consequently, we have

$$\limsup_{t\to\infty} \frac{|u(Z_t) - u(o)|}{\sqrt{t \log\log t}} \le C\|u\|_B$$

P-a.s. on $\{\langle M, M\rangle_\infty = \infty\}$. On the other hand $t \approx d(o, Z_t)$ as $t \to \infty$ (see [35, (3.2), p.254]). Therefore we have the desired inequality P-a.s. on $\{\langle M, M\rangle_\infty = \infty\}$. Thus the theorem was proved. Q.E.D.

As an immediate consequence of Theorem 11.2 we have the following

Corollary 11.3. *Let* $M = \{x \in \mathbf{R}^n : |x| < 1\}$ *and let* g *be the hyperbolic metric on* M. *Then for a harmonic Bloch function* u *on* (M, g),

$$\limsup_{t\to\infty} \frac{|u(Z_t) - u(o)|}{\sqrt{\log(1-|Z_t|)^{-1} \log\log\log(1-|Z_t|)^{-1}}} \le C\|u\|_B \quad a.s.\ P^o.$$

§12. Appendix 1 (A proof of Theorem BH2 (2))

Proof of Theorem BH2 (2). Let $\mathcal{C}_0$ (resp. $\mathcal{C}_0'$) be the cone with vertex $z(t)$ (resp. $z(t+1)$), direction tangent to $\partial C(z,t)$ (resp. $\partial C(z,t+1)$) of sufficiently small angle θ defined as in [1, p.518]. Consider sequences of cones $\mathcal{C}_1, \ldots, \mathcal{C}_k$ and $\mathcal{C}_1', \ldots, \mathcal{C}_l'$ obtained by iterated 1-shift of $\mathcal{C}_0$ and $\mathcal{C}_0'$ respectively. Then by our curvature condition, there exists a positive number ε depending only on κ_1 and κ_2 such that for an angle $\theta > \varepsilon$ the sequence of cones $\mathcal{C}_k^c, \ldots, \mathcal{C}_0^c$, $C(z, t+1/2)$, $\mathcal{C}_0', \ldots, \mathcal{C}_l'$ together with their vertices is a Φ-chain through $z(t+1/2)$ in the sense of Ancona [1], where Φ is depending only on κ_1 and κ_2. Moreover, by the remark after Proposition 15 in [1], we have that the sequence $(\mathcal{C}_l')^c, \ldots, (\mathcal{C}_0')^c$, $C(z,t+1/2)^c$, $\mathcal{C}_0, \ldots, \mathcal{C}_k$ together with their vertices is a Ψ-chian through $z(t+1/2)$ where Ψ depends only on Φ. Therefore from this observation it follows that the sets $M \cap C(z,t+1)^c$, $M \cap C(z,t+1/2)^c$ and $M \cap C(z,t)^c$ satisfy the assumption of [1, p.519, Theorem 5′]. Q.E.D.

Remark 5. It is easy to see that also Theorem BH1 for cones with more general aperture implies Theorem BH2.

§13. Appendix 2 (Unbounded BMO functions)

As is well known, the function $\log|x-1|$ on the unit sphere is unbounded but belongs to the classical BMO space. However, it seems to be difficult to construct a unbounded BMO function on the sphere at infinity. In this section we give a nonconstructive proof of existence of unbounded BMO function:

Proposition 13.1. *Suppose that* (X, ρ, μ) *is a space of homogeneous type in the sense of Coifman and Weiss* [15], *and that* $\mu(X) = 1$. *For* $x \in X$ *and* $r > 0$, *let* $B(x,r) = \{y \in X : \rho(x,y) < r\}$. *Assume that* $\mu(B(x,r)) > 0$ *for all* $x \in X$ *and* $r > 0$, *and that* $\lim_{r\to 0}\mu(B(x,r)) = 0$ *for every* $x \in X$. *Then* $L^\infty(X) \neq \mathrm{BMO}(X)$.

Proof. Suppose $L^\infty(X) = \mathrm{BMO}(X)$. Then it is easy to see that the norms $\|\cdot\|_{L^\infty}$ and $\|\cdot\|_{\mathrm{BMO}}$ must be equivalent. Therefore by H^1_{atom}-BMO duality and L^1-L^∞ duality, we have that the norms $\|\cdot\|_{L^1(\omega)}$ and $\|\cdot\|_{H^1_{\mathrm{atom}}}$ are equivalent on L^2. Since L^2 is dense in L^1 and H^1_{atom}, we have $H^1_{\mathrm{atom}} = L^1$. By [15], H^1_{atom} is isomorphic to the dual space of the Banach space VMO. Therefore L^1 also is isomorphic to the dual of VMO. However this is not the case as we will prove below (see Remark 6 after the proof). Using this isomorphism we can define on the space L^1 the topology induced from the weak $*$ topology of the dual of VMO. Moreover, with this topology, L^1 is a locally convex topological linear space. From Banach-Alaoglu theorem it follows that the closed unit ball $B(L^1)$ of L^1 is compact convex set in the induced topology. Therefore by Krein-Milman's theorem $B(L^1)$ must have at least one extreme point. However L^1 has no extreme points. For if $f \in B(L^1)$, then we can consider the following two cases: (Case 1) There exists a Borel set $E \subset X$ such that $\int_E |f|\,d\mu \in (0,1)$. (Case 2) For every Borel set $E \subset X$, $\int_E |f|\,d\mu = 1$ or $= 0$. In the first case, f is not an extreme point of $B(L^1)$, because $f = \alpha f_1 + (1-\alpha) f_2$, where $\alpha = \int_E |f|\,d\mu \in (0,1)$, $f_1 = \alpha^{-1} f\chi_E$, and $f_2 = (1-\alpha)^{-1} f\chi_{X\setminus E}$ (χ_A is the characteristic function of A). The second case does not happen: Let $\|f\|_{L^1} > 0$. Then we have $\mu(\{0 < |f| < \infty\}) > 0$. By Lebesgue's differential theorem, for almost every $x \in \{0 < |f| < \infty\}$,

$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f|\,d\mu = |f(x)| \in (0,\infty). \tag{48}$$

However, in the second case,

$$\frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f|\,d\mu = \frac{1}{\mu(B(x,r))} \quad \text{or} \quad 0,$$

and $1/\mu(B(x,r)) \to \infty$ as $r \to 0$. This contradicts (48). Q.E.D.

Remark 6. A. Pełczyński proved that for a σ-finite and non purely atomic measure space (Σ, μ), the space $L^1(\Sigma,\mu)$ is not isomorphic to any conjugate Banach spaces ([40]).

§14. Appendix 3 (Local Fatou-Doob theorem revisited)

In this section we study local version of Doob-Fatou type theorem on boundary behavior of harmonic functions, and yet another definition of Hardy spaces.

For $t > 0$, let $T_d^t(Q) = T_d(Q) \cap (M \setminus B(o,t))$ $(d > 0)$ and $\Gamma_\alpha^t(Q) = \Gamma_\alpha(Q) \cap (M \setminus B(o,t))$ $(\alpha \in \mathbf{R})$. We say a function f converges admissibly (resp. nontangentially) to l at $Q \in S(\infty)$ or has an admissible limit (resp. a nontangential limit) l if $\lim_{k\to\infty} f(x_k) = l$ for every $\alpha \in \mathbf{R}$ (resp. $d > 0$) and for every sequence $\{x_k\}_k \subset \Gamma_\alpha(Q)$ (resp. $T_d(Q)$) with $x_k \to Q$ as $k \to \infty$. We wirte the admissible limit (resp. nontangential limit) by ad-$\lim_{x\to Q} f(x)$ (resp. K-$\lim_{x\to Q} f(Q)$).

In 1989 we proved a local version of Fatou-Doob theorem under a technical assumption on admissible regions (see [5, Theorem 5.1]). However, by virtue of results in Cifuentes and Korányi [14] we can remove the assumption, and obtain the following theorem by the same way as in [5, Theorem 5.1]:

Theorem 14.1. *Let h be a positive harmonic function on M. Let E be a Borel subset of $S(\infty)$, and u a harmonic function on M. Assume that for each $Q \in E$, there exist $t > 0$ and $\alpha \in \mathbf{R}$ such that u/h is bounded below on $\Gamma_\alpha^t(Q)$. Then there exist $F_1, F_2 \subset S(\infty)$ such that $\omega(F_1) = \mu_h(F_2) = 0$, where μ_h is the Martin representing measure for h, and that u/h converges admissibly at every point in $E \setminus (F_1 \cup F_2)$.*

Recently, F. Mouton [38] proved local versions of Fatou theorem and of Calderón-Stein type theorem. We note that the former follows also from Theorem 14.1 with $h = 1$.

As an application of Theorem 14.1 we prove the equivalence of $H^p(\omega)$ and $H^p(M)$ which is defined in §3:

Theorem 14.2. *Suppose $1 \le p \le \infty$. Let $u \in H^p(M)$. Then there exists $f \in H^p(\omega)$ such that $\tilde{f} = u$ on M.*

Proof. Here we prove only the case of $p = 1$. We can apply the following proof to other $p \in (1, \infty)$. Because of Theorem 14.1 (or [38, Theorem 5.1] and Theorem CK1), for ω-a.e. $Q \in S(\infty)$, the admissible limit $f(Q) =$ ad-$\lim_{x\to Q} u(x)$ exists. Furthermore, $u_t(Q) = u(\gamma_{oQ}(t))$ is continuous on $S(\infty)$ and $f(Q) = \lim_{t\to\infty} u_t(Q)$ ω-a.e. $Q \in S(\infty)$ and $|u_t(Q)| \le N_0(u)(Q)$. Therefore we have that the function $f : Q \mapsto f(Q)$ is measurable and $f \in L^1(\omega)$.

It remains to prove that $u = \tilde{f}$. Let $z \in M$. By (9), $N_0(f) \in L^1(\omega^z)$. Since Theorem 6.2 holds true also for ω^z, we have that $\sup_{0\le t<\infty} |u(Z_t)| \in L^1(W, P_z)$. Therefore by the martingale convergence theorem there exists $F_z \in L^1(W, P_z)$ such that $\lim_{t\to\infty} u(Z_t) = F_z$ and $u(Z_t) = E_z[F_z/\mathcal{F}_t^z]$ (P_z-a.s.). For $Q \in S(\infty)$, let (P_z^Q, Z_t) be the conditioned Brownian motion to exit M at Q (see [38, 3.3] for the definition and basic properties). Then

$$\int_{S(\infty)} P_z^Q(\{\lim_{t\to\infty} u(Z_t) = F_z\})\, d\omega^z(Q) = P_z(\{\lim_{t\to\infty} u(Z_t) = F_z\}) = 1.$$

Hence for ω^z-a.e. $Q \in S(\infty)$, $P_z^Q(\{\lim_{t\to\infty} u(Z_t) = F_z\}) = 1$. From this we prove the following assertion:

Assertion. *There exists a constant $G_{Q,z}$ such that $\lim_{t\to\infty} u(Z_t) = G_{Q,z}$ P_z^Q-a.e.*

Proof of Assertion. Let $I_j = [j, j+1)$, and $L_j = \{\lim_{t\to\infty} u(Z_t)$ exists in $I_j\}$ ($j \in \mathbf{Z}$). Then L_j is asymptotic in the sense of [38, p.482], and therefore the 0-1 law implies that $P_z^Q(L_j) = 1$ or 0. Therefore there exists a unique number $j \in \mathbf{Z}$ such that $P_z^Q(L_j) = 1$ and $P_z^Q(L_i) = 0$ for $i \neq j$. We write the interval I_j by H_1. Consider the dyadic decomposition of H_1, namely $H_{11} = [j, j+1/2)$ and $H_{12} = [j+1/2, j+1)$. Let $L_{1i} = \{\lim_{t\to\infty} u(Z_t)$ exists in $H_{1i}\}$ ($i = 1, 2$). Then again by 0-1 law we have either $P_z^Q(L_{11}) = 1$ or $P_z^Q(L_{12}) = 1$. We denote by H_2 the interval H_{1i} with $P_z^Q(L_{1i}) = 1$. Continuing this procedure we get a decreasing sequence of intervals $\{H_k\}_{k=1,2,\ldots}$ such that $P_z^Q(\{\lim_{t\to\infty} u(Z_t)$ exists in $H_k\}) = 1$. Since there exists a point $G_{Q,z} \in \mathbf{R}$ such that $\{G_{Q,z}\} = \bigcap_k H_k$, the assertion is proved.

End of the proof of Assertion.

Because of (9) and Theorem CK1, K-$\lim_{x\to Q} u(x) = f(Q)$ for ω^z-a.e. $Q \in S(\infty)$. Hence by [38, Corollary 4.4] we have that $f(Q) = G_{Q,z}$ for ω^z-a.e. $Q \in S(\infty)$.

Since $Z_\infty = \lim_{t\to\infty} Z_t = Q$ a.s. P_z^Q for every $Q \in S(\infty)$, we have that for ω^z-a.e. $Q \in S(\infty)$,

$$\begin{aligned} 1 &= P_z^Q(\{\lim_{t\to\infty} u(Z_t) = G_{Q,z}\}) = P_z^Q(\{\lim_{t\to\infty} u(Z_t) = f(Q)\}) \\ &= P_z^Q(\{\lim_{t\to\infty} u(Z_t) = f(Z_\infty)\}). \end{aligned}$$

Therefore $P_z(\{\lim_{t\to\infty} u(Z_t) = f(Z_\infty)\}) = 1$. This implies that $F_z = f(Z_\infty)$ P_z-a.s. Thus

$$u(z) = E_z[u(Z_0)] = E_z[E_z[F_z/\mathcal{F}_0^z]] = E_z[f(Z_\infty)] = \int_{S(\infty)} f\, d\omega^z.$$

Q.E.D.

From this theorem it follows that $H^p(M)$ is naturally identified with $H^p(\omega)$.

References

[1] A. Ancona, Negatively curved manifolds, elliptic operators, and the Martin boundary, Ann. of Math., **125** (1987), 495–536.

[2] M. T. Anderson, The Dirichlet problem at infinity for manifolds of negative curvature, J. Diff. Geometry, **18** (1993), 701–721.

[3] M. T. Anderson and R. Schoen, Positive harmonic functions on complete manifolds of negative curvatures, Ann. of Math., **121** (1985), 429–461.

[4] H. Arai, Measure of Carleson type on filtrated probability spaces and the corona theorem on complex Brownian spaces, Proc. Amer. Math. Soc., **96** (1986), 643–647.

[5] H. Arai, Harmonic analysis on negatively curved manifolds I, Proc. Japan Acad., **63** (1987), 239–242.

[6] H. Arai, Boundary behavior of functions on complete manifolds of negative curvature, Tohoku Math. J., **41** (1989), 307–319.

[7] H. Arai, Kähler diffusions, Carleson measures and BMOA functions of several complex variables, Complex Variables, Theory and Appl., **22** (1993), 255–266.

[8] H. Arai, Degenerate elliptic operators, Hardy spaces and diffusions on strongly pseudoconvex domains, Tohoku Math. J., **46** (1994), 469–498.

[9] R. F. Bass, "Probabilistic Techniques in Analysis", Springer-Verlag, New York, Berlin, Heidelberg, 1995.

[10] D. L. Burkholder, R. F. Gundy and M. L. Silverstein, A maximal function characterization of the class H^p, Trans. Amer. Math. Soc., **157** (1971), 137–153.

[11] J. S. Choa and B. R. Choe, A Littlewood-Paley type identity and a characterization of BMOA, Complex Variables, **17** (1991), 15–23.

[12] P. Cifuentes, H^p-classes on rank one symmetric spaces of noncompact type, I. Nontangential and probabilistic maximal functions, Trans. Amer. Math. Soc., **294** (1986), 133–149.

[13] P. Cifuentes, H^p-classes on rank one symmetric spaces of noncompact type, II. Nontangential maximal function and area integral, Bull. Sci. math. 2^e série **108** (1984), 355–371.

[14] P. Cifuentes and A. Korányi, Admissible convergence in Cartan-Hadamard manifolds, submitted for publication.

[15] R. R. Coifman and G. Weiss, Extensions of Hardy spaces and their use in analysis, Bull. Amer. Math. Soc., **83** (1977), 569–645.

[16] A. Debiard, Comparison des espaces H^p géométrique et probabilistes au dessus de l'espace hermitien hyperbolique, Bull. Sci. Math., **103** (1979), 305–351.

[17] A. Debiard, Espaces H^p au-dessus de l'espace hermitien hyperbolique de $\mathbf{C}^n$ $(n > 1)$ II, J. Functional Analysis, **40** (1981), 185–265.

[18] R. Durrett, "Brownian Motion and Martingales in Analysis", Wadsworth, Belmont, CA, 1984.

[19] P. Eberlein and B. O'Neill, Visibility manifolds, Pacific J. Math., **46** (1973), 45–109.

[20] E. Fabes and C. Kenig, On the Hardy space H^1 of a C^1 domain, Ark. Mat., **19** (1981), 1–22.

[21] E. Fabes, C. Kenig and U. Neri, Carleson measures, H^1 duality and weighted BMO in nonsmooth domains, Indiana Univ. Math., **30** (1981), 547–581.

[22] C. Fefferman and E. M. Stein, H^p spaces of several variables, Acta Math., **129** (1972), 137–193.

[23] D. Geller, Some results in H^p theory for the Heisenberg group, Duke Math. J., **47** (1980), 365–390.

[24] R. K. Getoor and M. J. Sharpe, Conformal martingales, Invent. Math., **16** (1972), 271–308.

[25] M. Jevtić, On the Carleson measures characterization of BMO functions on the unit sphere, Proc. Amer. Math. Soc., **123** (1995), 3371–3377.

[26] D. S. Jerison and C. E. Kenig, Boundary behavior of harmonic functions in non-tangentially accessible domains, Adv. in Math., **46** (1982), 80–147.

[27] Y. Kifer, Brownian motion and positive harmonic functions on complete manifolds of non-positive curvature, in "From Local Times to Global Geometry, Control and Physics", Pitman Reseach Notes in Math. Ser. **150**, 1986, pp. 187–232.

[28] A. Korányi, Boundary behavior of Poisson integrals on symmetric spaces, Trans. Amer. Math. Soc., **140** (1969), 393–409.

[29] A. Korányi and J. C. Taylor, Fine convergence and admissible convergence for symmetric spaces of rank one, Trans. Amer. Math. Soc., **263** (1981), 169–181.

[30] S. G. Krantz and D. Ma, Bloch functions on strongly pseudoconvex domains, Indian Univ. Math. J., **37** (1988), 145–163.

[31] S. G. Krantz and S.-Y. Li, A note on Hardy spaces and functions of bounded mean oscillation on domains in $\mathbf{C}^n$, Michigan Math. J., **41** (1994), 51–71.

[32] P. Li and L.-F. Tam, Linear growth harmonic functions on a complete manifolds, J. Diff. Geometry, **29** (1989), 421–425.

[33] T. Lyons, A synthetic proof of Makarov's law of the iterated logarithm, Bull. London Math. Soc., **22** (1990), 159–162.

[34] N. G. Makarov, On the distortion of boundary sets under conformal mappings, Proc. London Math. Soc., **51** (1985), 369–384.

[35] P. Malliavin, Diffusions et géométrie différentielle globale, C.I.M.E., **25** (1975), 207–279.

[36] P. A. Meyer, Un cours sur les intégrales stochastiques, in "Séminare de Probabilités, X (Seconde partie: Théorie des intégrales stochastiques,

Univ. Strasbourg, Strasbourg, année universitaire 1974/1975)", Lecture Notes in Math., **511**, Springer, Berlin, 1976, pp. 245–400.

[37] P. A. Meyer, Le dual de $H^1(\mathbf{R}^\nu)$, Lect. Notes in Math., **581** (1977), 132–195.

[38] F. Mouton, Comportement asymptotique des fonctions harmoniques en courbure négative, Comment. Math. Helv., **70** (1995), 475–505.

[39] K. Muramoto, Bloch functions on the unit disk and martingale, Math. J. Toyama Univ., **13** (1990), 45–50.

[40] A. Pełcziński, On Banach spaces containing $L^1(\mu)$, Studia Math. XXX (1968), 231–246.

[41] Ch. Pommerenke, "Boundary Behaviour of Conformal Maps", Springer-Verlag, 1992.

[42] D. Revuz and M. Yor, "Continuous Martingales and Brownian Motion, Third Ed.", Springer, 1999.

[43] L. Roger and D. Williams, "Diffusions, Markov Processes, and Martingales, vol. 1 (2nd ed.), vol. 2", John Wiley & Sons, Chichester, New York, Brisbane, Tronto, Singapore, 1994, 1987.

[44] W. Rudin, "Real and Complex Analysis, Third Ed.", McGraw-Hill Book Co., 1986.

[45] R. Schoen and S. T. Yau, "Lectures on Differential Geometry", International Press, 1994.

[46] D. Sullivan, The Dirichlet problem at infinity for a negatively curved manifold, J. Diff. Geom., **18** (1983), 723–732.

[47] D. C. Urllich, A Bloch functions in the unit ball with no radial limits, Bull. London Math. Soc., **20** (1988), 337–341.

[48] N. Th. Varopoulos, A probabilistic proof of the Garnett-Jones theorem on BMO, Pacific J. Math., **90** (1980), 201–221.

[49] S.-T. Yau, Harmonic functions on a complete Riemannian manifolds, Comm. Pure Appl. Math., **28** (1975), 201–228.

[50] S.-T. Yau, On the heat kernel of a complete Riemannian manifold, J. Math. Pure Appl., **57** (1978), 191–201.

[51] T. M. Wolniewicz, Isomorphism between Hardy spaces on complex balls, Ark. Mat., **27** (1989), 155–168.

Graduate School of Mathematical Sciences
The University of Tokyo
3-8-1 Komaba
Meguro-ku
Tokyo, 153-8914
Japan
harai@mint.ocn.ne.jp

Advanced Studies in Pure Mathematics 31, 2001
Taniguchi Conference on Mathematics Nara '98
pp. 51–74

Quantum Vertex Algebras

Richard E. Borcherds

Contents

§1. Introduction

The purpose of this paper is to make the theory of vertex algebras trivial. We do this by setting up some categorical machinery so that vertex algebras are just "singular commutative rings" in a certain category. This makes it easy to construct many examples of vertex algebras, in particular by using an analogue of the construction of a twisted group ring from a bicharacter of a group. We also define quantum vertex algebras as singular braided rings in the same category and construct some examples of them. The constructions work just as well for higher dimensional analogues of vertex algebras, which have the same relation to higher dimensional quantum field theories that vertex algebras have to one dimensional quantum field theories.

One way of thinking about vertex algebras is to regard them as commutative rings with some sort of singularities in their multiplication. In algebraic geometry there are two sorts of morphisms: regular maps that are defined everywhere, and rational maps that are not defined everywhere. It is useful to think of a commutative ring R as having a regular multiplication map from $R \times R$ to R, while vertex algebras only have some sort of rational or singular multiplication map from $R \times R$ to

---

Received March 8, 1999.

Supported by a Royal Society professorship. This paper was written at the Max-Planck institute in Bonn.

R which is not defined everywhere. One of the aims of this paper is to make sense of this, by defining a category whose bilinear maps can be thought of as some sort of maps with singularities.

The main idea for constructing examples of vertex algebras in this paper is a generalization of the following well-known method for constructing twisted group rings from bicharacters of groups. Suppose that L is a discrete group (or monoid) and R is a commutative ring. Recall that an R-valued bicharacter of L is a map $r : L \times L \to R$ such that

$$\begin{aligned} r(1,a) &= r(a,1) = 1, \\ r(ab,c) &= r(a,c)r(b,c), \\ r(a,bc) &= r(a,b)r(a,c). \end{aligned}$$

If r is any R-valued bicharacter of L then we define a new associative multiplication $\circ$ on the group ring $R[L]$ by putting $a \circ b = abr(a,b)$. We call $R[L]$ with this new multiplication the twisted group ring of L. The point is that this rather trivial construction can be generalized from group rings to bialgebras in additive symmetric tensor categories. We will construct vertex algebras by applying this construction to "singular bicharacters" of bialgebras in a suitable additive symmetric tensor category.

Section 2 describes how to generalize the twisted group ring construction to bialgebras, and constructs several examples of singular bicharacters that we will use later. Much of Section 2 uses an extra structure on the spaces underlying many common vertex algebras that is often overlooked. It is well known that these spaces often have natural ring structures, but what is less well known is that this can usually be extended to a cocommutative bialgebra structure. The comultiplication turns out to be very useful for keeping track of the behavior of vertex operators; this is not so important for vertex algebras, but is very useful for quantum vertex algebras. It also allows us to interpret these spaces as the coordinate rings of gauge groups.

Section 3 contains most of the hard work of this paper. We have to construct a category in which the commutative rings are more or less the same as vertex algebras. The motivation for the construction of this category comes from classical and quantum field theory (though it is not necessary to know any field theory to follow the construction). The idea is to construct categories which capture all the formal operations one can do with fields. For examples, fields can be added, multiplied, differentiated, multiplied by functions on spacetime, and we can change variables and restrict fields. All of these operations are trivial but there are so many of them that it takes some effort to write down all the compatibility conditions between them. The categories constructed in Section 3 are really just a way of writing down all these compatibility conditions explicitly. The main point of doing this is the definition at the end of Section 3, where we define (A, H, S) vertex algebras to be the commutative rings in these categories. Here A is a suitable additive category (for example the category of modules over a commutative ring), H is a suitable bialgebra in A (and can be thought of as a sort of group ring of the group of automorphisms of spacetime), and S is something that controls the sort of singularities we allow.

One of the main differences between the (A, H, S) vertex algebras defined in Section 3 and previous definitions is as follows. Vertex algebras as usually defined consist of a space $V(1)$ with some extra operations, whose elements can be thought of a fields depending one one spacetime variable. On the other hand (A, H, S) vertex algebras include spaces $V(1, 2, \ldots, n)$ which can be thought of as fields depending on n spacetime variables for all n. The lack of these fields in several variables seems to be one reason why classical vertex algebras are so hard to handle: it is necessary to reconstruct these fields, and there seems to be no canonical way to do this. However if these fields are given in advance then a lot of these technical problems just disappear.

Section 4 puts everything together to construct many examples of vertex algebras. The main theorem of this paper is Theorem 4.2, which shows how to construct a vertex algebra from a singular bicharacter of a commutative and cocommutative bialgebra. As examples, we show that the usual vertex algebra of an even lattice can be constructed like this from the Hopf algebra of a multiplicative algebraic group, and the vertex algebra of a (generalized) free quantum field theory can be constructed in the same way from the Hopf algebra of an additive algebraic group. (This shows that the vertex algebra of a lattice is in some sense very close to a free quantum field theory: they have the same relation as multiplicative and additive algebraic groups.)

The vertex algebras we construct in this paper do not at first sight look much like classical vertex algebras: they seem to be missing all the structure such as vertex operators, formal power series, contour integration, operator product expansions, and so on. We show that all this extra structure can be reconstructed from the more elementary operations we provide for vertex algebras. For example, the usual locality property of vertex operators follows from the fact that we define vertex algebras as *commutative* rings in some category.

All the machinery in Sections 2 and 3 has been set up so that it generalizes trivially to quantum vertex algebras and higher dimensional analogues of vertex algebras. For example, we define quantum vertex algebras to be braided (rather than commutative) rings in a certain category, and we can instantly construct many examples of them from non-symmetric bicharacters of bialgebras. By changing a certain bialgebra H in the construction, we immediately get the "vertex G algebras" of [B98], which have the same relation to higher dimensional quantum field theories that vertex algebras have to one dimensional quantum field theories.

Finally in Section 5 we list some open problems and topics for further research.

Some related papers are [F-R] and [E-K], which give alternative definitions of quantum vertex algebras. These definitions are not equivalent to the ones in this paper, but define concepts that are closely related (at least in the case of 1 dimensional spacetime) in the sense that the interesting examples for all definitions should correspond. Soibelman has introduced other foundations for quantum vertex algebras, which seem to be related to this paper. There is also a preprint [B-D] which defines vertex algebras as commutative rings or Lie algebras in suitable multilinear categories. (Soibelman pointed out to me that multi categories seem to have been first introduced by Lambek in [L].) It might be an interesting question to study the relationship of this paper to [B-D]. One major difference is that the paper [B-D] extends the genus 0 Riemann surfaces that appear in vertex algebra theory to higher genus Riemann surfaces, while in this paper we extend them instead to higher dimensional groups.

I would like to thank S. Bloch, I. Grojnowski, J. M. E. Hyland, and Y. Soibelman for their help.

Notation

A	An additive symmetric tensor category.
C	A symmetric tensor category, with tensor product $\cup$; usually Fin or $Fin^{\not\equiv}$.
Δ	The coproduct of a bialgebra, or a propagator.
$D^{(i)}$	An element of the formal group ring of the one dimensional additive formal group.
η	The counit of a bialgebra.
Fin	The category of finite sets.
$Fin^{\not\equiv}$	The category of finite sets with an inequivalence relation.
Fun	A functor category.
H	A cocommutative bialgebra in A.
I, J	Finite sets.
L	An integral lattice.

M	A commutative ring in A, or a commutative cocommutative bialgebra.
r	A bicharacter.
R	A commutative ring or an R-matrix.
R-mod	The category of R-modules.
S	A commutative ring in some category, especially $Fun(C, A, T^*(H))$.
S^*	A symmetric algebra.
T	A cocommutative Hopf algebra in $Fun(C^{op}, A)$.
T_*, T^*	$T_*(M)(I) = \bigotimes_{i \in I} M$, $T^*(H)(I) = \bigotimes_{i \in I} H$. See Definition 3.3.
U, V	Objects of $Fun(C, A, T^*(H), S)$.

§2. Twisted group rings

We let R be any commutative ring. Recall that a bialgebra is an algebra with a compatible coalgebra structure, with the coproduct and counit denoted by Δ and η, and a Hopf algebra is a bialgebra with an antipode. If a is an element of a coalgebra then we put $\Delta(a) = \sum a' \otimes a''$.

Recall from the introduction that any bicharacter r of a group L can be used to define a twisted group ring. We now extend this idea from group rings $R[L]$ to cocommutative bialgebras.

Definition 2.1. Suppose that M and N are bialgebras over R and S is a commutative R-algebra. Then we define a bimultiplicative map from $M \otimes N$ to S to be a linear map $r : M \otimes N \to S$ such that

$$r(1 \otimes a) = \eta(a), \qquad r(a \otimes 1) = \eta(a),$$
$$r(ab \otimes c) = \sum r(a \otimes c')r(b \otimes c''),$$
$$r(a \otimes bc) = \sum r(a' \otimes b)r(a'' \otimes c),$$

where $\Delta(a) = \sum a' \otimes a''$, $\Delta(c) = \sum c' \otimes c''$, and η is the counit of M or N. We define an S-valued bicharacter of M to be a bimultiplicative map from $M \otimes M$ to S. We say the bicharacter r is symmetric if $r(a \otimes b) = r(b \otimes a)$ for all $a, b \in M$.

The S-valued bicharacters form a monoid, which is commutative if M is co-commutative. The identity bicharacter is defined by $r(a \otimes b) = \eta(a) \otimes \eta(b)$, and the product rs of two bicharacters r and s is given by

$$rs(a \otimes b) = \sum r(a' \otimes b')s(a'' \otimes b'').$$

If M is a Hopf algebra with antipode s then any S-valued bicharacter r has an inverse r^{-1} defined by

$$r^{-1}(a \otimes b) = r(s(a) \otimes b)$$

so the S-valued bicharacters form a group.

Example 2.2. Suppose that $M = R[L]$ is the group ring of a group L, considered as a bialgebra in the usual way (with $\Delta(a) = a \otimes a$ for $a \in L$). Any R^*-valued bicharacter of L can be extended to a linear function from $M \otimes M$ to R, and this is an R-valued bicharacter of M. This identifies the bicharacters of the group L with the bicharacters of its group ring M.

In order to define quantum vertex algebras we need a generalization of commutative rings, called braided rings. The idea is that we should be able to write $ab = \sum b_i a_i$ for suitable a_i and b_i related in some way to a and b. For example, for a commutative ring we would have $a_i = a$, $b_i = b$. The definition of the elements a_i and b_i is given in terms of an R-matrix with $R(a \otimes b) = \sum a_i \otimes b_i$, where an R-matrix is defined as follows.

Definition 2.3. An R-matrix for a ring M with multiplication map $m : M \otimes M \to M$ in an additive symmetric tensor category consists of a map $R : M \otimes M \to M \otimes M$ satisfying the following conditions.

1. (R is compatible with 1.) $R(1 \otimes a) = 1 \otimes a$, $R(a \otimes 1) = a \otimes 1$.
2. (R is compatible with multiplication.) $m_{23}R_{12}R_{13} = R_{12}m_{23} : M \otimes M \otimes M \to M \otimes M$ and $m_{12}R_{23}R_{13} = R_{13}m_{12} : M \otimes M \otimes M \to M \otimes M$.
3. (Yang-Baxter equation.) $R_{12}R_{13}R_{23} = R_{23}R_{13}R_{12}$.

Here R_{13} is R restricted to the first and third factors of $M \otimes M \otimes M$, and so on.

Definition 2.4. A braided ring M in an additive symmetric tensor category is a ring M with an R-matrix R such that

$$mR = m\tau : M \otimes M \longmapsto M$$

(where $\tau : a \otimes b \mapsto b \otimes a$ is the twist map and $m : a \otimes b \mapsto ab$ is the product).

Example 2.5. Suppose that M is $Z/2Z$ graded as $M = M_0 \oplus M_1$, and define R by $R(a \otimes b) = (-1)^{\deg(a)\deg(b)} a \otimes b$. Then M is a braided ring with R-matrix R if and only if M is a super commutative ring.

Lemma/Definition 2.6. *Suppose that r is an R-valued bicharacter of a commutative cocommutative bialgebra M. Define a new multiplication $\circ$ on M by*

$$a \circ b = \sum a'b'r(a'' \otimes b'')$$

(where $\Delta(a) = \sum a' \otimes a''$, $\Delta(b) = \sum b' \otimes b''$). Then this makes M into a ring, called the twisting of M by r. If r is symmetric then the twisting of M by r is commutative. If r is invertible (which is true whenever M is a Hopf algebra) then the twisting of M by r is a braided ring.

Proof. The element 1 is an identity for twisting of M by r because R is compatible with 1. The twisting is an associative ring because R satisfies the Yang-Baxter equation and is compatible with multiplication. It is easy to check that the twisting is commutative if r is symmetric and M is commutative. Finally we have to check that M has an R matrix if r is invertible. Define a bicharacter r' by

$$r'(a \otimes b) = \sum r(a' \otimes b')r^{-1}(b'' \otimes a'').$$

We define the R matrix by

$$R(a \otimes b) = a' \otimes b'r'(b'' \otimes a'')$$

where $\Delta(a) = \sum a' \otimes a''$, and $\Delta(b) = \sum b' \otimes b''$. It is easy to check that this satisfies the conditions for an R matrix for the twisting of M by r. This proves Lemma 2.6.

Example 2.7. Suppose that L is a free abelian group or free abelian monoid with a basis $\alpha_1, \ldots, \alpha_n$, and suppose that we are given elements $r(\alpha_i, \alpha_j) \in S^*$ for some commutative R-algebra S. We write e^α for the element of the group ring of L corresponding to the element $\alpha \in L$. Then we can extend r to a unique S-valued bicharacter of the ring $M = R[L]$ by putting

$$r\Big(\prod_i e^{m_i\alpha_i} \otimes \prod_j e^{n_j\alpha_j}\Big) = \prod_{1 \le i,j \le n} r(\alpha_i, \alpha_j)^{(m_i, n_j)}.$$

Example 2.8. Suppose that S is a commutative R algebra and that Φ is a free R-module, considered as an abelian Lie algebra. We let M be the universal enveloping algebra of Φ (in other words the symmetric algebra of Φ), so M is a commutative cocommutative Hopf algebra. Suppose that r is any linear map from $\Phi \otimes \Phi$ to S. Then we can extend r an S-valued bicharacter of M by putting

$$r(\phi_1 \cdots \phi_m \otimes \phi'_1 \cdots \phi'_n) = \begin{cases} 0, & \text{if } m \neq n, \\ \displaystyle\sum_{\sigma \in S_m} \prod_{i=1}^{m} r(\phi_i \otimes \phi'_{\sigma(i)}), & \text{if } m = n, \end{cases}$$

for $\phi_i, \phi'_i \in \Phi$ (where S_m is the symmetric group of permutations of $1, 2, \ldots, m$).

Example 2.9. Suppose that r is any bicharacter of a cocommutative bialgebra M. We can define an R-matrix for M by putting

$$R(a \otimes b) = \sum a' \otimes b' r(a'' \otimes b'').$$

If R is any R-matrix for a ring M we can define a new associative multiplication on M by putting

$$a \circ b = mR(a \otimes b) : M \otimes M \longmapsto M$$

where $m : M \otimes M \to M$ is the old multiplication. The composition of these two operations is just the twisting of M by r.

In the rest of this section we describe the construction of universal rings acted on by bialgebras, which we will need for the construction of vertex algebras. These universal rings can be thought of as something like the coordinate rings of function spaces or gauge groups.

Lemma/Definition 2.10. *Suppose that M is a commutative algebra over some ring and H is a cocommutative coalgebra. Then there is a universal commutative algebra $H(M)$ such that there is a map $h \otimes m \mapsto h(m)$ from $H \otimes M$ to $H(M)$ with*

$$h(mn) = \sum h'(m)h''(n), \quad h(1) = \eta(h).$$

If H is a bialgebra then H acts on the commutative ring $H(M)$. If M is a commutative and cocommutative bialgebra (or Hopf algebra) then so is $H(M)$.

Proof. The existence of $H(M)$ is trivial; for example, we can construct it by writing down generators and relations. Equivalently we can construct it as the quotient of the symmetric algebra $S(H \otimes M)$ by the ideal generated by the images of H and $H \otimes M \otimes M$ under the maps describing the relations. If H is a cocommutative bialgebra then it acts on $H(M)$ by $h_1(h_2(m)) = (h_1h_2)m$. If M has a coproduct $M \to M \otimes M$ then this induces a map $M \to H(M) \otimes H(M)$. As $H(M) \otimes H(M)$ is a commutative algebra acted on by H, this map extends to a map from $H(M)$ to $H(M) \otimes H(M)$ by the universal property of $H(M)$. It is easy to check that this coproduct makes $H(M)$ into a bialgebra. This proves Lemma 2.10.

The ring $H(M)$ has the following geometric interpretation. Pretend that H^* is the coordinate ring of a variety G. Then $\mathrm{Spec}(H(M))$ can be thought of as a sort of function space of all maps from G to $\mathrm{Spec}(M)$. If H is a bialgebra then we can pretend that it is the group ring of a group G, and the action of H on $H(M)$ then corresponds to the natural action of G on this function space induced by the action of G on itself by left multiplication. If in addition M is a cocommutative Hopf algebra, then $\mathrm{Spec}(M)$ is an affine algebraic group. The space $\mathrm{Spec}(H(M))$ is also an affine algebraic group, and can be thought of as the gauge group of all maps from G to $\mathrm{Spec}(M)$.

Example 2.11. Suppose that H is the supercommutative bialgebra with a basis 1, d, with $d^2 = 0$, $\Delta(d) = d \otimes 1 + 1 \otimes d$, such that d has odd degree. If M is any supercommutative ring then $H(M)$ is the ring of differential forms over M (where of course we replace "commutative" by "supercommutative" in Lemma 2.10).

Example 2.12. Suppose that M is a polynomial algebra $R[\phi_1, \ldots, \phi_n]$. Let H be the commutative cocommutative Hopf algebra over R with basis $D^{(i)}$ for $i \geq 0$, where $D^{(i)}D^{(j)} = \binom{i+j}{i}D^{(i+j)}$ and $\Delta(D^{(i)}) = \sum_j D^{(j)} \otimes D^{(i-j)}$. (We can think of H as the formal group ring of the one dimensional additive formal group. If R contains the rational numbers then $D^{(i)} = D^i/i!$ (where $D = D^{(1)}$) and H is just the universal enveloping algebra $R[D]$ of a one dimensional Lie algebra.) Then $H(M)$ is the ring of polynomials in the variables $D^{(i)}(\phi_j)$ for $i \geq 0$, $1 \leq j \leq n$. More generally, if we take M to be a symmetric algebra $S^*(\Phi)$ for an R-module Φ, then $H(S^*(M)) = S^*(H \otimes \Phi)$.

Example 2.13. Suppose that L is a lattice and $R[L]$ its group ring and suppose that H is the formal group ring of the one dimensional additive group, as in Example 2.12. Then $H(R[L])$ is the module underlying the vertex algebra of the lattice L. If instead we take H to be the polynomial ring $R[D]$ (with $\Delta(D) = D \otimes 1 + 1 \otimes D$) then $H(R[L])$ is isomorphic to the tensor product $R(L) \otimes S^*(L(1) \oplus L(2) \oplus \cdots)$ of the group ring $R[L]$ and the symmetric algebra of the sum of an infinite number of copies $L(n)$ of $L \otimes R$. This tensor product is also commonly used to construct the vertex algebra of a lattice. If R contains the rational numbers then it is equivalent to the first construction because $R[D]$ is then the same as the H defined in 2.12. However in non-zero characteristics it does not work quite so well; for example, we cannot define formal contour integrals as in Example 4.7, because this requires divided powers of D.

We now show that bicharacters of M are more or less the same as $H \otimes H$-invariant bicharacters of $H(M)$.

Lemma 2.14. *Suppose that M and N are bialgebras and r is a bimultiplicative map from $M \otimes N$ to S, where S is a commutative algebra acted on by the bialgebra H. Then r extends uniquely to a H invariant bimultiplicative map from $H(M) \otimes N$ to S.*

Proof. By adjointness we get an algebra homomorphism from M to the algebra $Hom(N, S)$ of linear maps from the coalgebra N to the algebra S. By the universality property of $H(M)$ this extends uniquely to an H invariant homomorphism from $H(M)$ to $Hom(N, S)$, which by adjointness gives a map from $H(M) \otimes N$ to S such that $r(m_1 m_2 \otimes n) = \sum r(m_1 \otimes n')r(m_2 \otimes n'')$ (where $\Delta(n) = \sum n' \otimes n''$). To finish the proof we have to check that $r(m \otimes n_1 n_2) = \sum r(m' \otimes n_1)r(m'' \otimes n_2)$. The set of m with this property contains M because by assumption r is bimultiplicative on $M \otimes N$. It is also easy to check that it is closed under multiplication and under the action of H. Therefore it contains the smallest H-invariant subalgebra of $H(M)$ containing M, which is the whole of $H(M)$. This proves Lemma 2.14.

Lemma 2.15. *Suppose that H is a cocommutative bialgebra and S is a commutative algebra acted on by $H \otimes H$. Suppose that M is a commutative and cocommutative bialgebra with an S-valued bicharacter r. Then r extends uniquely to a $H \otimes H$-invariant S-valued bicharacter $r : H(M) \otimes H(M) \to S$ of $H(M)$.*

Proof. We apply Lemma 2.14 to get a bimultiplicative H-invariant map from $H(M) \otimes M$ to S. Then we apply Lemma 2.14 again to get a bimultiplicative $H \otimes H$-invariant map from $H(M) \otimes H(M)$ to S. This proves Lemma 2.15.

§3. Construction of some categories

In this section we define a category $Fun(Fin^{\not\equiv}, A, H, S)$ in which we can carry out the "twisted group ring" construction in order to produce vertex algebras. The definition of this category is strongly motivated by classical and quantum field theory, and commutative rings in this category are formally quite similar to quantum field theories.

In the rest of this paper we fix an additive tensor category A that is cocomplete and such that colimits commute with tensor products. (In fact we do not need all colimits in A; it would be sufficient for most applications to assume that A has countable colimits.) For example, A could be the category R-mod of modules over a commutative ring. Note that most of the constructions and definitions of Section 2 work for any category A with the properties above.

Definition 3.1. We define Fin to be the category of all finite sets, with morphisms given by functions. We define $Fin^{\not\equiv}$ to be the category whose objects are finite sets with an equivalence relation $\equiv$, and whose morphisms are the functions f preserving inequivalence; in other words, if $f(a) \equiv f(b)$ then $a \equiv b$. We define $\cup$ on Fin and $Fin^{\not\equiv}$ to be the disjoint union (where in $Fin^{\not\equiv}$, elements of I and J in the disjoint union $I \cup J$ are inequivalent). This makes Fin and $Fin^{\not\equiv}$ into (non-additive) symmetric tensor categories.

We will write objects of $Fin^{\not\equiv}$ by using colons to separate the equivalence classes.

We could replace Fin and $Fin^{\not\equiv}$ by smaller equivalent categories; for example we could restrict the objects of Fin to be the finite sets of the form $\{1, 2, \dots, n\}$.

Note that $\cup$ is a coproduct in Fin but is not a coproduct in $Fin^{\not\equiv}$; in fact, $Fin^{\not\equiv}$ does not have coproducts. For example the coproduct of a one point set and a two point set with two equivalence classes does not exist.

Definition 3.2. If C is a category we define the category $Fun(C, A)$ to be the category of functors V from C to A. The category $Fun(C, A)$ is additive and has a symmetric tensor product given by the pointwise tensor product $(U \otimes V)(I) = U(I) \otimes V(I)$.

In applications the category C will be one of Fin, $Fin^{\not\equiv}$, or their opposite categories Fin^{op}, $Fin^{\not\equiv op}$.

Definition 3.3. Suppose that M is any commutative ring in A. We define $T_*(M)$ in $Fun(Fin, A)$ by $T_*(M)(I) = \bigotimes_{i \in I} M$, where the action of $T_*(M)$ on morphisms of Fin is induced in the obvious way by the product and unit of M. (For example, if f is the morphism from $\{1, 2\}$ to itself with $f(1) = f(2) = 2$, then $T_*(M)(f)$ takes $x_1 \otimes x_2$ to $1 \otimes x_1 x_2$.) If H is a cocommutative coalgebra in A then we define $T^*(H)$

in $Fun(Fin^{op}, A)$ in a similar way, using the coproduct and counit of H to define the action of $T^*(H)$ on morphisms.

Example 3.4. If M is a commutative ring in A then $T_*(M)$ is a commutative ring in $Fun(Fin, A)$. If in addition M is a commutative cocommutative bialgebra then so is $T_*(M)$.

Example 3.5. If V is a commutative ring in $Fun(Fin^{\not\equiv}, A)$ then we can think of $V(I)$ as the space of (nonsingular) quantum fields $\phi(x_1, x_2, \ldots)$ depending on $|I|$ spacetime variables.

The space of fields in one spacetime variable is acted on by the group of automorphisms G of spacetime, and similarly the space of fields of $|I|$ spacetime variables is acted on by $|I|$ commuting copies of G. We now add a similar structure to the objects of $Fun(Fin, A)$. It is convenient to use a cocommutative bialgebra H instead of a group G; we can think of this bialgebra H as analogous to the group ring of the automorphisms of spacetime (or maybe to the universal enveloping algebra of the Lie algebra of infinitesimal automorphisms of spacetime).

Definition 3.6. Suppose that T is a cocommutative bialgebra in $Fun(C^{op}, A)$. (In applications, T will be of the form $T^*(H)$ for a cocommutative bialgebra of A.) We define a T module in $Fun(C, A)$ to be an object V of $Fun(C, A)$ such that $V(I)$ is a module over $T(I)$ for all I and such that $f_*(f^*(g)(v)) = g(f_*(v))$ for $v \in V(I)$, $g \in T(J)$, $f : I \to J$. The action of T on the tensor product of two T modules is defined in the usual way using the coalgebra structure of the $T(I)$'s. We define $Fun(C, A, T)$ to be the additive symmetric tensor category of T modules in $Fun(C, A)$.

Example 3.7. Suppose that V is any commutative ring in A acted on by the cocommutative bialgebra H. Then $T_*(V)$ is a commutative ring in $Fun(Fin, A, T^*(H))$.

Recall that we can define the category of modules over any commutative ring in any additive symmetric tensor category, and it is again an additive symmetric tensor category.

Definition 3.8. Suppose that T is a cocommutative bialgebra in $Fun(C^{op}, A)$ and suppose that S is a commutative ring in $Fun(C, A, T)$. We define $Fun(C, A, T, S)$ to be the additive symmetric tensor category of modules over S.

Example 3.9. Suppose that we define S by letting $S(I)$ be the smooth functions depending on $|I|$ variables in spacetime. Then we would expect a field theory to be a module over S because we should be able to multiply a field by a smooth function to get a new field.

Commutative rings in $Fun(Fin, A, T^*(H), S)$ as defined above behave rather like classical field theories, or at least they have most of their formal properties. However quantum field theories do not fit into this framework. The problem is that in quantum field theory it is no longer true that the product of two nonsingular fields is a nonsingular field. For example, a typical formula in free quantum field theory is

$$\phi(x_1)\phi(x_2) =: \phi(x_1)\phi(x_2) :+ \Delta(x_1 - x_2)$$

where the propagator $\Delta(x)$ usually has a singularity at $x = 0$. In particular if we take $x_1 = x_2$ we find that the product of two fields depending on x_1 is not defined. Instead, we can take the product of two fields depending on different variables x_1 and x_2, and it lies in the space $V(1:2)$ of fields that are defined whenever x_1 and x_2 are "apart" in some sense.

The category $Fun(C, A, T^*(H), S)$ has a natural tensor product $\otimes$ which can be used to define multilinear maps. We will now define a new tensor product in $Fun(C, A, T^*(H), S)$ by defining a new concept of multilinear maps, called singular multilinear maps. We assume that C is a symmetric tensor category (not necessarily additive) with the tensor product denoted by $\cup$. (As the notation suggests, this will often be some sort of disjoint union.)

Definition 3.10. We let T be a cocommutative bialgebra in $Fun(C^{op}, A)$, and we let S be a commutative ring in $Fun(C, A, T)$. Suppose that $U_1, U_2, \ldots$ and V are objects of $Fun(C, A, T, S)$. We define a singular multilinear map from $U_1, U_2, \ldots$ to V to be a set of maps from $U_1(I_1) \otimes_A U_2(I_2) \cdots$ to $V(I_1 \cup I_2 \cdots)$ for all $I_1, I_2, \ldots \in C$, satisfying the following conditions.

1. The maps commute with the action of T.
2. The maps commute with the actions of $S(I_1), S(I_2), \ldots$.
3. If we are given any morphisms from I_1 to I_1', I_2 to I_2', ..., then the following diagram commutes:

$$\begin{array}{ccc} U_1(I_1) \otimes U_2(I_2) \cdots & \longrightarrow & V(I_1 \cup I_2 \cdots) \\ \downarrow & & \downarrow \\ U_1(I_1') \otimes U_2(I_2') \cdots & \longrightarrow & V(I_1' \cup I_2' \cdots) \end{array}$$

As A is co-complete and co-limits commute with taking tensor products the singular multilinear maps are representable, so we define the "singular tensor products" $U_1 \odot U_2 \cdots$ to be the objects representing the singular multilinear maps. It is possible to write down an explicit formula for these singular tensor products as follows.

$$(U_1 \odot U_2 \cdots)(I) = \varinjlim_{I_1 \cup I_2 \cdots \mapsto I} (U_1(I_1) \otimes_A U_2(I_2) \cdots) \otimes_{S(I_1) \otimes S(I_2) \cdots} S(I)$$

where the limit is a direct limit taken over the following category. The objects $I_1 \cup I_2 \cdots \mapsto I$ of the category consist of objects $I_1, I_2, \ldots$ of C together with a morphism from $I_1 \cup I_2 \cdots$ to I. A morphism from $I_1 \cup I_2 \cdots \mapsto I$ to $I_1' \cup I_2' \cdots \mapsto I$ consists of morphisms from I_1 to I_1', I_2 to I_2', ..., making the following diagram commute:

$$\begin{array}{ccc} I_1 \cup I_2 \cdots & \longrightarrow & I \\ \downarrow & & \| \\ I_1' \cup I_2' \cdots & \longrightarrow & I. \end{array}$$

J. M. E. Hyland told me that the product $\odot$ is similar to the "Day product" in category theory. The construction of $\odot$ can be extended to the case when C is a "symmetric multi-category" rather than a symmetric tensor category. Soibelman remarked that the conditions for V to be an algebra for $\odot$ are similar to the conditions for the functor V from C to A to be a functor of tensor categories.

Example 3.11. Suppose that $\cup$ is a coproduct in C; for example, we could take C to be Fin and $\cup$ to be disjoint union. Then singular tensor products are the same as pointwise tensor products. In later examples we will take C to be $Fin^{\not\equiv}$ and $\cup$ to be disjoint union, which is not a coproduct in $Fin^{\not\equiv}$.

The two tensor products $\odot$ and $\otimes$ are related in several ways, as follows. There is a canonical morphism from $U \odot V$ to $U \otimes V$, so that any ring is automatically a singular ring. Also there is a canonical "interchange" morphism

$$(U \otimes V) \odot (W \otimes X) \longrightarrow (U \odot W) \otimes (V \odot X).$$

(Unlike the case of the interchange map for natural transformations, this interchange map is not usually an isomorphism.) The interchange map can be used to show that if U and V are singular rings then so is $U \otimes V$.

We define singular rings, singular Lie algebras, and so on, in $Fun(C, A, T, S)$ to be rings, Lie algebras, and so on using the singular tensor product. We define singular bialgebras a little bit differently: the product uses the singular tensor product, but the coproduct uses the pointwise tensor product $\otimes$. Note that for this to make sense we need to know that the pointwise tensor product of two singular algebras is a singular algebra; see the paragraph above. In general, we should use the pointwise tensor product $\otimes$ for "coalgebra" structures, and the singular tensor product $\odot$ for "algebra" structures.

If S is a commutative ring in $Fun(Fin^{\not\equiv}, A, T^*(H))$ then by restriction it is also a commutative ring in $Fun(Fin, A, T^*(H))$ (using the functor which gives any finite set the equivalence relation where all elements are equivalent.) We can embed the category $Fun(Fin, A, T^*(H), S)$ into $Fun(Fin^{\not\equiv}, A, T^*(H), S)$ by defining $V(I_1 : I_2 \cdots) = V(I_1 \cup I_2 \cdots) \otimes_{S(I_1) \otimes S(I_2) \cdots} S(I_1 : I_2 \cdots)$ for $I_1, I_2, \ldots \in Fin$. In particular singular multilinear maps are defined in $Fun(Fin, A, T^*(H), S)$. (Note that singular tensor products representing singular multilinear maps do not usually exist in $Fun(Fin, A, T^*(H), S)$, though they do exist in the larger category $Fun(Fin^{\not\equiv}, A, T^*(H), S)$.)

The main point of all this category theory is the following definition:

Definition 3.12. Suppose that A is an additive symmetric tensor category, H is a cocommutative bialgebra in A, and S is a commutative ring in $Fun(Fin^{\not\equiv}, A, T^*(H))$. We define an (A, H, S) vertex algebra to be a singular commutative ring in $Fun(Fin, A, T^*(H), S)$. We define a quantum (A, H, S) vertex algebra to be a singular braided ring in $Fun(Fin, A, T^*(H), S)$.

Soibelman remarked that all the examples of quantum (A, H, S) vertex algebras in this paper have the extra property that the R matrix satisfies $R_{12}R_{21} = 1$, so perhaps this condition should be added to the definition of a quantum (A, H, S) vertex algebra.

Note that the vertex algebra is in $Fun(Fin, A, T^*(H), S)$ rather than $Fun(Fin^{\not\equiv}, A, T^*(H), S)$, although we can of course embed the former category in the latter if we wish. The reason for using $Fun(Fin, A, T^*(H), S)$ rather than $Fun(Fin^{\not\equiv}, A, T^*(H), S)$ is that we wish to have control over the connection between (say) $V(1, 2)$ and $V(1 : 2)$.

§4. Examples of vertex algebras

In this section we construct some examples of (A, H, S) vertex algebras by applying the twisted group ring construction of Section 2 to the categories constructed in Section 3. We also show how these are related to classical vertex algebras.

Lemma 4.1. *Suppose that r is an $H \otimes H$-invariant $S(1 : 2)$-valued bicharacter of a commutative cocommutative bialgebra $H(M)$ in A. Then H can be extended to a singular bicharacter of $T_*(H(M))$, which we also denote by r.*

Proof. We define r by

$$r\Big(\bigotimes_{i\in I} a_i \otimes \bigotimes_{j\in J} b_j\Big) = \sum \prod_{i\in I}\prod_{j\in J} r(a_i^{(j)} \otimes b_j^{(i)})$$

where $\Delta^{|J|-1}(a_i) = \sum \bigotimes_{j\in J} a_i^{(j)}$, $\Delta^{|I|-1}(b_j) = \sum \bigotimes_{i\in I} b_j^{(i)}$, and $r(a_i^{(j)}\otimes b_j^{(i)})$ is considered as an element of $S(I\cup J)$ using the obvious map from $S(i:j)$ to $S(I\cup J)$. Some routine checking then proves Lemma 4.1.

The following theorem is the main theorem of this paper. It shows how to construct many examples of (A, H, S) vertex algebras, by giving a sort of generalization of the construction of the vertex algebra of a lattice.

Theorem 4.2. *Suppose that H is a cocommutative bialgebra in A and S is a commutative ring in $Fun(Fin^{\not\equiv}, A, T^*(H))$. Assume that we are given an $S(1:2)$-valued bicharacter r of a commutative and cocommutative bialgebra M in A. The bicharacter r of M extends to a bicharacter of $T_*(H(M))$ as in Lemmas 2.15 and 4.1, which we also denote by r. Then the twisting of $T_*(H(M))$ by r is a quantum (A, H, S) vertex algebra if r is invertible, and is an (A, H, S) vertex algebra if r is symmetric.*

Proof. By Lemma 2.10 and Example 3.4, $T_*(H(M))$ is a commutative cocommutative bialgebra in $Fun(Fin, A, T^*(H), S)$. By Lemmas 2.15 and 4.1 the bicharacter r extends to a singular bicharacter of $T_*(H(M))$ with values in S. By Lemma 2.6 (extended to additive tensor categories) the twisting of $T_*(H(M))$ by r is a braided ring if r is invertible, and is a commutative ring if r is symmetric. Theorem 4.2 now follows from the Definition 3.12 of (quantum) (A, H, S) vertex algebras.

The following theorem describes the relation between the (A, H, S) vertex algebras of this paper, and ordinary vertex algebras.

Theorem 4.3. *Suppose we take H to be the formal group ring of the one dimensional additive formal group, as in Example 2.12. Define S by $S(I)$ = the R-algebra generated by $(x_i - x_j)^{\pm 1}$ for i and j not equivalent (so $S = R$ if all elements of I are equivalent). If V is a $(R\text{-mod}, H, S)$ vertex algebra, then $V(1)$ is an ordinary vertex algebra over the ring R.*

Proof. For every element u_1 of $V(1)$ we have to construct a vertex operator $u_1(x_1)$ taking $V(1)$ to $V(1)[[x_1]][x_1^{-1}]$. We do this as follows.

If $u_2 \in V(2)$ then $u_1u_2 \in V(1:2) = V(1,2) \otimes S(1:2) = V(1,2)[(x_1 - x_2)^{\pm 1}]$. There is a map from $V(1,2)$ to $V(1)[[x_1, x_2]]$ taking w to the "Taylor series expansion" $\sum_{i,j} f_{12\to 1}(D_1^{(i)} D_2^{(j)} w) x_1^i x_2^j$. (Here $f_{12\to 1}$ is the map from $V(1,2)$ to $V(1)$ induced by the morphism of finite sets taking both 1 and 2 to 1, and D_1 and D_2 indicate the two different actions of H on $V(1,2)$.) This induces a map from $V(1,2)[(x_1 - x_2)^{\pm 1}]$ to $V(1)[[x_1, x_2]][(x_1 - x_2)^{-1}]$, and we denote the image of u_1u_2 under this map by $u_1(x_1)u_2(x_2)$. Then we define the vertex operator $u_1(x_1)$ by $u_1(x_1)u_2 = u_1(x_1)u_2(0) \in V(1)[[x_1]][x_1^{-1}]$.

This defines the vertex operators of elements of $V(1)$; now we have to check that they formally commute. We can define expressions like $u_1(x_1)u_2(x_2)u_3(x_3)\cdots \in V(1)[[x_1, \dots]][\prod(x_i - x_j)^{-1}]$ in the same way as above. The fact that V is commutative implies that $u_1(x_1)u_2(x_2)u_3(0) = u_2(x_2)u_1(x_1)u_3(0)$. This in turn implies that the vertex operators $u_1(x_1)$ and $u_2(x_2)$ commute in the sense that

$$(x_1 - x_2)^N (u_1(x_1)u_2(x_2) - u_2(x_2)u_1(x_1))u_3 = 0$$

for N a sufficiently large integer, depending on u_1 and u_2. So we have constructed commuting vertex operators for all elements of $V(1)$, and this can easily be used to show that $V(1)$ is a vertex algebra. This proves Theorem 4.3.

Example 4.4. Take L to be an even integral lattice. Choose a bicharacter c such that $c(\alpha, \beta) = (-1)^{(\alpha,\beta)} c(\beta, \alpha)$. (There are many ways to do this. For example we can choose a basis $\alpha_1, \alpha_2, \dots$ and define $c(\alpha_i, \alpha_j)$ to be 1 if $i \geq j$ and $(-1)^{(\alpha_i,\alpha_j)}$ if $i < j$.) Define a symmetric $R[(x_1 - x_2)^{\pm 1}]$-valued bicharacter r of L by

$$r(\alpha, \beta)(x_1, x_2) = c(\alpha, \beta)(x_1 - x_2)^{(\alpha,\beta)}.$$

If V is the $(R\text{-mod}, H, S)$ vertex algebra constructed in Theorem 4.2 with underlying object $T_*(H(R[L]))$ then $V(1)$ is just the usual vertex algebra of the even integral lattice L. If L is any integral lattice (not necessarily even) then we can do a similar construction with the following changes. We choose c so that $c(\alpha, \beta) = (-1)^{(\alpha,\beta)+(\alpha,\alpha)(\beta,\beta)} c(\beta, \alpha)$. The bicharacter r is no longer symmetric but is supersymmetric, so we end up with a vertex superalgebra rather than a vertex algebra.

Example 4.5. Now we write down some quantum deformations of Example 4.4. Let L be an even lattice as in Example 4.4, let q be an invertible element of the commutative ring R, and let A be the category of R modules. We define S by $S(I) = R$ if I has only one equivalence class, and $S(I)$ = the R-algebra generated by $(x_i - q^n x_j)$ for i and j not equivalent, n an integer, if I has more than 1 equivalence class. Choose a basis $\alpha_1, \ldots, \alpha_n$ for L and define r using Lemma 2.7 by putting

$$r(\alpha_i, \alpha_j) = c(\alpha, \beta) \prod_{k=1}^{(\alpha_i,\alpha_j)} (x_1 - q^{(\alpha_i,\alpha_j)-2k} x_2)$$

where c is the bicharacter of Example 4.4. By applying Theorem 4.2 we get a $(R\text{-mod}, H, S)$ quantum vertex algebra. We see that

$$(x_1 - q^{(\alpha_i,\alpha_j)} x_2) e^{\alpha_1}(x_1) e^{\alpha_2}(x_2) = (q^{(\alpha_i,\alpha_j)} x_1 - x_2) e^{\alpha_2}(x_2) e^{\alpha_1}(x_1).$$

This is similar to many of the formulas of statistical mechanics in the book [J-M].

Example 4.6. We show how to construct (A, H, S) vertex algebras corresponding to generalized free quantum field theories. Suppose that Φ is a module over a commutative ring in A and H is a commutative cocommutative bialgebra in A. Then any linear map Δ from $\Phi \otimes \Phi$ to $S(1:2)$ gives a quantum (A, H, S) vertex algebra as follows. Use Example 2.8 to extend r to a $S(1:2)$-valued bicharacter of the symmetric algebra M of Φ. Then use Theorem 4.2 to make $T_*(H(M))$ into a quantum (A, H, S) vertex algebra. If r is symmetric then this is a (A, H, S) vertex algebra, and is closely related to generalized free quantum field theories, at least when H is finite dimensional abelian. (To obtain analogues of free quantum field theories in odd dimensions or dimension 2 we should allow slightly more general sorts of singularities, such as half integral powers or logarithms of $(x_1 - x_2)^2$ rather than just poles.) The function r gives the propagator of free fields, and the Greens functions $\langle|\phi_1(x_1)\cdots\phi_n(x_n)|\rangle$ can be recovered as $\eta(\phi_1(x_1)\cdots\phi_n(x_n))$ where η is the counit of $H(M)$ and $\phi_1, \ldots, \phi_n$ are elements of Φ.

Take H to be the additive formal group of dimension d for some positive even integer d. If we take Φ to be a one dimensional free module over R spanned by an element ϕ and put $r(\phi \otimes \phi) = (\sum(x_{1,i} - x_{2,i})^2)^{1-d/2}$ then V is the "H vertex algebra of a free scalar field" constructed in [B98]. It is obvious that we can just write down many quantum deformation of this H vertex algebra just by varying r; for example, we could take $r(\phi \otimes \phi) = (\sum(x_{1,i} - qx_{2,i})^2)^{1-d/2}$.

Example 4.7. In the theory of vertex algebras we often get contour integrals such as

$$\int_{x_1} a_1(x_1) a_2(x_2) a_3(x_3)\, dx_1.$$

We will show how to define such contour integrals for (A, H, S) vertex algebras, where H and S are as in Theorem 4.3. Take $a_i \in V(i)$, where $V(i)$ can be identified with $V(1)$. We know that $a_1a_2a_3 \in V(1:2:3)$ using the multiplication of V. We also know that $V(1 : 2 : 3) = V(1,2,3)[(x_1 - x_2)^{\pm 1}, (x_2 - x_3)^{\pm 1}, (x_1 - x_3)^{\pm 1}]$, so we can write $a_1a_2a_3$ as a finite sum of terms of the form

$$a_{123}(x_1 - x_2)^i (x_1 - x_3)^j (x_2 - x_3)^k.$$

Next we expand $(x_1 - x_3)^j$ as a possibly infinite series

$$(x_1 - x_3)^j = \sum_{n \geq 0} \binom{j}{n} (x_1 - x_2)^n (x_2 - x_3)^{j-n}.$$

Finally we replace each term $a_{123}(x_1 - x_2)^i (x_2 - x_3)^k$ by

$$f_*(D_1^{(i)}(a_{123}))(x_2 - x_3)^k \in V(2:3)$$

where f is the function from $\{1,2,3\}$ to $\{2,3\}$ with $f(1) = f(2) = 2$, $f(3) = 3$. This algebraically defined contour integral has most of the properties one would expect. For example we have the identity

$$\begin{aligned}\int a_1(x_1)\,dx_1 \int a_2(x_2)\,dx_2 a_3 - \int a_2(x_2)\,dx_2 \int a_1(x_1)\,dx_1 a_3 \\ = \int \left(\int a_1(x_1)\,dx_1 a_2 \right)(x_2)\,dx_2 a_3\end{aligned}$$

which can be used to prove the usual vertex algebra identities. Of course this identity depends on the simple choice of H and S we made; for more complicated choices of H and S we will usually get more complicated identities. In particular contour integrals can be defined in terms of the more elementary operations of a (A, H, S) vertex algebra. One reason for using the bialgebra H with divided powers (see Example 2.12) rather then the universal enveloping algebra $R[D]$ is that the divided powers are needed to define the contour integrals.

Example 4.8. Take H as in Example 2.12, and let $S(I)$ be R if I has at most one equivalence class, and the ring generated by the elements $(x_i - q^n x_j)^{\pm 1}$ for $i \not\equiv j$ and I having more than 1 equivalence class. Then if V is a quantum (A, H, S) vertex algebra, we can think of $V(1)$ as being some sort of "quantum vertex algebra". We will not give a definition of quantum vertex algebras here, because the philosophy of this paper is that (quantum) vertex algebras should be replaced by (quantum) (A, H, S) vertex algebras. Several sets of axioms for quantum vertex algebras have been proposed by various authors in [E-K], [F-R].

Example 4.9. The (ordinary) tensor product of two (ordinary) vertex algebras is a vertex algebra. The analogue of this for (A, H, S) vertex algebras is trivial to prove: the pointwise tensor product of any two singular commutative rings in $Fun(Fin^{\neq}, A, H, S)$ is a singular commutative ring, and the pointwise tensor product of two objects of $Fun(Fin, A, T^*(H), S)$ is still in $Fun(Fin, A, T^*(H), S)$, so the pointwise tensor product of two (A, H, S) vertex algebras is an (A, H, S) vertex algebra. Note that the singular tensor product of two (A, H, S) vertex algebras is a singular commutative ring in $Fun(Fin^{\neq}, A, T^*(H), S)$, but need not be in $Fun(Fin, A, T^*(H), S)$, so the singular tensor product of two (A, H, S) vertex algebras need not be an (A, H, S) vertex algebra.

Example 4.10. We can obtain many variations of vertex algebras by changing H and S. For example we could take H to be the universal enveloping algebra of the Virasoro algebra to get things similar to "vertex operator algebras". If we take H to be the tensor product of two copies of the Virasoro algebra then (A, H, S) vertex algebras are closely related to conformal field theory and string theory. If we let H be the universal enveloping algebra of various superalgebras then we get (A, H, S) vertex algebras related to supersymmetry.

§5. Open problems

In this section we list some suggestions for further research.

Problem 5.1. Are there natural quantum deformations of other well known vertex algebras, such as the monster vertex algebra [B86], [F-L-M], the vertex algebra of the lattice $II_{25,1}$ [B86], [K97], and the vertex algebras of highest weight representations of affine Lie algebras and the Virasoro algebra [F-Z], [K97]? Etingof and Kazhdan [E-K] construct "quantum vertex operator algebras" corresponding to the vertex algebras of affine Lie algebras, and it seems likely that their construction could be extended to give examples satisfying the definitions in this paper. Frenkel and Jing [F-J] previously constructed vertex operators related to of quantum affine Lie algebras.

Problem 5.2. Ordinary vertex algebras can be used to construct many examples of generalized Kac-Moody algebras. Is there a relation between quantum vertex algebras and some sort of quantized generalized Kac-Moody algebras, possibly those defined in [K95]?

Problem 5.3. The similarity of the formulas in solvable lattice models in [J-M] and quantum vertex algebras suggests that there may be some relation between these subjects.

Problem 5.4. We have constructed vertex algebras from bicharacters of bialgebras that are both commutative and cocommutative. If a bialgebra is cocommutative but not commutative then the bicharacters are usually not all that interesting (for the much same reason that one dimensional characters of a non-abelian group are not usually interesting). However there are nontrivial examples of bicharacters of bialgebras that are neither commutative or cocommutative. Can these be used to construct some sort of vertex algebras?

Problem 5.5. Construct (R-mod$, H, S$) vertex algebras corresponding to the other standard examples of vertex algebras, such as the vertex algebras of affine and Virasoro algebras ([F-Z]), or the monster vertex algebra ([F-L-M]) or the vertex algebra of differential operators on a circle ([K97]).

Problem 5.6. Many of the constructions and definitions in Section 3 do not use the fact that the category A is additive. Is there any use for these constructions in the non-additive case?

Problem 5.7. Do these constructions for braided rather than symmetric tensor categories? In particular it should be possible to allow nonintegral powers of $x_i - x_j$, which often arise from non-integral lattices or from conformal field theory.

Problem 5.8. A cobraided Hopf algebra (as defined in in [K, Definition VIII. 5.1]) is a Hopf algebra with a bicharacter r with the extra property that $\mu^{op} = r * \mu * \bar{r}$. This suggests that it might be possible to replace commutative, cocommutative bialgebras by something more general, maybe cobraided bialgebras. In particular Theorem 4.2 should be extended to the case when M is cobraided rather than cocommutative.

Problem 5.9. Instead of twisting a group ring by a bicharacter, we can also twist it by a 2-cocycle (preferably normalized). We can define "multiplicative 2-cocycles" of arbitrary cocommutative bialgebras with values in any algebra S acted on by the bialgebra, and use these to construct more general twistings. We can also define multiplicative n-cochains, cocycles, and coboundaries, and use these to define multiplicative analogues $H^n(M, S^*)$ of cohomology groups. Note that the usual (additive) cohomology $H^n(M, S)$ of bialgebras depends only on the underlying associative algebra and the counit of M and on the module structure of S, and should not be confused with these multiplicative cohomology groups $H^n(M, S^*)$ that also depend on the coproduct of M and the algebra structure of S. Find some examples of vertex algebras constructed using singular 2-cocycles rather than singular bicharacters. There are many examples that can be constructed like this in a formal (and not very interesting) way from a perturbative quantum field theory.

Problem 5.10. It is possible to construct singular 2-cocycles which look formally similar to the Greens functions of perturbative quantum field theories. At the moment this just seems to be little more than a formal triviality, but may be worth investigating further.

Problem 5.11. I. Grojnowski and S. Bloch independently suggested replacing the Hopf algebra H of Example 4.4 by the formal group ring of the formal group of an elliptic curve. Over the rationals this makes no difference, but over finite fields or the integers we seem to get something different. The underlying space of the vertex algebra we get can be thought of as the coordinate ring of the gauge group of maps from a (formal) elliptic curve to an algebraic torus. The problem is to find a use for this construction!

Problem 5.12. Develop the theory of categories with two symmetric tensor products satisfying the conditions suggested in Section 3 (and maybe some others), and find more examples of them. Soibelman pointed out that Beilinson and Drinfeld [B-D] have some categories which have both a tensor product and a separate multilinear structure.

Problem 5.13. The study of orbifolds of vertex algebras (in other words, fixed subalgebras under finite automorphism groups) is notoriously hard (see [D-M] for example), though this ought to be an easy and natural operation. The difficulties appear to be caused partly by the fact that vertex algebras seem to have something missing from their structure. Does the theory of orbifolds for (A, H, S) vertex algebras (with their extra structure of fields of several spacetime variables) become any easier?

Problem 5.14. Soibelman suggested that the examples of associative algebras of automorphic forms in the meromorphic tensor category of [So, Theorem 8] might be some sort of (A, H, S) vertex algebras. These may be related to the algebras in [K96].

References

[B86] R. E. Borcherds, Vertex algebras, Kac-Moody algebras and the monster, Proc. Nat. Acad. Sci. U.S.A., **83** (1986), 3068–3071.

[B98] R. E. Borcherds, Vertex algebras, in "Topological field theory, primitive forms and related topics", (Kyoto, 1996), edited by M. Kashiwara, A. Matsuo, K. Saito and I. Satake, Progress Math., **160**, Birkhäuser Boston, Inc., Boston, MA., ISBN:0-8176-3975-6, 1998, pp. 35–77. preprint q-alg/9706008.

[B-D] A. Beilinson and V. Drinfeld, Chiral algebras I, preprint.

[D-M] C. Dong and G. Mason, On quantum Galois theory, Duke Math. J., no. **2** (1997), 305–321.

[E-K] P. Etingof and D. Kazhdan, Quantization of Lie bialgebras, V. preprint math. QA/9808121.

[F-R] E. Frenkel and N. Yu. Reshetikhin, Towards Deformed Chiral Algebras, submitted to the Proceedings of the Quantum Group Symposium at the XXIth International Colloquium on Group Theoretical Methods in Physics, Goslar (1996). preprint q-alg/9706023.

[F-J] I. B. Frenkel and N. H. Jing, Vertex representations of quantum affine algebras, Proc. Nat. Acad. Sci. U.S.A., no. **24** (1988), 9373–9377.

[F-L-M] I. B. Frenkel, J. Lepowsky and A. Meurman, "Vertex operator algebras and the Monster", Pure and Applied Mathematics, **134**, Academic Press, Inc., Boston, MA, 1988. ISBN:0-12-267065-5.

[F-Z] I. B. Frenkel and Y. Zhu, Vertex operator algebras associated to representations of affine and Virasoro algebras, Duke Math. J., no. **1** (1992), 123–168.

[J-M] M. Jimbo and T. Miwa, "Algebraic analysis of solvable lattice models", CBMS Regional Conference Series in Mathematics, **85**, Published for the Conference Board of the Mathematical Sciences, Washington, DC. by the American Mathematical Society, Providence, RI, 1995. ISBN:0-8218-0320-4.

[J] N. Jing, Quantum Z-algebras and representations of quantum affine algebras, Comm. Algebra, **28**, no. 2 (2000), 829–844. preprint math. QA/9806043.

[K97] V. Kac, "Vertex algebras for beginners", University Lecture Series, **10**, American Mathematical Society, Providence, RI, 1997. ISBN:0-8218-0643-2.

[K95] S.-J. Kang, Quantum deformations of generalized Kac-Moody algebras and their modules, J. Algebra, **175,** no. 3 (1995), 1041–1066.

[K96] M. M. Kapranov, Eisenstein series and quantum affine algebras, Algebraic geometry, **7**. J. Math. Sci., New York, **84**, no. 5 (1997), 1311–1360. preprint alg-geom/9604018.

[K] C. Kassel, "Quantum groups", Graduate Texts in Mathematics, **155**, Springer-Verlag, New York, 1995. ISBN:0-387-94370-6.

[L] J. Lambek, Deductive systems and categories, II, in "Standard constructions and closed categories", 1969, Category Theory, Homology Theory and their Applications, I (Battelle Institute Conference, Seattle, Wash., 1968, Vol. **1**), Springer, Berlin, pp. 76–122.

[So] Y. Soibelman, Meromorphic tensor categories, Internat. Math. Res. Notices, no. **19** (1999), 1067–1079. preprint q-alg/9709030. Meromorphic braided category arising in quantum affine algebras, preprint math/9901003.

D.P.M.M.S.
16 Mill Lane
Cambridge, CB2 1SB
England
`reb@dpmms.cam.ac.uk`

Current address:
University of California
Department of Mathematics
Evans Hall #3840
Berkeley, CA 94720-3840
U.S.A.
`reb@math.berkeley.edu`

Advanced Studies in Pure Mathematics 31, 2001
Taniguchi Conference on Mathematics Nara '98
pp. 75–91

Floer Homology and Gromov-Witten Invariant over Integer of General Symplectic Manifolds – Summary –

Kenji Fukaya[1] and Kaoru Ono[2]

Abstract.

In this article we give a summary of an improvement our earier result [FO2] on Arnold's conjecture about the number of periodic orbits of periodic Hamiltonian system. In [FO2], we gave an estimate in terms of Betti numbers. In this article, we include torsion coefficients. We also define an "integer part" of the Gromov-Witten invariant.

§1. Introduction

Let (X^{2n}, ω) be a compact symplectic manifold and $h : X \times S^1 \to \mathbb{R}$ be a smooth function. We put $h_t(x) = h(x,t)$. Let V_{h_t} be the Hamiltonian vector field generated by h_t. Let $\Phi_t : X \to X$ be the family of symplectic diffeomorphisms such that

$$\frac{d\Phi_t}{dt} = V_{h_t} \circ \Phi_t, \quad \Phi_0 = \mathrm{id}.$$

We assume that the graph $\mathrm{Graph}(\Phi_1)$ of $\Phi_1 \subset X \times X$ is transversal to the diagonal Δ_X. The intersection $\Delta_X \cap \mathrm{Graph}(\Phi_1)$ can be identified with the fixed point set $\mathrm{Fix}(\Phi_1)$ of Φ_1. Our main result is an estimate of the order of $\mathrm{Fix}(\Phi_1)$ in terms of the Betti numbers and the torsion coefficients of X.

We define the universal Novikov ring Λ by

$$\Lambda = \Big\{ \sum c_i T^{\lambda_i} \;\Big|\; c_i \in \mathbb{Z},\, \lambda_i \in \mathbb{R},\, \lim_{i\to\infty} \lambda_i = \infty \Big\}.$$

Received August 13, 1999.

[1]Partially supported by Grant in-Aid for Scientific Research no. 09304008 and no. 11874013.

[2]Partially supported by Grant in-Aid for Scientific Research no. 11440015.

Here T is a formal parameter. We remark that the modulo 2 Conley-Zehnder index μ of elements of $\mathrm{Fix}(\Phi_1)$ is well-defined (see [F]). We put (for $i \in \mathbb{Z}_2$),

$$CF_i(X,h) = \bigoplus_{p\in \mathrm{Fix}(\Phi_1),\, \mu(p)=i} \Lambda[p].$$

The main result explained in this article is the following theorem, which is a version of Arnold's conjecture [A1], [A2].

Theorem 1. *There exist homomorphisms $\partial_i : CF_i(X,h) \to CF_{i-1}(X,h)$ such that $\partial_i\partial_{i+1} = 0$ and*

$$\frac{\mathrm{Ker}\,\partial_i}{\mathrm{Im}\,\partial_{i+1}} \simeq \sum_{i\equiv k \bmod 2} H_k(M;\mathbb{Z}) \otimes_{\mathbb{Z}} \Lambda. \tag{1}$$

Remark 1. If we replace $\mathbb{Z}$ by $\mathbb{Q}$, Theorem 1 was proved by Fukaya-Ono [FO1, 2], Liu-Tian [LT], Ruan [R]. In case when X is semi-positive, Theorem 1 was proved by Hofer-Salamon [HS] and Ono [O]. (They are generalizations of celebrated results by Conley-Zehnder [CZ] and Floer [F].)

In this article, we show an outline of a proof of Theorem 1. The detail will appear elsewhere.

§2. A brief review of Floer homology and negative multiple cover problem

It is known to experts that, if one can define the fundamental chain over $\mathbb{Z}$ of the moduli space of pseudoholomorphic curves with appropriate properties, then we can prove Theorem 1. We first explain it briefly. Let J_X be an almost complex structure on X compatible with ω. We put

$$\mathrm{Orb}(h) = \left\{ \ell : S^1 \to X \;\middle|\; \frac{d\ell}{dt} = V_{h_t}(\ell(t)) \right\}$$

We can identify $\mathrm{Orb}(h)$ with $\mathrm{Fix}(\Phi_1)$. For $\ell_1, \ell_2 \in \mathrm{Orb}(h)$, we put

$$\widetilde{\mathcal{M}}(\ell_1,\ell_2) = \left\{ \varphi : \mathbb{R}\times S^1 \to X \;\middle|\; \begin{array}{l} \dfrac{\partial\varphi}{\partial\tau} = J_X\left(\dfrac{\partial\varphi}{\partial t} - V_{h_t}\right), \\ \lim\limits_{\tau\to-\infty} \varphi(\tau,t) = \ell_1(t), \\ \lim\limits_{\tau\to+\infty} \varphi(\tau,t) = \ell_2(t). \end{array} \right\}$$

$\mathbb{R}$ acts on $\widetilde{\mathcal{M}}(\ell_1,\ell_2)$ by the translation along $\mathbb{R}$ factor. Let $\mathcal{M}(\ell_1,\ell_2)$ be the quotient space. For $\varphi \in \widetilde{\mathcal{M}}(\ell_1,\ell_2)$, we define its energy by

$$E_h(\varphi) = \frac{1}{2}\int \left(\left\| \frac{\partial\varphi}{\partial\tau} \right\|^2 + \left\| \frac{\partial\varphi}{\partial t} - V_{h_t} \right\|^2 \right) dt d\tau.$$

We put

$$\mathcal{M}(\ell_1,\ell_2;E) = \{\varphi \in \mathcal{M}(\ell_1,\ell_2) \mid E_h(\varphi) = E\}.$$

Gromov's compactness theorem [G] implies that $\mathcal{M}(\ell_1,\ell_2;E)$ is nonempty only for $E = E_1, E_2, \ldots$ such that $0 = E_1 < E_2 < E_3 < \cdots$, $\lim E_i \to \infty$.

The virtual dimension of $\mathcal{M}(\ell_1,\ell_2;E)$ depends on the component. Let $\mathcal{M}(\ell_1,\ell_2;E;k)$ be the union of the components of virtual dimension k.

Suppose we have a "perturbation" of $\mathcal{M}(\ell_1,\ell_2;E;k)$ for $k = 0,1$ with the following properties.

(2.1) $\mathcal{M}(\ell_1,\ell_2;E;0)$ consists of finitely many points. Each point φ of $\mathcal{M}(\ell_1,\ell_2;E;0)$ is given an orientation $\epsilon_\varphi = \pm 1$.

(2.2) $\mathcal{M}(\ell_1,\ell_2;E;1)$ can be compactified to an oriented one dimensional manifold whose boundary is

$$\bigcup_{E'+E''=E} \bigcup_{\ell_3} \mathcal{M}(\ell_1,\ell_3;E';0) \times \mathcal{M}(\ell_3,\ell_2;E'';0).$$

We then put

$$\partial[\ell_1] = \sum_i \sum_{\ell_2} \sum_{[\varphi]\in\mathcal{M}(\ell_1,\ell_2;0;E_i)} \epsilon_\varphi T^{E_i}[\ell_2].$$

(2.1) implies that the coefficient of the right hand side belongs to Λ. Then (2.2) implies $\partial\partial = 0$. We need some more properties to show the isomorphism (1). We omit the discussion about it in this article.

There is a trouble to find a perturbed moduli space satisfying (2.1) and (2.2). The main problem is the equivariant transversality at infinity, which we recall very briefly here. (A bit more detailed summary is in the introduction of [FO2].)

Let us consider a divergent sequence $\varphi_i \in \mathcal{M}(\ell_1,\ell_2;E;1)$. One possibility of its "limit" is an element of $\mathcal{M}(\ell_1,\ell_3;E';0) \times \mathcal{M}(\ell_3,\ell_2;E'';0)$. This is the component of the boundary of a compactification of $\mathcal{M}(\ell_1,\ell_2;E;1)$ described in (2.2). However there is another possibility. Namely φ_i may "converege" to a map $\varphi\sharp(\psi\circ\pi)$. Here $\varphi \in \mathcal{M}(\ell_1,\ell_2;E';*)$, $\psi : S^2 \to M$ is a pseudoholomorphic map, and $\pi : S^2 \to S^2$ is a degree

k holomorphic map. ($E = E' + k([S^2] \cap \varphi^*\omega)$.) We assume also that $\psi(S^2)$ intersects with the image of φ. $\sharp$ denotes the connected sum. The trouble is especially serious in the case when $\varphi\sharp(\psi \circ \pi)$ has a nontrivial symmetry. If moreover $c^1(M) \cap \psi(S^2)$ is negative, we find that there is no perturbation, in the usual sense, to make $\varphi\sharp(\psi \circ \pi)$ transversal. This trouble is called the negative multiple cover problem. We studied it in [FO2], where we used a multivalued perturbation and hence we worked over rational coefficient. The purpose of this article is to explain an outline of a way to overcome this trouble without using rational coefficient.

§3. Period-doubling bifurcation and Stiefel-Whitney class

Let us describe a toy model which shows how the rational coefficient occurs in a natural way. In this toy model, we consider a moduli space of maps $S^1 \to Y$ in place of $\Sigma^2 \to X$. Let Y be the Möbius band $\mathbb{R} \times [0,1]/\!\sim$ where $(x, 1) = (f_\epsilon(x), 0)$ and $f_\epsilon(x)$ is a diffeomorphism of $\mathbb{R}$ such that $f_\epsilon(x) = -(1+\epsilon)x + x^3$ in a neighborhood of 0. We consider the vector field $V_\epsilon = \partial/\partial y$. (Here y is the coordinate of the second factor.) Let $\mathcal{M}_\epsilon(2)$ be the moduli space of the solutions of

$$\frac{d\ell}{dt} = V_\epsilon$$

whose homology class is 2 times the generator of $H_1(Y;\mathbb{Z}) \simeq \mathbb{Z}$. $\mathcal{M}_\epsilon(2)$ can be identified with the fixed point set of $f_\epsilon \circ f_\epsilon$ divided by the $\mathbb{Z}_2$ action induced by f_ϵ on it. Since

$$f_\epsilon \circ f_\epsilon(x) = (1+\epsilon)^2 x - (2+4\epsilon)x^3 + \cdots,$$

in a neighborhood of 0, it follows that the fixed point set of $f_\epsilon \circ f_\epsilon$ consists of one point for $\epsilon < 0$ and of 3 points for $\epsilon > 0$. Taking into acount $\mathbb{Z}_2$ action, we find that $\mathcal{M}_\epsilon(2)$ consists of one point with multiplicity $-1/2$ for $\epsilon < 0$, and of two points with multiplicity -1, $+1/2$, respectively, for $\epsilon > 0$. Hence the total multiplicity is preserved. (Namely $-1/2 = -1 + 1/2$.) At first sight, it seems impossible to keep this independence of total multiplicity without introducing rational coefficient.

This phenomenon is called the period-doubling bifurcation and is famous in the study of dynamical system. (Taubes [T] also discussed it in the context of pseudohomolomorhpic tori in 4 manifolds.) Moreover period-doubling bifurcation can occur repeatedly and multiplicity will become 2^{-m}.

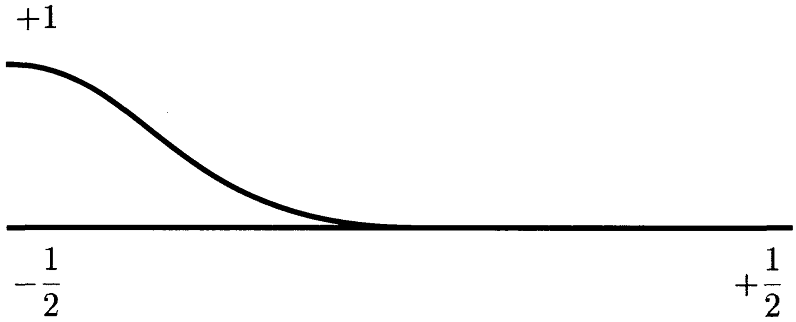


Figure 1.

There is also a similar bifurcation related to cyclic groups of order $\geq$ 3. We will discuss it later in §5.

Let us now go back to our problem. First we compactify $\mathcal{M}(\ell_1, \ell_2; E; k)$ by adding isomorphism classes of maps from singular Rieman surfaces. (See [FO2, §19], where it is called stable connecting orbits.) We denote by $\mathcal{CM}(\ell_1, \ell_2; E; k)$ the compactification. Now the main technical result established in [FO2] is:

Theorem 2 ([FO2, Theorem 19.14]). *$\mathcal{CM}(\ell_1, \ell_2; E; k)$ has Kuranishi structure with corners.*

The precise definition of Kuranishi structure is in [FO2, §5]. We birefly recall it here for reader's convenience. $\mathcal{CM}(\ell_1, \ell_2; E; k)$ is said to have a Kuranishi structure if, for each $x \in \mathcal{CM}(\ell_1, \ell_2; E; k)$, there exists an open subset $U_x \in \mathbb{R}^{m_x}$, a finite group Γ_x (the group of automorphisms of x) such that Γ_x acts on U_x and the action is linear.

We also assume that there exist a Γ_x module $\mathcal{E}_x$ and a Γ_x equivariant map $s_x : U_x \to \mathcal{E}_x$, such that

$$s_x^{-1}(0)/\Gamma_x \simeq \text{a neighborhood of } x \text{ in } \mathcal{CM}(\ell_1, \ell_2; E; k).$$

We need to assume various compatibility conditions for these deta, which are omitted here. We call U_x the Kuranishi neighborhood, $\mathcal{E}_x$ the obstruction bundle and s_x the Kuranishi map.

The idea in [FO2] to find a $\mathbb{Q}$ chain is to perturb s_x by using multivalued perturbation. This method does not work for the purpose of this article. So we first try to go as much as single valued perturbation goes. We then obtain the following Proposition 1. To state it we need some notations. Let s'_x be a (single valued) perturbation of s_x satisfying appropriate compatibility conditions. (See [FO, §6].) We put

$$\mathcal{CM}'(\ell_1, \ell_2; E; k) = \bigcup s_x'^{-1}(0)/\Gamma_x.$$

We write it $\mathcal{CM}'$ in case no confusion can occur. Let G be a finite group. We put

$$\mathcal{CM}'(G) = \{x \in \mathcal{CM}' \mid \Gamma_x \simeq G\},$$
$$\mathcal{G}(G) = \bigcup_{x \in \mathcal{CM}'(G)} \Gamma_x.$$

$\mathcal{G}(G)$ is a local system on $\mathcal{CM}'(G)$.

Proposition 1. *The following holds for generic s'_x.*

(3.1) *$\mathcal{CM}'(G)$ is a smooth manifold with corners.*

(3.2) *There exists two vector bundles $\mathcal{E}_1(G)$, $\mathcal{E}_2(G)$ on $\mathcal{CM}'(G)$. $\mathcal{G}(G)$ acts on them. There exists also a $\mathcal{G}(G)$ equivariant bundle map $s_G : \mathcal{E}_1(G) \to \mathcal{E}_2(G)$ between them. (s_G may not be linear in general.)*

(3.3) *Let $x \in \mathcal{CM}'(G)$ and $\mathcal{E}_{1,x}(G)$, $\mathcal{E}_{2,x}(G)$ be fibers. We regard them as G vector spaces. Then they do not contain trivial component. (Note that this condition implies that s_G sends zero section to zero section.)*

(3.4) *The intersection of $s_G^{-1}(0)/\mathcal{G}(G)$ and a neighborhood of zero section in $\mathcal{E}_1(G)$ is identified to a neighborhood of $\mathcal{CM}'(G)$ in $\mathcal{CM}'$.*

(3.5) *Moreover, for each $x \in \mathcal{CM}'(G)$, its Kuranishi neighborhood U_x is identified to a neighborhood of x in $\mathcal{E}_1(G)$. The obstruction bundle is isomorphic to $\mathcal{E}_2(G)$ and the Kuranishi map is identified to the restriction of s_G to U_x.*

The proof will be given in [FO3]. Hereafter we write $\mathcal{CM}(G)$ etc. in place of $\mathcal{CM}'(G)$ etc.

Remark 2. We remark that, to show Proposition 1, we need to use abstract perturbation. In fact, the conclusion of Proposition 1 is not satisfied by any perturbation of the almost complex structure of M. The reason is that, if we perturb only almost complex structure, then multiple covered spheres may not be made transversal even in the case when its automorphism group is trivial.

Note the condition that the peudoholomorphic sphere is somewhere injective in the sense of McDuff [M] is related to but is different from the condition that pseudoholomorphic sphere does not have nontrivial symmetry.

We are going to show how we use Proposition 1 to avoid period-doubling bifurcations.

To clarify the idea, we first consider the simplest case. Namely we assume that $\mathcal{CM}(G)$ is nonempty only for $G = 1$ or $G = \mathbb{Z}_2$. We put $\mathcal{CM}(1) = N$, $\mathcal{CM}(\mathbb{Z}_2) = M$.

We first remark that (3.3) of Proposition 1 implies that $\mathcal{E}_1(1)$, $\mathcal{E}_2(1)$ are trivial. Namely N is transversal. In other words, the actual dimension of N is equal to its virtual dimension. On the other hand, the dimension of M can be higher than that.

We have $\mathbb{Z}_2$ vector bundles $\mathcal{E}_1(\mathbb{Z}_2)$, $\mathcal{E}_2(\mathbb{Z}_2)$ over M and $s_{\mathbb{Z}_2} : \mathcal{E}_1(\mathbb{Z}_2) \to \mathcal{E}_2(\mathbb{Z}_2)$. (The local system is trivial in this case.) We write $\mathcal{E}_1$, $\mathcal{E}_2$, s in place of $\mathcal{E}_1(\mathbb{Z}_2)$, $\mathcal{E}_2(\mathbb{Z}_2)$, $s_{\mathbb{Z}_2}$ for simplicity. Note that the action of $\mathbb{Z}_2$ on the fibers of $\mathcal{E}_1$, $\mathcal{E}_2$ is $\times -1$. ((3.3) of Proposition 1.) Hence the leading term of $\mathbb{Z}_2$ equivariant map $s : \mathcal{E}_1 \to \mathcal{E}_2$ is linear. So, by replaicing s, we may assume that s is linear in a neighborhood of 0 section. (This is not the case when the group G is more complicated.) We put

$$\Xi = \{x \in M \mid s_x : \mathcal{E}_{1x} \to \mathcal{E}_{2x} \text{ is not injective}\}. \tag{4}$$

By definition, it is easy to see that $M \cap \overline{N} = \Xi$. Namely Ξ is the set of points where period-doubling bifurcation occurs.

We can prove the following lemma by an easy dimension counting.

Lemma 1.

$$\operatorname{codim} \Xi = \operatorname{rank} \mathcal{E}_2 - \operatorname{rank} \mathcal{E}_1 + 1.$$

Note the virtual dimension of our moduli space is $\dim M + \operatorname{rank} \mathcal{E}_1 - \operatorname{rank} \mathcal{E}_2$. Therefore

$$\dim N = \dim M + \operatorname{rank} \mathcal{E}_1 - \operatorname{rank} \mathcal{E}_2 = \dim \Xi + 1.$$

It follows that $\dim \partial N = \dim \Xi$. In other words, N contains other boundary components than those stated in (2.2).

To clarify the topological backgroud, we prove the following:

Proposition 2. *Let M be an oriented closed manifold, $\mathcal{E}_1$, $\mathcal{E}_2$ be oriented vector bundles on it, and $s : \mathcal{E}_1 \to \mathcal{E}_2$ be a generic bundle homomorphism. (s is linear.) We assume that* $\operatorname{rank} \mathcal{E}_2 - \operatorname{rank} \mathcal{E}_1$ *is even. Define Ξ by* (4). *Then we have the following*:

(5.1) *Ξ has an orientation and determines a cycle over $\mathbb{Z}$.*

(5.2) *The Poincaré dual to $[\Xi]$ is δy. Here*

$$\delta : H^k(M; \mathbb{Z}_2) \longrightarrow H^{k+1}(M; \mathbb{Z})$$

is the Bockstein operator associated to the exact sequence $0 \to \mathbb{Z} \stackrel{\times 2}{\to} \mathbb{Z} \to \mathbb{Z}_2 \to 0$, and y is a polynomial of the Stiefel-Whitney classes of $\mathcal{E}_1$, $\mathcal{E}_2$.

Proof. First we define an orientation of Ξ. We put

$$\Xi_2 = \{x \in M \mid \dim \operatorname{Ker} s_x \geq 2\}.$$

It is easy to see that $\dim \Xi - \dim \Xi_2 \geq 2$. So it suffices to define an orientation only on $\Xi - \Xi_2$. (It is also easy to see that $\Xi - \Xi_2$ is a smooth manifold for generic s.) Let $x \in \Xi - \Xi_2$. *Choose* an orientation of $\operatorname{Im} s_x \subset \mathcal{E}_{2,x}$. Take $V_x \subset \mathcal{E}_{1,x}$ such that $s_x : V_x \to \operatorname{Im} s_x$ is an isomorphism. ($\operatorname{rank} V_x = \operatorname{rank} \mathcal{E}_{1,x} - 1$.) The orientation of $\operatorname{Im} s_x$ induces one on V_x. This orientation together with the orientation on $\mathcal{E}_{1,x}$ determine an orientation of one dimensional vector space $\mathcal{E}_{1,x}/V_x$. Let e_x be the oriented basis of the complement of $\mathcal{E}_{1,x}$ in V_x. We extend V_x and e_x to a neighborhood of x and denote it by V and e. Then $s(e)$ determines a section $\overline{e}$ of the bundle $\mathcal{E}_2/s(V)$. It is easy to see that the intersection of Ξ and a neighborhood of x is $\overline{e}^{-1}(0)$. Since the orientations of V and $\mathcal{E}_2$ determine the orientation of $\mathcal{E}_2/s(V)$, we obtain an orientation of $\overline{e}^{-1}(0)$ and of Ξ in a neighborhood of x.

We remark that this orientation of Ξ is independent of the orientation of $\operatorname{Im} s_x$ we have chosen. In fact, if we change the orientation of $\operatorname{Im} s_x$, then the orientation of V will be reversed. Hence we need to replace e by $-e$. On the other hand, the orientation of $\mathcal{E}_2/s(V)$ also will be reversed. Therefore, the orientation on $(-\overline{e})^{-1}(0) = \Xi$ does not change.

It follows that we obtain a global orientation of $\Xi - \Xi_2$.

Next we show the property (5.2). We choose a generic section t of $\mathcal{E}_2$. It induces a section $\overline{t}$ of $\mathcal{E}_2/s(\mathcal{E}_1)$. We remark that $\mathcal{E}_2/s(\mathcal{E}_1)$ is a vector bundle on $M - \Xi$. We put

$$Y = \overline{t}^{-1}(0) \cap (M - \Xi).$$

Let $\overline{Y}$ be its closure in M. Since $\mathcal{E}_2/s(\mathcal{E}_1)$ is oriented, it follows that Y is oriented.

Lemma 2. *$\overline{Y}$ is a $\mathbb{Z}$ chain and satisfies $\partial \overline{Y} = 2\Xi$.*

Proof. Let $x \in \Xi - \Xi_2$. Let U be an neighborhood of x. We choose V_x, V, e_x and e as before. We then obtain an isomorphism $\mathcal{E}_2/s(V)|_{\Xi \cap U} \simeq N_\Xi M$. (Here N denotes the normal bundle.) Hence the restriction of $\mathcal{E}_2/s(\mathcal{E}_1) \simeq (\mathcal{E}_2/s(V))/\overline{e}$ to $\partial N_\Xi M$ is isomorphic to the fiberwise tangent bundle of $\partial N_\Xi M \to \Xi$. The fiber is $S^{\operatorname{rank} \mathcal{E}_2 - \operatorname{rank} \mathcal{E}_1}$. Hence the Euler number of the fiber is 2. (Here we use the assumption that $\operatorname{rank} \mathcal{E}_2 - \operatorname{rank} \mathcal{E}_1$ is even.) t induces a section of $(\mathcal{E}_2/s(V))/\overline{e}$. The induced section is close to constant on U. The lemma follows.

Lemma 2 implies that $[\overline{Y}]$ is a $\mathbb{Z}_2$ cycle and that $[\Xi]$ is a Bockstein image of $[\overline{Y}]$. The proof of Proposition 2 is complete.

The following figure illustrates the relation of Lemma 2 to the period-doubling bifurcation. We remark that t is *not* $\mathbb{Z}_2$ equivariant. Hence it is a multisection in the sense of [FO2]. Therefore $\bar{t}^{-1}(0)$ has the multiplicity $1/2$. $\bar{t}^{-1}(0)/2 + N$ is the $\mathbb{Q}$ cycle constructed in [FO2]. The orientation of $\bar{t}^{-1}(0)$ changes at the point where $\overline{N}$ intersect with it. Hence $\bar{t}^{-1}(0)/2 + N$ becomes a $\mathbb{Q}$ cycle in a similar way as the toy model we discussed before.

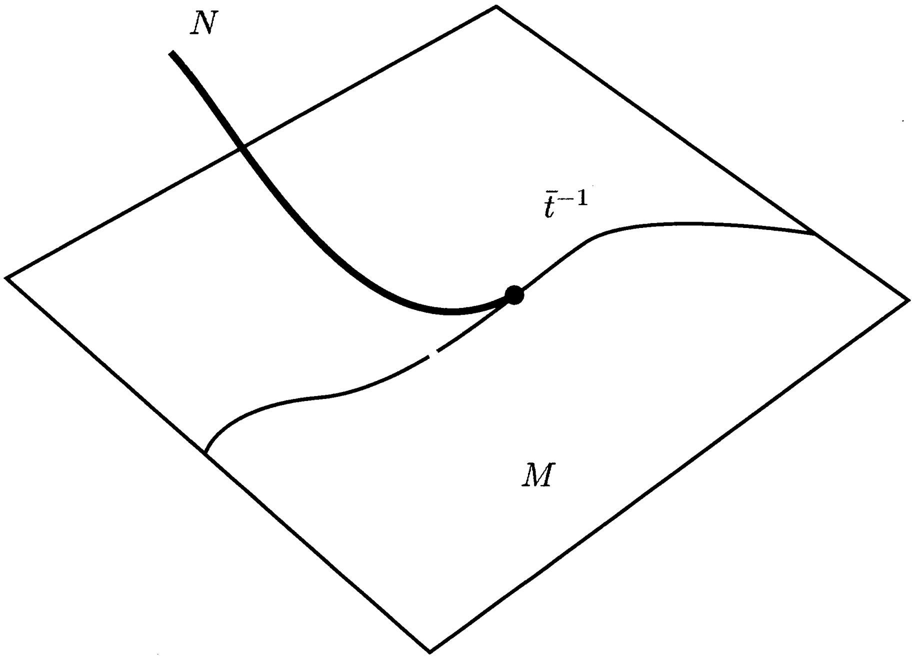


Figure 2.

§4. Normally complex linear perturbation

Proposition 2 suggests, to avoid period-doubling bifurcation, we need to lift $\mathbb{Z}_2$ characteristic classes to a class defined over $\mathbb{Z}$. This is impossible for general oriented vector bundle. However, for complex vector bundle, any $\mathbb{Z}_2$ characteristic class can be lifted to a class defined over $\mathbb{Z}$ in a canonical way, since the cohomology group of complex Grassmannian is torsion free. In fact, we need to perform the construction in the chain level in order to define Floer homology. (Compare [FO2, §20].) For this purpose, we proceed as follows.

Let $\mathcal{E}_1, \mathcal{E}_2 \to M$ be complex vector bundles on an oriented manifold M. (We do not need to assume that M has a complex structure.) Let $s : \mathcal{E}_1 \to \mathcal{E}_2$ be a generic *complex linear* bundle homomorphism. We put

$$\Xi = \{x \in M \mid s_x : \mathcal{E}_{1,x} \to \mathcal{E}_{2,x} \text{ is not injective}\}.$$

Lemma 3.

$$\operatorname{codim}_{\mathbb{R}} \Xi = \operatorname{rank}_{\mathbb{R}} \mathcal{E}_2 - \operatorname{rank}_{\mathbb{R}} \mathcal{E}_1 + 2.$$

The proof is a simple dimension counting. We remark that the right hand side of Lemma 3 is the right hand side of Lemma 1 plus 1. This is a good news.

Now we go back to the Kuranishi structure of Theorerm 2.

Proposition 3. *$[\mathcal{E}_1(G)] - [\mathcal{E}_2(G)] \in KO(\mathcal{CM}(G))$ is in the image of $K(\mathcal{CM}(G))$.*

We proved in [FO2, §16] that Kuranishi structure on the moduli space of stable pseudoholomorphic maps is stably almost complex. (See [FO2, §5] for the definition of stably almost complexity.) In case of the moduli space of stable connecting orbits, the same is true. (We can reduce its proof to the case of closed Rieman surface. We will discuss it in [FO3].) Proposition 3 is a consequence of this fact.

Proposition 3 implies that there exists a vector bundle $\mathcal{F}$ over $\mathcal{CM}(G)$ such that $\mathcal{E}_1(G) \oplus \mathcal{F}$ and $\mathcal{E}_2(G) \oplus \mathcal{F}$ are complex vector bundles. In fact, we can choose $\mathcal{F}$ so that if $x = [\Sigma, \varphi] \in \mathcal{CM}(G)$, then the fiber $\mathcal{F}_x$ is a subspace of $\Gamma(\Sigma, \varphi^* TX \otimes \Lambda^{0,1}(\Sigma))$. So the construction of Kuranishi structure in [FO2] implies that we may change it such that $\mathcal{E}_1(G)$, $\mathcal{E}_2(G)$ will become complex vector bundles for the new Kuranishi structure.

Now we modify s in a neighborhood of 0 section so that it is complex linear there. (We can not change s outside a neighborhood of 0 section, because we need to modify s so that its zero point sets can be patched with N.)

We remark that the modified s is also $\mathbb{Z}_2$ equivariant. The following lemma then is an immediate consequence of Lemma 3.

Lemma 4. *We assume that the virtual dimension of $\mathcal{CM}$ is 0 or 1. We modify s so that it is complex linear in a neighborhood of 0 section. Then Ξ is empty.*

It follows from Lemma 4 that

$$\overline{N} \cap M = \emptyset.$$

We will write $\mathcal{N}(\ell_1, \ell_2; k; E)$, $\mathcal{M}(\ell_1, \ell_2; k; E)$ in place of N, M, in case they are components of $\mathcal{CM}(\ell_1, \ell_2; k; E)$. We then have

$$\begin{aligned} &\sharp \mathcal{CM}(\ell_1, \ell_2; 0; E) \\ &\quad = \sharp \mathcal{N}(\ell_1, \ell_2; 0; E) + \frac{[\mathcal{M}(\ell_1, \ell_2; 0; E)] \cap e(\mathcal{E}_2(\mathbb{Z}_2)/s(\mathcal{E}_1(\mathbb{Z}_2)))}{2}. \end{aligned}$$

Here the left hand side is the fundamental chain of $\mathcal{CM}(\ell_1, \ell_2; 0; E)$ (which is a rational number) in the sense of Kuranishi structure. $\sharp$ in the right hand side is the order counted with sign. $e(\mathcal{E}_2(\mathbb{Z}_2)/s(\mathcal{E}_1(\mathbb{Z}_2))$ is the Euler class of the bundle. In fact, it is not precise to use this notation, since $\mathcal{M}(\ell_1, \ell_2; 0; E)$ may have a boundary. So, to be precise, by using generic section $\bar{t}$ of $\mathcal{E}_2(\mathbb{Z}_2)/s(\mathcal{E}_1(\mathbb{Z}_2))$, we obtain

$$\sharp\mathcal{CM}(\ell_1, \ell_2; 0; E) = \sharp\mathcal{N}(\ell_1, \ell_2; 0; E) + \frac{\sharp(\bar{t}^{-1}(0))}{2}.$$

We recall that the boundary operator we defined in [FO2, §20] is

$$\partial^{old}[\ell_1] = \sum_{\ell_2, E} \sharp\mathcal{CM}(\ell_1, \ell_2; 0; E) T^E [\ell_2].$$

The coefficient in the right hand side is in $\Lambda \otimes \mathbb{Q}$. We define our new boundary operator by

$$\partial^{new}[\ell_1] = \sum_{\ell_2, E} \sharp\mathcal{N}(\ell_1, \ell_2; 0; E) T^E [\ell_2].$$

By applying Lemma 4 to $\mathcal{N}(\ell_1, \ell_2; 1; E)$, we can prove $\partial^{new}\partial^{new} = 0$.

We thus explained the definition of the boundary operator in the case when $\mathcal{CM}(G)$ is nonempty only for $G = 1$, $\mathbb{Z}_2$.

The following figure shows how the moduli space in Figure 1 will be modified.

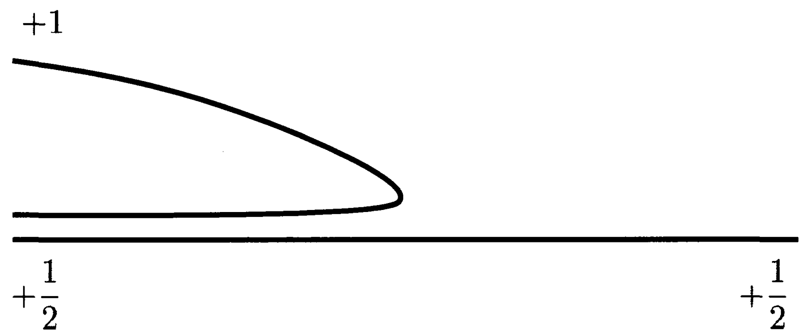


Figure 3.

§5. Another example of bifurcation in the case of cyclic group

Before discussing the case when the group G is general, we mention another example of bifurcation. We consider the case when $\mathcal{CM}(G)$ is empty unless $G = 1$, $\mathbb{Z}_3$. We put $N = \mathcal{CM}(1)$, $M = \mathcal{CM}(\mathbb{Z}_3)$. Let $\dim N = \operatorname{virdim} \mathcal{CM} = 1$, and $M = [0, 1]$. Let us assume that

$\mathcal{E}_1(\mathbb{Z}_3) = \mathcal{E}_2(\mathbb{Z}_3) = M \times \mathbb{C}$. We suppose also that the generator of $\mathbb{Z}_3$ acts by $\times \exp(4\pi\sqrt{-1}/3)$ on $\mathcal{E}_1(\mathbb{Z}_3)$, and by $\times \exp(2\pi\sqrt{-1}/3)$ on $\mathcal{E}_2(\mathbb{Z}_3)$. Let τ be the coordinate of M. We consider $s_\tau : \mathbb{C} \to \mathbb{C}$ such that

$$s_\tau(z) = \begin{cases} \overline{z}, & \tau = 0, \\ z^2, & \tau = 1. \end{cases}$$

if z is in a neighborhood of 0 and $s_\tau(z) = z^2$ if $|z| > 1$. s_τ determines a $\mathbb{Z}_3$-equivariant map $s : \mathcal{E}_1(\mathbb{Z}_3) \to \mathcal{E}_2(\mathbb{Z}_3)$. A neighborhood of M in $\mathcal{CM}$ is identified with

$$\{(z,\tau) \mid s_\tau(z) = 0\}/\mathbb{Z}_3.$$

It is easy to see that this moduli space is described as in Figure 4 below.

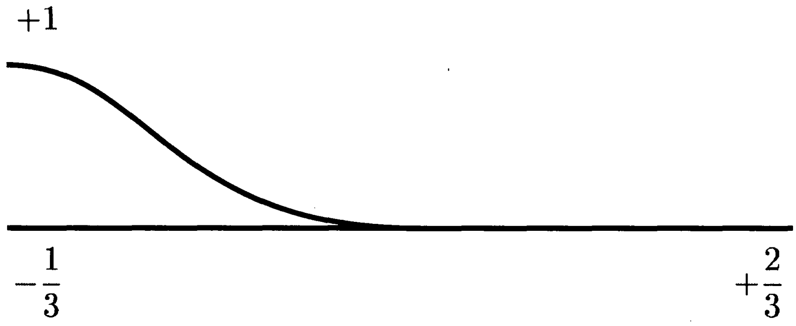


Figure 4.

From this example, it is easy to see that, in the case when the group G is not $\mathbb{Z}_2$, we may not be able to take s so that it is complex linear in a neighborhood of zero section.

We also remark that, if we take s to be generic, then $z \mapsto c\overline{z}$ is the leading term. However we insist s to be holomorphic (or complex polynomial) at each fiber. For example, in this particular case, we take $z \mapsto cz^2$.

§6. The general case

We now go back to the study of $\mathcal{CM}$. The proof of the general case is based on the following Proposition 4. We need some notations. Let M be a manifold and $\mathcal{G}$ be a local system of finite group. Let $\mathcal{E}_1$, $\mathcal{E}_2$ be complex vector bundles on which $\mathcal{G}$ acts. We assume (3.3). We assume moreover that the action of $\mathcal{G}$ on $\mathcal{E}_1$ is effective. Let D be a sufficiently large integer.

Proposition 4. *Let $s : \mathcal{E}_1 \to \mathcal{E}_2$ be a smooth bundle map such that $s_x : \mathcal{E}_{1x} \to \mathcal{E}_{2x}$ is a (complex) polynomial map of degree $\leq D$ for* *each $x \in M$ at a neighborhood is* 0 *section. We assume that s is generic among such maps. We put*

$$N = \{v \in \mathcal{E}_1 \mid s(v) = 0,\, I_v = \{1\}\}.$$

(Here $I_v = \{g \in \mathcal{G}_x \mid gv = v\}$, $v \in \mathcal{G}_x$.) Then we have

$$\dim(\overline{N} - N) \leq \dim M + \operatorname{rank} \mathcal{E}_1 - \operatorname{rank} \mathcal{E}_2 - 2.$$

Sketch of the proof. Let $V_1 = \mathcal{E}_{1x}$, $V_2 = \mathcal{E}_{2x}$ be fibers. Let $\mathrm{Poly}_G^D(V_1, V_2)$ be the set of all G-equivariant polynomial maps $P : V_1 \to V_2$ of degree $\leq D$. There is an evaluation map $ev : \mathrm{Poly}_G^D(V_1, V_2) \times V_1 \to V_2$. We put

$$V_{1\,free} = \{v \in V_1 \mid I_v = \{1\}\},$$
$$Y = ev^{-1}(0) \cap (\mathrm{Poly}_G^D(V_1, V_2) \times V_{1\,free}).$$

Lemma 5. *If the action of G on V_1 is effective, then, for sufficiently large D, the space Y is a smooth manifold of dimension*

$$\dim Y = \dim V_1 + \dim \mathrm{Poly}_G^D(V_1, V_2) - \dim V_2.$$

In other words, ev is a submersion on $\mathrm{Poly}_G^D(V_1, V_2) \times V_{1\,free}$.

Lemma 5 follows easily from the following sublemma:

Sublemma. *Let $p \in V_1$ and $w \in V_2$. We assume $I_p = \{1\}$. Then there exists a G equivariant polynomial map $P : V_1 \to V_2$ such that $P(p) = w$.*

Proof. We may assume that V_2 is an irreducible G module. We put

$$W = \bigoplus_{\gamma \in G} \mathbb{C}[\gamma].$$

and define a G action on it by

$$g\Big(\sum c_\gamma[\gamma]\Big) = \sum c_\gamma[\gamma g^{-1}].$$

Since W is a regular representation of G, there exists a surjective G linear map $\Psi : W \to V_2$. We choose $w_\gamma \in \mathbb{C}$ such that:

$$\Psi\Big(\sum w_\gamma[\gamma]\Big) = w.$$

Since $I_p = \{1\}$, there exists a ($\mathbb{C}$ valued) polynomial f on V_1 such

$$f(\gamma p) = w_\gamma$$

for each $\gamma \in G$. We put

$$P(x) = \Psi\left(\sum_\gamma f(\gamma x)[\gamma]\right).$$

It is straightforward to see that P has the required property[1].

We put $X = \overline{Y} - Y$. The space X is an algebraic variety. We have:

$$\dim_{\mathbb{C}} X \le \dim_{\mathbb{C}} V_1 + \dim_{\mathbb{C}} \mathrm{Poly}_G^D(V_1, V_2) - \dim_{\mathbb{C}} V_2 - 1.$$

Two bundles $\mathcal{E}_1, \mathcal{E}_2 \to M$ induce a bundle $\mathrm{Poly}_G^D(\mathcal{E}_1, \mathcal{E}_2) \to M$ whose fiber is $\mathrm{Poly}_G^D(V_1, V_2)$. We also have a bundle $\mathcal{X} \to M$ whose fiber is X. The projection $X \subset \mathrm{Poly}_G^D(V_1, V_2) \times V_{1\,free} \to \mathrm{Poly}_G^D(V_1, V_2)$ induces a bundle map

$$\pi : \mathcal{X} \longrightarrow \mathrm{Poly}_G^D(\mathcal{E}_1, \mathcal{E}_2).$$

Since X is an algebraic variety, it has simplicial decomposition. Using it we can find a section $\mathfrak{s} : M \to \mathrm{Poly}_G^D(\mathcal{E}_1, \mathcal{E}_2)$ which is of general position to $\pi(\mathcal{X})$. It follows that

$$\dim_{\mathbb{R}}\{x \in M \mid \mathfrak{s}(x) \in \pi(\mathcal{X})\} \le \mathrm{rank}_{\mathbb{R}}\,\mathcal{E}_1 + \dim_{\mathbb{R}} M - \mathrm{rank}_{\mathbb{R}}\,\mathcal{E}_2 - 2.$$

(Note that dimension and rank here are real dimension and real rank.) $\mathfrak{s}$ induces a bundle map $s : \mathcal{E}_1 \to \mathcal{E}_2$ which is a polynomial map on each fibers. It is easy to see that

$$\{x \in M \mid \mathfrak{s}(x) \in \pi(\mathcal{X})\} \simeq \overline{N} - N.$$

Proposition 4 follows.

We apply Proposition 4 to $\mathcal{E}_1(G)$, $\mathcal{E}_2(G)$, $\mathcal{CM}(G)$. We remark that $\mathrm{rank}\,\mathcal{E}_1 + \dim M - \mathrm{rank}\,\mathcal{E}_2$ is the virtual dimension of $\mathcal{CM}$. We modify

[1]Our first idea of the proof of Theorem 1 was to show Lemma 5 under additional assumption that G is abelian, and then use resolution of singularity to reduce the general case to this case. After Theorem 1 had been anounced by the first named author in several conferences, we realized that there is a simpler argument (which we gave above) without using resolution of singularity. We thank Prof. Hambleton who suggested that Proposition 4 may hold without assuming G to be abelian.

$s_G : \mathcal{E}_1(G) \to \mathcal{E}_2(G)$, so that it will be the bundle map constructed by Proposition 4 in a neighborhood of 0 section. Then it is easy to see that $N/\mathcal{G}(G)$ is identified with the intersection of $\mathcal{CM}(1)$ and a neighborhood of $\mathcal{CM}(G)$. Therefore, Proposition 4 implies

$$\dim_{\mathbb{R}} \overline{\mathcal{CM}(1)} \cap \mathcal{CM}(G) \leq \dim_{\mathbb{R}} \mathcal{CM}(1) - 2, \tag{6}$$

for $G \neq 1$. (Note $\dim_{\mathbb{R}} \mathcal{CM}(1)$ is equal to the virtual dimension of $\mathcal{CM}$.)

We modify s_G by an induction of the stratum so that (6) is satisfied.

Now let us consider the case when the virtual dimension of $\mathcal{CM}$ is 0 or 1. Then (6) means that $\mathcal{CM}(1)$ is compact. Hence using it in place of $\mathcal{CM}$, we obtain ∂ such that $\partial^2 = 0$. This is an outline of the proof of Theorem 1.

§7. Gromov-Witten invariant

Our construction in this article can be applied to the moduli space of marked stable maps also. Then we obtain a homology class defined over integer. The result can be summarized as in Theorem 3 below. Let X be an $2n$-dimensional compact symplectic manifold and $\beta \in H_2(X;\mathbb{Z})$. Let

$$GW_{g,m}(X;\beta) \in H_{2m+2\beta c^1+2(3-n)(g-1)}(\mathcal{CM}_{g,m} \times X^m;\mathbb{Q})$$

be the Gromov-Witten invariant. (Here g is the genus m is the number of marked point. $\mathcal{CM}_{g,m}$ is the Deligne-Mumford compactification of the moduli space of stable curves.) (See [FO2, §17] for a definition of Gromov-Witten invariant.)

Theorem 3. *There exists a decomposition*

$$GW_{g,m}(X;\beta) = GW_{g,m}(X;\beta)_{simple} + GW_{g,m}(X;\beta)_{multiple}$$

with the following properties.

(1) $GW_{g,m}(X;\beta)_{simple}$ *is a homology class defined over integer.*

(2) $GW_{g,m}(X;\beta)_{simple}$ *is invariant of the deformation of* X *(as far as it is smooth).*

(3) $GW_{0,3}(X;\beta)_{simple}$ *defines an associative product on* $H^*(X;\Lambda)$.

Some of the other axioms by Kontsevich-Manin [KM] (see also [FO2, §23]) may hold for $GW_{g,m}(X;\beta)_{simple}$. The authors did not check yet which holds and which does not hold.

Problem. Are there any universal formula to calculate $GW_{g,m}(X;\beta)_{multiple}$ in terms of $GW_{g',m'}(X;\beta')_{simple}$ with $g' \leq g$, $m' \leq m$?

We remark that such a formula is known in the case when X is a Calabi-Yau 3 fold and $g = 0$. (See [Ma].)

Our method of this article can be used also in the case of moduli space of pseudoholomorphic disks. Combined with [FKO_3], it gives applications to the problem of Lagrangian intersection. We will discuss it later in [FO3].

References

[A1] V. Arnold, Sur une propriétés des application globalement canonique de la méchanique classique, Comptes Rendus de l'Académie des Science, Paris, **261** (1965), 3719–3722.

[A2] V. Arnold, Mathematical development arising Hilbert problems, in "Proceedings of Symposia Pure Math. **28**", Amer. Math. Soc., Providence, RI, 1976, p. 66.

[CZ] C. Conley and E. Zehnder, The Birkhoff-Lewis fixed point theorem and a conjecture by V. I. Arnold, Invent. Math., **73** (1983), 33–49.

[F] A. Floer, Symplectic fixed point and holomorphic sphere, Commun. Math. Phys., **120** (1989), 575–611.

[FKO_3] K. Fukaya, M. Kontsevich, Y. Oh, H. Ohta and K. Ono, Lagrangian intersection Floer theory – anomaly and abstruction –, in preparation.

[FO1] K. Fukaya and K. Ono, Arnold conjecture and Gromov-Witten invariants for general symplectic manifolds, in "Arnoldfest", Fields Institute Communications **24**, Amer. Math. Soc., 1999, pp. 173–190.

[FO2] K. Fukaya and K. Ono, Arnold conjecture and Gromov-Witten invariants, Topology, **38** (1999), 933–1048.

[FO3] K. Fukaya and K. Ono, in preparation.

[G] M. Gromov, Pseudoholomorphic curves in symplectic geometry, Invent. Math., **82** (1985), 307–347.

[HS] H. Hofer and D. Salamon, Floer homology and Novikov ring, in "Floer's memorial volume", Birkhäuser, Basel, 1995, pp. 483–524.

[KM] M. Kontsevich and Y. Manin, Gromov-Witten classes, quantum cohomology and enumerative geometry, Comm. Math. Phys., **164** (1994), 525–562.

[LT] G. Liu and G. Tian, Floer homology and Arnold conjecture, J. Differential Geom., **49** (1998), 1–74.

[Ma] Y. Manin, Generating functions in algebraic geometry and sums over trees, in "The moduli space of curves", (R. Dijkgraaf, C. Faber and G. van der Geer, eds.), Birkhäuser, Basel, 1994.

[M] D. McDuff, Elliptic methods in symplectic geometry, Bull. Amer. Math. Soc., **23** (1990), 311–358.

[O] K. Ono, On the Arnold conjecture for weakly monotone symplectic manifolds, Invent. Math., **119** (1995), 519–537.

[R] Y. Ruan, Virtual neighborhood and pseudoholomorphic curve, preprint.

[T] C. Taubes, Counting pseudo holomorphic submanifolds in dimension 4, J. Differential Geom., **44** (1996), 818–893.

Kenji Fukaya
Department of Mathematics
Faculty of Science
Kyoto University
Kitashirakawa
Kyoto, 606-8224
Japan

Kaoru Ono
Department of Mathematics
Hokkaido University
Sapporo, 060-0810
Japan

Advanced Studies in Pure Mathematics 31, 2001
Taniguchi Conference on Mathematics Nara '98
pp. 93–125

Very Singular Diffusion Equations

Mi-Ho Giga, Yoshikazu Giga[1] and Ryo Kobayashi[2]

§1. Introduction

In the modeling of nonequilibrium phase transition it is often interesting to consider motion of phase-boundaries driven by singular surface energy. This topic was initiated by J. Taylor [T] and independently by S. Angenent and M. Gurtin [AG] who formulated motion of faceted curves moved by 'crystalline energy'. The governing equation is formally written in a quasilinear diffusion equation. However, because of singularity of energy, the diffusion effect is so strong that it may not be local. Even the notion of solution is not clear in general. There are two ways to handle such very singular diffusion equations systematically as a limit of diffusion equation with smooth energy. The first one is variational approach or the theory of nonlinear semigroups initiated by Y. Kōmura [Ko] and developed by many mathematicians for many years. It provides mathematical formulation of various important problems including the Stefan problem and the Hele-Shaw problem as explained in a book of A. Visintin [V]. The application of this theory to motion with facets is found in [FG] and is further developed by [EGS]; the theory developed in [HZ] is in the line of this approach. The second one is an approach by extending the theory of viscosity solutions initiated by M.-H. Giga and Y. Giga [GG1], [GG3], [GG4]. The first method applies to problem for arbitrary dimensions but the method needs the divergence structure of equations. The second method is so far limited in one space dimension and spatially homogeneous problem. However, it does not require divergence structure of the equation so that it applies equations

Received May 27, 1999.

[1] Partly supported by Nissan Science Foundation and the Grant-in-Aid for Science Research No. 10304010, No. 08874005, the Japan Ministry of Education, Science and Culture.

[2] Partly supported by the Grant-in-Aid for Science Research No. 09650074, the Japan Ministry of Education, Science and Culture.

describing the motion of curves moved by singular surface energy [GG4]. The bibliography of [GG1] and [GG3] includes many references to recent work on the motion by crystalline energy and singular energy. The reader is referred to [GG1] and [GG3] for related references as well as the background of problems.

In this paper, as an example of very singular diffusion equations, we consider a quasilinear equation

$$\frac{\partial u}{\partial t} = \frac{1}{b}\operatorname{div}\left(a\frac{\nabla u}{|\nabla u|}\right), \tag{1.1}$$

where a and b are a given positive function. This equation is interpreted as the gradient system by taking energy

$$E(u) = \int_\Omega a(x)|\nabla u(x)|\,dx \tag{1.2}$$

with respect to the norm $\|f\| = \left(\int_\Omega b(x)|f(x)|^2\,dx\right)^{1/2}$, where Ω is a domain in $\mathbf{R}^n$. In other words (1.1) is written as a gradient system

$$\frac{\partial u}{\partial t} = -(\operatorname{grad}_{\|\cdot\|} E)(u),$$

where $\operatorname{grad}_{\|\cdot\|}$ denotes the gradient with respect to $\|\cdot\|$. When a and b are identically equal to one, $E(u)$ is called a coarea functional and (1.1) is called the coarea gradient flow equation. Its Dirichlet problem is studied in [HZ], where it is proved that the asymptotic limit of solution as time tends to infinity exists and enjoys the relvent minimizing properties for time-independent boundary data. As we observe from (1.2) the energy density $a(x)|p|$ in (1.2) is not differentiable at $p = 0 \in \mathbf{R}^n$ so that the equation has singularity at $|\nabla u| = 0$ in (1.1). The equation (1.1) is also important to describe motion of multi-grain problem studied in [KWC], where b is proportional to a.

The goal of this paper is to review the variational approach to (1.1) to define correct notion of a solution and to give several examples of solutions which develops "facet" or "plateau" (flat portion of the graph of solutions) as an effect of nonlocal diffusion. There is another review paper [KG] on this subject for physicists and material scientists. The present paper describes underlying mathematical basis on this subject for mathematicians.

For one dimensional version of (1.1) (with Dirichlet boundary condition) we derive an interesting necessary and suffieicnt condition so that "plateau" of a solution is preserved when a and b are not necessarily constant. For spatially homogeneous problem, i.e., for constant

a and b, plateau does not break. For piecewise constant initial data we give an explicit way to represent solution even after the time when some plateau merges. We also prove that the number of peaks does not increase during evolution; similar property is well-known for usual diffusion equations ([A], [M], ...). These results seem to be new. Roughly speaking, one of sufficient conditions for nonbreaking of plateau is the concavity of a with respect to 'metric' induced by b at each maximal interval where the solution is constant. This condition is fulfilled for the system proposed by [KWC] where the equation (1.1) is coupled except the fact that a and b now depend on time. We note that a different type of spatially inhomogeneous problem has been studied in [GG2]. At the end of this paper we also note that solution may become discontinuous instanteneously if a is not spatially homogeneous because of strong diffusion (when $b \equiv 1$).

§2. Variational formulation

We recall an abstract formulation for a gradient flow equation for a convex energy. Let H be a real Hilbert space equipped with an inner product $\langle \cdot , \cdot \rangle$. Let φ be a real-valued convex function defined on a convex subset $D(\varphi)$ of H. For technical convenience we extend φ outside $D(\varphi)$ by setting its values as $+\infty$ (which is larger than any real number). The extended function is still denoted φ and $D(\varphi)$ is called the domain of φ.

For application H is taken a space of functions and φ is its energy. It is important to analyse variation or gradient of φ. However, unfortunately φ need not be differentiable in H. For a convex function φ the notion of subdifferential substitutes the notion of gradient. A *subdifferential* of φ at $v \in D(\varphi)$ is the set of all $f \in H$ that satisfies

$$\varphi(v+h) - \varphi(v) \geq \langle f, h \rangle \tag{2.1}$$

for all $h \in H$. The subdifferential of φ at v is denoted $\partial\varphi(v)$. For example, for $\varphi(v) = |v|$ for $v \in \mathbf{R}$ we have $\partial\varphi(v) = \{1\}$ for $v > 0$; $\partial\varphi(v) = [-1, 1]$ for $v = 0$; $\partial\varphi(v) = \{-1\}$ for $v < 0$, if the set $\mathbf{R}$ of all real numbers is regarded as a Hilbert space equipped with the standard inner product. Of course if φ is differentiable at v, $\partial\varphi(v)$ is a singleton and consists of the gradient of φ at v.

We now recall a fundamental theorem for unique existence of solutions for the gradient system

$$\frac{du}{dt} \in -\partial\varphi(u), \quad u|_{t=0} = u_0. \tag{2.2}$$

Unique Existence Theorem. *Assume that φ is convex with nonempty domain $D(\varphi)$ in H and that φ is lower semicontinuous in H, i.e.,*

$$\varphi(v) \leq \liminf_{w \to v} \varphi(w) \quad \text{for all } v \in H,$$

where $w \to v$ denotes the convergence in the norm in H. For each $u_0 \in H$ there is a unique solution u of (2.2) in the sense that

(i) *u is continuous from the time interval $[0, \infty)$ to H and u is absolutely continuous with values in H on each compact set in $(0, \infty)$.*

(ii)

$$\frac{du}{dt}(t) \in -\partial\varphi(u(t)) \quad \textit{for almost all } t \geq 0, \tag{2.3}$$

$$u(0) = u_0. \tag{2.4}$$

Here, the time derivative $f = \frac{du}{dt}(t)$ is defined by the unique element of H that satisfies

$$\lim_{s \to 0} \left\| \frac{u(t+s) - u(t)}{s} - f \right\| = 0,$$

where $\|v\| = \langle v, v \rangle^{1/2}$ is the norm of v in H.

For the proof the reader is referred to a book [Ba]. To apply this theorem to our problem we need to interpret our energy E in (1.2) as a lower semicontinuous convex function on a suitable Hilbert space. We consider the Dirichlet problem in this paper.

Lemma (Lower semicontinuous interpretation). *Let a be a positive continuous function defined in a smoothly bounded domain Ω in $\mathbf{R}^n$. Assume that both a and $1/a$ are bounded in Ω. For a given (Lipschitz) continuous boundary data g on $\partial\Omega$ let $\tilde{g}$ denote a Lipschitz extension of g to $\mathbf{R}^n$. For $v \in H = L^2(\Omega)$ let $\tilde{v}$ be its extension to $\mathbf{R}^n$ such that $\tilde{v}(x) = \tilde{g}(x)$ for $\mathbf{R}^n \backslash \Omega$. Then the functional*

$$\varphi(v) = \int_{\overline{\Omega}} a(x) |\nabla \tilde{v}(x)| \, dx \tag{2.5}$$

with $D(\varphi) = BV(\Omega)$ is convex and lower semicontinuous in the Hilbert space $H = L^2(\Omega)$, where $BV(\Omega)$ denotes the space of functions with bounded variation in Ω. The functional $\varphi(v)$ is independent of the way of extension $\tilde{g}$ of g; it depends on g.

The convexity is easy to check. Note that if one assigns $\varphi \equiv \infty$ outside the set $\{v \in BV(\Omega) \,;\, v|_{\partial\Omega} = g\}$, then φ is not lower semicontinuous in $L^2(\Omega)$. The proof of the lower semicontinuity is standard in the theory of BV [Giu] so we omit it. Note that we have to give a meaning of (2.5) when ∇v is merely a Radon measure before starting the proof. Note also that (2.5) also measures $v|_{\partial\Omega} - g$ on $\partial\Omega$.

For the functional φ defined by (2.5) we derive an explicit form of the gradient form at least in the formal level. The subdifferential depends on the metric of H and there are various inner products of $L^2(\Omega)$. For a given positive continuous function b in Ω we set an inner product.

$$\langle v, w\rangle_b = \int_\Omega v(x)w(x)b(x)\,dx, \quad v, w \in L^2(\Omega).$$

If b identically equals one, this inner product is nothing but the standard inner product of $L^2(\Omega)$. The norm induced by $\langle\,\cdot\,,\,\cdot\,\rangle_b$ is equivalent to the standard one provided that both b and $1/b$ are bounded on Ω.

It is not easy to express the subdifferential when ∇v is a measure as presented in [Te] for a different but related problem. We give a simpler version.

Lemma (Subdifferentials). *Assume the same hypotheses of the preceeding lemma concerning Ω, a, g and φ. Assume that v is Lipschitz continuous (so that ∇v is bounded on Ω) and that $v|_{\partial\Omega} = g$. Let $\partial\varphi(v)$ denote the subdifferential of φ at v with respect to the inner product $\langle\,\cdot\,,\,\cdot\,\rangle_b$ where both $b > 0$ and $1/b$ are assumed to be bounded continuous. Then $f \in \partial\varphi(v)$ if and only if there is a locally integrable function ξ on Ω that satisfies*

$$f = -\frac{1}{b}\operatorname{div}(a\xi), \quad \xi(x) \in \partial j(\nabla v(x)) \tag{2.6}$$

for almost all $x \in \Omega$, where $j(p) = |p|$ for $p \in \mathbf{R}^n$ and ∂j is the subdifferential of j with respect to the standard inner product of $\mathbf{R}^n$.

This is an easy corollary of the result in [AD] by modifying $j(p)$ for large p so that $j(p)/|p| \to \infty$ as $|p| \to \infty$ as in [FG]. The proof that (2.6) implies $f \in \partial\varphi(v)$ directly follows from the definition of subdifferential while the converse is nontrivial. This lemma asserts that the equation (2.2) (with φ given by (2.5)) is formally written as

$$\frac{\partial u}{\partial t} = \frac{1}{b}\operatorname{div}\left(a\frac{\nabla u}{|\nabla u|}\right), \quad u|_{t=0} = u_0, \tag{2.7}$$

with the Dirichlet boundary condition $u|_{\partial\Omega} = g$, since

$$\partial j(p) = \begin{cases} \{p/|p|\}, & \text{for } p \neq 0, \\ B_1, & \text{for } p = 0, \end{cases} \tag{2.8}$$

where B_1 is a closed unit ball centered at the origin in $\mathbf{R}^n$. By a solution of (2.7) with $u|_{\partial\Omega} = g$ we mean that it is a solution of (2.2) with φ given by (2.5) defined in $H = L^2(\Omega)$ equipped with the inner product $\langle\,\cdot\,,\,\cdot\,\rangle_b$. Note that the boundary condition is hidden in φ. The unique existence theorem for (2.2) asserts that the equation (2.7) with $u|_{\partial\Omega} = g$ is uniquely solvable. If $a \equiv b \equiv 1$, (2.7) is called the coarea gradient flow equation in [HZ] which is qualitatively different from the level set mean curvature flow equation

$$\frac{\partial u}{\partial t} = |\nabla u| \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right)$$

analysed in [CGG], [ES] since (2.7) turns to have a nonlocal effects while the last equation does not have such an effert; see [G] for review on the level set equations. If a and b depend on time, (2.7) is not written in the form of (2.2) unless we use time dependent inner product and energy. Fortunately, an abstract theory including such situation has been developed by A. Damlamian [D]. We thank Professor N. Kenmochi for pointing out this reference.

We come back to the abstract setting in the unique existence theorem. The condition (2.3) can be interpreted as a variational inequality. By definition (2.1) the condition (2.3) is equivalent to a variational inequality:

$$\varphi(v) - \varphi(u(t)) \geq \left\langle v - u(t), -\frac{du}{dt}(t)\right\rangle \quad \text{for all } v \in H \text{ and almost all } t \geq 0. \tag{2.9}$$

It is not difficult to see the uniqueness of solution of (2.2). Indeed, let w fulfill (2.9), i.e.,

$$\varphi(v) - \varphi(w(t)) \geq \left\langle v - w(t), -\frac{dw}{dt}(t)\right\rangle \quad \text{for all } v \in H \text{ and almost all } t \geq 0. \tag{2.9$'$}$$

Setting $v = w(t)$ in (2.9) and $v = u(t)$ in (2.9)$'$ and adding (2.9) and (2.9)$'$ yields

$$\begin{aligned} 0 &\geq \left\langle w(t) - u(t), \frac{dw}{dt}(t) - \frac{du}{dt}(t)\right\rangle \\ &= \frac{1}{2}\frac{d}{dt}\|w(t) - u(t)\|^2 \quad \text{for almost all } t \geq 0. \end{aligned}$$

This implies the *contraction property*:

$$\|w(t) - u(t)\| \le \|w_0 - u_0\| \quad \text{for all } t \ge 0,$$

where w_0 is the initial data $w(0)$ of w. The contraction property immediately implies the *uniqueness* of solution of (2.2).

The evolution law (2.2) looks ambiguous since $\partial\varphi$ is multivalued. But by the uniqueness of solution the solution knows how to evolve. Does du/dt choose a special element of subdifferential $\partial\varphi(u(t))$? The next theorem which is well-known gives an answer. For the proof see e.g. [Ba].

Theorem on characterization of the speed. *Let u be the solution of (2.2) with $u_0 \in H$. Then u is right differentiable for all $t > 0$. Let d^+u/dt denote the right derivative. Then $f = -\frac{d^+u}{dt}(t)$ is the canonical restriction (or minimal section) of $\partial\varphi(u(t))$, i.e., $f \in \partial\varphi(u(t))$ and*

$$\|f\| = \min\{\|q\| \ ; \ q \in \partial\varphi(u(t))\}.$$

Conversely, if a continuous function u from $[0, \infty)$ to H is right differentiable at all $t > 0$ and $-\frac{d^+u}{dt}(t)$ is the canonical restriction of $\partial\varphi(u(t))$, then u is the solution of (2.2) with initial data $u_0 = u(0)$.

Since φ is convex and lower semicontinuous, the set $\partial\varphi(v)$ is always a closed convex set (which may be empty), so the canonical restriction is unique which is denoted $\partial\varphi^0(v)$ $(\in \partial\varphi(v))$.

We only give a formal proof for the characterization of the speed at $t_0 > 0$ by assuming that $\frac{du}{dt}(t_0)$ exists with the property that $\frac{du}{dt}(t_0) \in -\partial\varphi(u(t_0))$ and du/dt is right continuous at $t = t_0$. For the detailed proof see e.g. [Ba]. We set $t = t_0 + s$, $s > 0$ in (2.9) and $v = u(t_0)$ to get

$$\varphi(u(t_0)) - \varphi(u(t_0 + s)) \ge \left\langle u(t_0) - u(t_0 + s), -\frac{du}{dt}(t_0 + s) \right\rangle$$

For any $\zeta \in \partial\varphi(u(t_0))$ by definition

$$\langle \zeta, u(t_0) - u(t_0 + s) \rangle \ge \varphi(u(t_0)) - \varphi(u(t_0 + s)).$$

These two inequalities yield

$$\left\langle \frac{du}{dt}(t_0 + s), u(t_0 + s) - u(t_0) \right\rangle \le \langle -\zeta, u(t_0 + s) - u(t_0) \rangle.$$

Dividing both sides by s and sending s to zero yields

$$\left\| \frac{du}{dt}(t_0) \right\|^2 \le \left\langle -\zeta, \frac{du}{dt}(t_0) \right\rangle \le \|\zeta\| \left\| \frac{du}{dt}(t_0) \right\|$$

by the Schwarz inequality since du/dt is assumed to be right continuous at $t = t_0$. Thus for any $\zeta \in \partial\varphi(u(t_0))$ we have

$$\left\| \frac{du}{dt}(t_0) \right\| \leq \|\zeta\|.$$

Since $-\frac{du}{dt}(t_0) \in \partial\varphi(u(t_0))$, this minimality implies that $-\frac{du}{dt}(t_0) = \partial\varphi^0(u(t_0))$.

The solution of (2.2) we obtain is a nice stability property for perturbation of energy φ.

Stability Theorem. *Assume that φ_ε converges to φ in the sense of Mosco as $\varepsilon \to 0$, i.e., for any $v \in H$, $\varphi(v) \leq \liminf_{\varepsilon\to 0} \varphi_\varepsilon(v_\varepsilon)$ for any v_ε with $\|v_\varepsilon - v\| \to 0$ and there is a sequence $\tilde{v}_\varepsilon$ converges weakly in H that $\varphi(v) = \lim_{\varepsilon\to 0} \varphi_\varepsilon(\tilde{v}_\varepsilon)$. Assume that $\|u_0^\varepsilon - u_0\| \to 0$. Let u_0^ε be the solution of*

$$\frac{du}{dt} \in -\partial\varphi_\varepsilon(u), \quad u|_{t=0} = u_0^\varepsilon,$$

where φ_ε is a lower semicontinuous, convex function. Then for every $T > 0$

$$\lim_{\varepsilon\to 0} \sup_{0\leq t\leq T} \|u^\varepsilon(t) - u(t)\| = 0.$$

The result is due to J. Watanabe [W] based on a result of H. Brezis-A. Pazy [BP]. It asserts that our solution (2.2) can be obtained as a limit of approximate problems if the energy is approximated in a proper way. In practice it gives a way to calculate numerically a solution approximately by approximating φ by a smoother energy. We approximate φ in (2.5) by a smoother energy so that the approximate equation enjoys a comparison principle. By the stability theorem the comparison principle is inherited to (2.7).

Comparison principle. *Assume that φ is defined by* (2.5) *and $H = L^2(\Omega)$ equipped with inner product $\langle\ ,\ \rangle_b$. For solutions u and v of $du/dt \in -\partial\varphi(u)$, $u \leq v$ for all $t \geq 0$ if $u \leq v$ at $t = 0$.*

We are curious whether u^ε convergerges to u locally uniformly in space-time domain. So far such a convergence results is proved only for problem for one space dimension based on the theory of viscosity solutions [GG1], [GG3], when the equation is spatially homogeneous. In our problem (2.7) is spatially homogeneous if a and b are constants. We only give a simplest version of a general convergence result including nondivergence type equations proved in [GG3].

Uniform convergence. *Assume that $W_\varepsilon(p)$ converves to $|p|$ as $\varepsilon \to 0$ locally uniformly and that the second derivative $W''_\varepsilon > 0$ and W_ε is smooth. Assume that u_0^ε and u_0 is continuous in a closed bounded interval Ω in $\mathbf{R}$ with $u_0^\varepsilon = u_0 = 0$ on the boundary $\partial\Omega$ of Ω. Assume that u_0^ε converges to u_0 uniformly in $\overline{\Omega}$. Let u^ε be the solution of*

$$\frac{\partial u}{\partial t} = (W'_\varepsilon(u_x))_x, \qquad u|_{t=0} = u_0^\varepsilon,\ u|_{\partial\Omega} = 0.$$

Then u^ε converges to a unique solution u of

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\frac{u_x}{|u_x|}\right), \qquad u|_{t=0} = u_0,\ u|_{\partial\Omega} = 0$$

as $\varepsilon \to 0$ and the convergence is locally uniform in $\overline{\Omega} \times [0,\infty)$, where $u_x = \partial_x u = \partial u/\partial x$. Both u^ε and u are at least continuous in $\overline{\Omega} \times [0,\infty)$.

§3. Examples of solutions

We consider one-dimensional version of the equation (2.7). The equation is of form

$$\frac{\partial u}{\partial t} = \frac{1}{b}\frac{\partial}{\partial x}\left(a\frac{u_x}{|u_x|}\right). \tag{3.1}$$

3.1. Spatially homogeneous equation

To see the nonlocal effect of strong diffusion we first consider the spatially homogeneous equation:

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\frac{u_x}{|u_x|}\right), \tag{3.2}$$

which is of course an example of (3.1). We consider (3.2) for $x \in \Omega$ and $t > 0$ where $\Omega = (x_0, x_1)$ is an bounded open interval. For boundary condition we impose zero Dirichlet data but this is just to fix our problem. The problem (3.2) with zero Dirichlet data and initial data is formulated by (2.2) by taking

$$\begin{aligned} &H = L^2(\Omega), \quad \varphi(v) = \int_\Omega |v_x|\,dx + |v(x_0)| + |v(x_1)|, \\ &D(\varphi) = BV(\Omega), \end{aligned} \tag{3.3}$$

where H is equipped with the inner product

$$\langle f, g\rangle_1 = \int_\Omega f(x)g(x)\,dx.$$

We shall calculate subdifferential of φ at v when $v \in D(\varphi)$ is a Lipschitz continuous function with $v(x_0) = v(x_1) = 0$ having the property

$$\begin{cases} \text{monotone increasing on} & [x_0, \alpha], \\ \text{constant on} & [\alpha, \beta], \\ \text{monotone decreasing} & [\beta, x_1], \end{cases}$$

where $\Omega = (x_0, x_1)$ and $x_0 < \alpha < \beta < x_1$. By Lemma on subdifferentials we see $f \in \partial\varphi(v) \subset H$ if and only if there exists ξ satisfying

$$f = -\xi_x,$$
$$\xi(x) \begin{cases} = 1, & \text{if } v \text{ is increasing near } x, \\ = -1, & \text{if } v \text{ is decreasing near } x, \\ \in [-1, 1], & \text{otherwise.} \end{cases}$$

Since f is in $L^2(\Omega)$, ξ must be continuous on Ω. Thus the graph of ξ is as in Figure 1 (a). If $f = -\xi_x$ is the canonical restriction of $\partial\varphi(v)$, ξ must minimize

$$\int_\alpha^\beta |\xi_x|^2 \, dx$$

under constraint $|\xi| \leq 1$ on (α, β) with $\xi(\alpha) = 1$, $\xi(\beta) = -1$. The minimizer is an affine function on (α, β) so that $\xi_x = -2/(\beta - \alpha)$ as shown in Figure 1 (b). Thus

$$\partial\varphi^0(v)(x) = \begin{cases} 0, & x \in (x_0, \alpha) \text{ or } x \in (\beta, x_1), \\ +\frac{2}{\beta-\alpha}, & x \in (\alpha, \beta). \end{cases}$$

Note that $\partial\varphi^0(v)(x)$ is a nonlocal quantity determined by v for $x \in (\alpha, \beta)$.

Based on the calculation of $\partial\varphi^0(v)$ we seek the solution of (3.2) (i.e., the solution of (2.2) with (3.3)) for single peak initial data. Let u be a continuous function in $\overline{\Omega}$ of form

$$u_0(x) = \begin{cases} A(x), & x_0 \leq x \leq \alpha_0, \\ h_0, & \alpha_0 \leq x \leq \beta_0, \\ B(x), & \beta_0 \leq x \leq x_1, \end{cases} \tag{3.4}$$

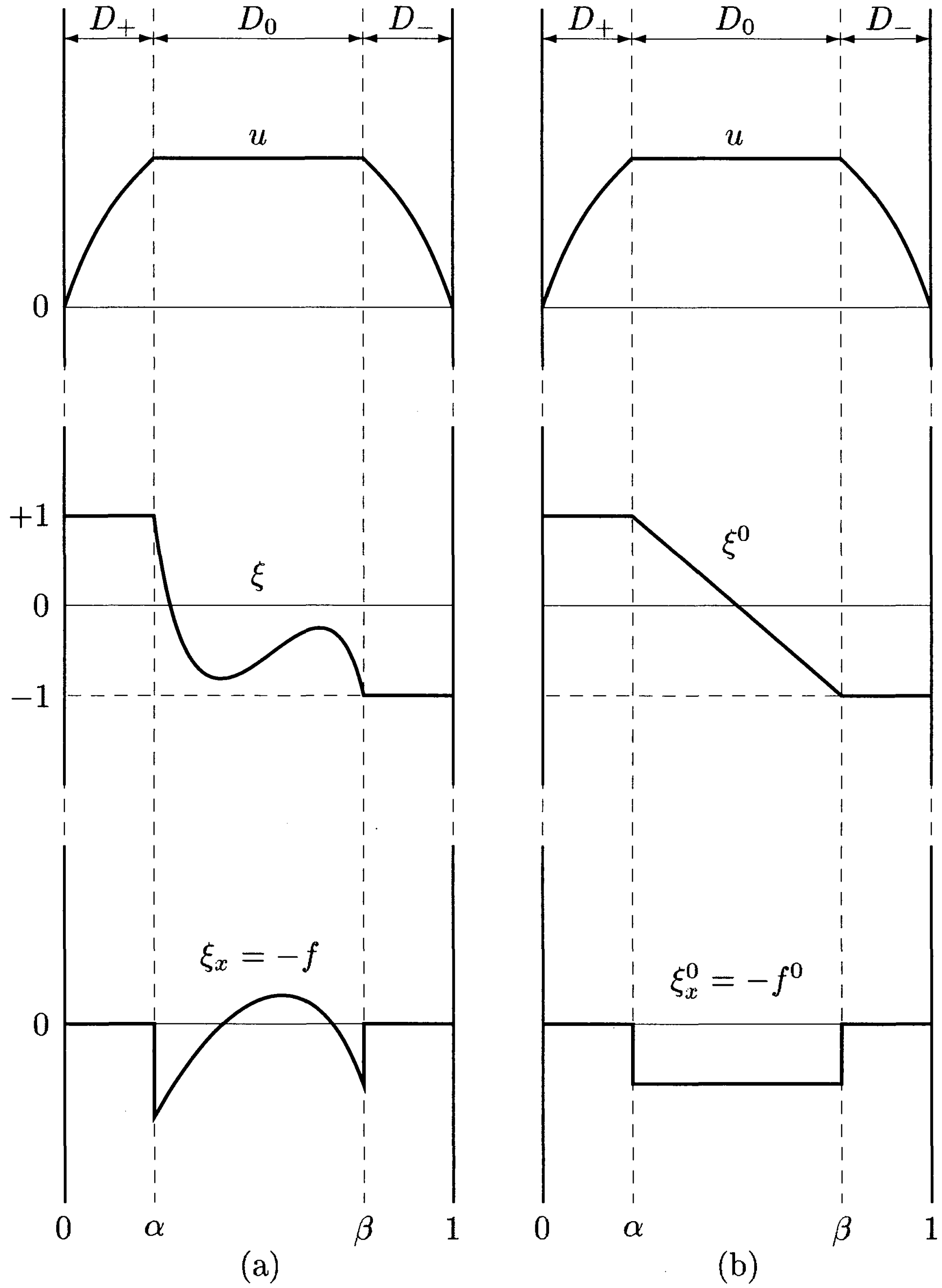


Fig. 1. The interval $[0,1]$ is divided into three intervals $D_+ = [0,\alpha)$, $D_0 = [\alpha,\beta]$ and $D_- = (\beta,1]$. In D_+, $\xi \equiv +1$ holds, and $\xi \equiv -1$ in D_-. (a) One example of ξ and $\xi_x = -f$ where f belongs to $\partial\varphi(u)$. (b) ξ^0 and $\xi_x^0 = -f^0$ where f^0 is a canonical restriction of $\partial\varphi(u)$.

with $\alpha_0 \leq \beta_0$, $A' > 0$, $B' < 0$, $A(\alpha_0) = B(\beta_0) = h_0$ and $A(x_0) = B(x_1) = 0$. We expect the solution of (3.2) is of form

$$u(x,t) = \begin{cases} A(x), & x_0 \leq x \leq \alpha(t), \\ h(t), & \alpha(t) \leq x \leq \beta(t), \\ B(x), & \beta(t) \leq x \leq x_1, \end{cases} \tag{3.5}$$

with $A(\alpha(t)) = B(\beta(t)) = h(t)$ until the time T such that $h(T) = 0$. By the characterization of the speed if

$$h_t(t) = -\partial\varphi^0(u(\,\cdot\,,t)) = -\frac{2}{\beta(t) - \alpha(t)}, \quad 0 < t < T, \tag{3.6}$$

then $u(x,t)$ solves (3.2). Fortunately there is continuous α, β on $[0,T]$ that satisfies (3.6). In fact

$$\begin{aligned} &\alpha(t) = A^{-1}(h(t)), \ \beta(t) = B^{-1}(h(t)), \ \alpha(0) = \alpha_0, \ \beta(0) = \beta_0, \\ &h(t) = S^{-1}(2t), \\ &S(k) = \int_k^{h_0} (B^{-1}(\eta) - A^{-1}(\eta))\, d\eta, \end{aligned}$$

where -1 represents the inverse of a function. By this choice of α, β, h, (3.5) is now the unique solution of (3.2) with initial data u_0 given by (3.4); after the time T we set $u(x,t) \equiv 0$. It is not difficult to study evolution of multi peak function by (3.2). The important feature of the shape is local maximum and minimum is flattened by nonlocal diffusion effects and it may merge. See [KG, Section 3] for such examples as well as numerical simulations.

Solution starting from u_0 of (3.4) is given in [HZ]; a similar example with initial data $-u_0$ is given in [GG1]. One way of proving (3.6) is based on observation given in [FG], where the speed $-2/(\beta - \alpha)$ of evolution equals the canonical restriction of φ; see also [EGS].

3.2. Equations with inhomogeneous diffusion

We now consider (3.1) for $x \in \Omega = (z_0, z_1)$ and $t > 0$ with inhomogeneous Dirichlet boundary condition. We consider a piecewise constant initial data

$$u_0(x) = h_i^0 \quad \text{on } (x_i, x_{i+1}), \ i = 0, \ldots, m-1, \ m \geq 2 \tag{3.7}$$

where $z_0 = x_0 < x_1 < x_2 < \cdots < x_m = z_1$. The values h_i^0 may be the same as h_{i+1}^0. The boundary condition we impose is $u = h_0^0$ at z_0, $u = h_{m-1}^0$ at z_1. We interpret (3.1) as (2.2) with (2.5) where $g = h_0^0$

at z_0 and $g = h^0_{m-1}$ at z_1 in the definition of $\varphi(v)$; the inner product of $H = L^2(\Omega)$ is chosen by $\langle\ ,\ \rangle_b$ as in §2. In general the solution u with initial data (3.7) may not be locally constant in x because of inhomogeneity of a and b. We shall seek conditions on a and b so that $u(\,\cdot\,,t)$ is piecewise constant and its jump discontinuities are included in $\{x_i\}_{i=1}^{m-1}$. If $h^0_i = h^0_{i+1} = \cdots = h^0_{i+k-1}$, $h^0_{i-1} \neq h^0_i$, $h^0_{i+k-1} \neq h^0_{i+k}$ we say the segment

$$(x_i, x_{i+k}) \times \{h^0_i\}, \quad k \geq 1$$

is a *plateau* (of height h^0_i) for a piecewise constant function u_0. We first consider the case that $h^0_{i-1} \neq h^0_i$ for $i = 1, \ldots, m-1$.

3.2.1. *Evolution before merging of plateaus*

Theorem on persistency. *Assume that a and b are positive and continuous on $\overline{\Omega} = [z_0, z_1]$. Assume that $a(x_1) \leq a(x)$ for all $x \in [x_0, x_1]$ and $a(x_{m-1}) \leq a(x)$ for all $x \in [x_{m-1}, x_m]$ and that*

(C_i)

$$\{a(x_{i+1}) - a(x_i)\} \Big/ \int_{x_i}^{x_{i+1}} b(\tau)\, d\tau \leq \{a(x) - a(x_i)\} \Big/ \int_{x_i}^{x} b(\tau)\, d\tau \qquad \text{for all } x \in (x_i,\ x_{i+1}]$$

holds for all $i = 1, \ldots, m-2$. Let u be the solution of (3.1) *with initial data u_0 given by* (3.7) *with Dirichlet boundary condition $u|_{\partial\Omega} = u_0|_{\partial\Omega}$. Assume that $h^0_i \neq h^0_{i-1}$ for $i = 1, \ldots, m-1$. Then for each $i = 0, 1, \ldots, m-1$ the speed $u_t(x,t)$ is independent of $x \in (x_i, x_{i+1})$ and $t \in (0, t_0)$, where t_0 is the first time that some plateau of $u(\,\cdot\,,t)$ merges to another one. Moreover $u_t(x,t) \equiv 0$ in (x_0, x_1) and (x_{m-1}, x_m) for $0 < t < t_0$. (Thus the function $u(\,\cdot\,,t)$ is piecewise constant and it jumps at $x_1, \ldots, x_{m-1}$ for $t \in (0, t_0)$.)*

Remark 1. If we use the length with respect to the metric $b\,dx$, the condition (C_i) is rewritten as

$$\{a_b(y_{i+1}) - a_b(y_i)\}/(y_{i+1} - y_i) \leq \{a_b(y) - a_b(y_i)\}/(y - y_i) \qquad \text{for all } y \in (y_i, y_{i+1}],$$

where $a_b(y) = a(x(y))$, $y_i = \int_{x_1}^{x_i} b(\tau)\, d\tau$ and $x(y)$ is the inverse function of $y(x) = \int_{x_1}^{x} b(\tau)\, d\tau$. In other words the convex hull of a_b on $[y_i, y_{i+1}]$ is affine.

Remark 2. It turns out that the condition (C_i) is necessary so that u_t is constant on (x_i, x_{i+1}) for any initial data u_0 given by (3.7) with $h_j^0 \neq h_{j-1}^0$ ($j = 1, \ldots, m-1$). Also $a(x) \geq a(x_1)$ on $[x_0, x_1]$ and $a(x) \geq a(x_{m-1})$ on $[x_{m-1}, x_m]$ are necessary so that u_t is zero on (x_0, x_1) and (x_{m-1}, x_m), respectively. We shall explain the reasons in the proof of Lemma 2.

Remark 3. It is clear that (C_i) is fulfilled if a is concave when b is constant on $[x_i, x_{i+1}]$. It is more difficult to see that (C_i) is fulfilled if a is concave on $[x_i, x_{i+1}]$ when b is proportional to a. We shall prove this statement at the end of this subsection §3.2.1.

Since u_0 is discontinuous, our Lemma on subdifferentials does not apply. We seek a nice element of subdifferential at a piecewise constant function.

Lemma 1. *Assume that a and b are positive and continuous on $[z_0, z_1]$. Assume that u_0 is given by* (3.7)*. Let $f \in L^2(\Omega)$ be of form*

$$f(x) = -\frac{1}{b(x)}\{a(x)\xi(x)\}_x \quad |\xi(x)| \leq 1,\ x \in \Omega = (z_0, z_1) \tag{3.8}$$

for some continuous ξ in Ω that satisfies

$$\xi(x_i) = \begin{cases} 1, & \text{if } h_{i-1}^0 < h_i^0, \\ -1, & \text{if } h_{i-1}^0 > h_i^0, \end{cases} \tag{3.9}$$

for $i = 1, 2, \ldots, m-1$. (If $h_{i-1}^0 = h_i^0$, no condition on $\xi(x_i)$ is imposed.) Then $f \in \partial\varphi(u_0)$, where the boundary data g is taken as the boundary value of u_0 and φ is given in (2.5)*. Conversely, if $f \in \partial\varphi(u_0)$ then f is of form* (3.8) *satisfying* (3.9)*.*

Proof of Lemma 1. We shall check (2.1) or

$$\langle v - u_0, f\rangle_b \leq \varphi(v) - \varphi(u_0)$$

for all $v \in D(\varphi)$. By definition

$$\langle v - u_0, f\rangle_b = -\int_\Omega (v - u_0)(a\xi)_x\, dx. \tag{3.10}$$

Since $|\xi| \leq 1$, integrating by parts we see

$$-\int_\Omega v(a\xi)_x\, dx = \int_\Omega v_x a\xi\, dx - u_0 a\xi|_{z_0}^{z_1} \leq \varphi(v) - u_0 a\xi|_{z_0}^{z_1}; \tag{3.11}$$

here v_x is regarded as a measure and $\varphi(v) = \int |v_x| a(x)\, dx$ is defined for a Radon measure v_x, for example

$$\varphi(u_0) = \sum_{i=1}^{m-1} a(x_i)|h_i - h_{i-1}|.$$

Since $\xi(x_i) = \pm 1$ depending on the sign of $h_i - h_{i-1}$, we see

$$\begin{aligned} \int_\Omega u_0(a\xi)_x\, dx &= u_0 a\xi|_{z_0}^{z_1} - \sum_{i=1}^{m-1} a(x_i)|h_i - h_{i-1}| \\ &= u_0 a\xi|_{z_0}^{z_1} - \varphi(u_0). \end{aligned} \tag{3.12}$$

The formula (3.10)–(3.12) now yields

$$\begin{aligned} \langle v - u_0, f\rangle_b &\le \varphi(v) - u_0 a\xi|_{z_0}^{z_1} + u_0 a\xi|_{z_0}^{z_1} - \varphi(u_0) \\ &= \varphi(v) - \varphi(u_0), \end{aligned}$$

which implies $f \in \partial\varphi(u_0)$.

Conversely, assume that $f \in \partial\varphi(u_0)$. Let ζ by a primitive of $-bf$. Since $f \in L^2(\Omega)$, ζ must be absolutely continuous in Ω. The condition $f \in \partial\varphi(u_0)$ is equivalent to

$$- \int_\Omega (v - u_0)\zeta_x\, dx \le \varphi(v) - \varphi(u_0). \tag{$*$}$$

We take various v in this inequality to derive properties of ζ. If $u_{0x} \not\equiv 0$, there is an index $i \in \{1, \dots, m-1\}$ such that $h_i^0 \neq h_{i-1}^0$. We may assume that $h_{i-1}^0 < h_i^0$. For $\hat{x} \in (z_0, z_1) \setminus \{x_j\}_{j=1}^{m-1}$ we take

$$v(x) = u_0(x) + \lambda \int_{z_0}^x (\delta(\tau - \hat{x}) - \delta(\tau - x_i)), \quad \lambda < h_i^0 - h_{i-1}^0$$

in ($*$) and integrate by parts to get $\lambda(\zeta(\hat{x}) - \zeta(x_i)) \le |\lambda| a(\hat{x}) - \lambda a(x_i)$. If we set $\zeta(x_i) = a(x_i)$, then this inequality yields $|\zeta(\hat{x})| \le a(\hat{x})$ for all $\hat{x} \in (z_0, z_1) \setminus \{x_j\}_{j=1}^{m-1}$ by taking λ positive or negative. By continuity this implies $|\zeta(x)| \le a(x)$ for all $x \in (z_0, z_1)$. If $h_{i+1}^0 > h_i^0$ we take $\hat{x} = x_{i+1}$ and plug above v in ($*$) to get $\lambda(\zeta(x_{i+1}) - \zeta(x_i)) \le \lambda(a(x_{i+1}) - a(x_i))$ for $\lambda \in \mathbf{R}$ with small $|\lambda|$. This yields $\zeta(x_{i+1}) = a(x_{i+1})$. Similarly, we have $-\zeta(x_{i+1}) = a(x_{i+1})$ if $h_{i+1}^0 < h_i^0$. Repeating this argument for both sides of x_i we conclude that f is of form (3.8) satisfying (3.9) when $u_{0x} \not\equiv 0$. (If u_0 is a constant, we normalize ζ so that $\max(\zeta - a) = 0$ to get $\zeta(x_*) = a(x_*)$ for some $x_* \in [z_0, z_1]$. We set v as above in ($*$) with $\hat{x} = x_*$, $x_1 = x$ and $\lambda = 1$ to get $-\zeta(x) \le a(x)$ which yields $|\zeta(x)| \le a(x)$.) Q.E.D.

Lemma 2. *Assume that u_0 is given by (3.7) with $h_i^0 \neq h_{i-1}^0$ for $i = 1, \ldots, m-1$. Assume the same hypothesis of theorem on persistency concerning a and b. Then there is continuous ξ on Ω satisfying (3.8) and (3.9) such that f given by (3.8) is constant $-\nu^i$ on each interval (x_i, x_{i+1}) $(i = 0, 1, \ldots, m-1)$ and that $\nu^0 = \nu^{m-1} = 0$. The constant ν^i is of form*

$$\nu^i = \frac{(a\xi)(x_{i+1}) - (a\xi)(x_i)}{\int_{x_i}^{x_{i+1}} b\, dx}, \quad i = 0, 1, \ldots, m-1,$$

which depends on u_0 only through the order of h_i, h_{i-1}, h_{i+1}.

Proof of Theorem on persistency. Since $h_i^0 \neq h_{i-1}^0$ $(i = 1, \ldots, m-1)$, by Lemmas 1 and 2 (and the unique existence theorem) we see

$$u(x,t) = h_i(t) \quad \text{on } (x_i, x_{i+1}) \text{ with } \frac{dh_i}{dt}(t) = \nu^i$$

for $i = 0, 1, \ldots, m-1$ is the unique solution of (3.1) with initial data u_0 given in (3.7), until the first time t_0 when $h_i(t_0) = h_{i+1}(t_0)$ for some $i = 0, 1, \ldots, m-1$. Actually we have used a version of uniqueness of a local solution for (2.2) on $[0, T)$ which is proved in the same way as the unique existence theorem in §2. (The time t_0 may be infinite. Indeed, if $a \equiv b \equiv 1$, and $h_i^0 < h_{i+1}^0$ for $i = 0, 1, \ldots, m-2$, then $u_0(x)$ itself is the solution of (3.1) with (3.7) and no plateau merges for all $t > 0$.) Q.E.D.

Remark 4. We did not use the fact that f in Lemma 2 is the canonical restriction of $\partial\varphi(u_0)$. By the Theorem on characterization of the speed we see that $f \in \partial\varphi(u_0)$ in Lemma 2 is actually equals $\partial\varphi^0(u_0)$ a posteriori.

In the rest of this subsection we shall prove Lemma 2.

Lemma 3 (Constant velocity profile). *Assume that a and b are positive and continuous on a nontrivial interval $[\alpha, \beta]$. The following two conditions are equivalent.*

(A) *$\{a(\beta) - a(\alpha)\}/\int_\alpha^\beta b(\tau)\, d\tau \leq \{a(x) - a(\alpha)\}/\int_\alpha^x b(\tau)\, d\tau$ for all $x \in (\alpha, \beta]$.*

(B) *For any* $\delta_1, \delta_2 \in \{-1, 1\}$ *there is a unique function* ξ *on* $[\alpha, \beta]$ *fulfilling the properties*:

$$a\xi \quad \text{is } C^1 \text{ on } [\alpha, \beta]; \tag{3.13}$$

$$|\xi(x)| \le 1 \quad \text{for all } x \in [\alpha, \beta]; \tag{3.14}$$

$$\xi(\alpha) = \delta_1, \quad \xi(\beta) = \delta_2; \tag{3.15}$$

$$\frac{1}{b}\frac{d}{dx}(a\xi)(x) = \frac{(a\xi)(\beta) - (a\xi)(\alpha)}{\int_\alpha^\beta b(\tau)\,d\tau} \quad \text{for all } x \in [\alpha, \beta]. \tag{3.16}$$

Proof. We first prove that (A) implies (B). We may assume that $\xi(\beta) = 1$ since the argument for $\xi(\beta) = -1$ is symmetric. We denote the constant $\{(a\xi)(\beta) - (a\xi)(\alpha)\}/\int_\alpha^\beta b$ by V. Integrating the equation $\partial_x(a\xi) = bV$, we have a representation formula for ξ

$$(a\xi)(x) - (a\xi)(\alpha) = V\int_\alpha^x b(\tau)\,d\tau.$$

The regularity of $a\xi$ is clear. It remains to check the constraint $|\xi(x)| \le 1$, which is equivalent to

$$-a(x) \le (a\xi)(\alpha) + V\int_\alpha^x b(\tau)\,d\tau \le a(x) \quad \text{for all } x \in [\alpha, \beta]. \tag{3.17}$$

If $\xi(\alpha) = 1$, the condition (A) is equivalent to the rightest inequality of (3.17). Since $\int_\alpha^x b \le \int_\alpha^\beta b$ so that $\{-a(x) - a(\alpha)\}/\int_\alpha^x b \le -a(\alpha)/\int_\alpha^\beta b \le \{a(\beta) - \alpha(\alpha)\}/\int_\alpha^\beta b$, the leftest inequality of (3.17) has been proved. If $\xi(\alpha) = -1$, the inequalities (3.17) read

$$\frac{-a(x) + a(\alpha)}{\int_\alpha^x b(\tau)\,d\tau} \le \frac{a(\beta) + a(\alpha)}{\int_\alpha^\beta b(\tau)\,d\tau} \le \frac{a(x) + a(\alpha)}{\int_\alpha^x b(\tau)\,d\tau}. \tag{3.18}$$

Since $\int_\alpha^x b \le \int_\alpha^\beta b$, the condition (A) implies

$$\frac{a(\beta) + a(\alpha)}{\int_\alpha^\beta b(\tau)\,d\tau} \le \frac{a(\beta) - a(\alpha)}{\int_\alpha^\beta b(\tau)\,d\tau} + \frac{2a(\alpha)}{\int_\alpha^\beta b(\tau)\,d\tau} \le \frac{a(x) - a(\alpha)}{\int_\alpha^x b(\tau)\,d\tau} + \frac{2a(\alpha)}{\int_\alpha^x b(\tau)\,d\tau}.$$

Thus we get the rightest inequality of (3.18). By the condition (A) we have

$$\frac{-a(x) + a(\alpha)}{\int_\alpha^x b(\tau)\,d\tau} \le \frac{a(\alpha) - a(\beta)}{\int_\alpha^\beta b(\tau)\,d\tau} \le \frac{a(\beta) + a(\alpha)}{\int_\alpha^\beta b(\tau)\,d\tau}$$

which implies the leftest hand of (3.18). We thus obtained (3.17).

The converse follows from (3.17) by taking $\xi(\alpha) = \xi(\beta) = 1$.

Q.E.D.

Lemma 4 (Constant velocity profile – boundary version). *Assume that a and b are positive and continuous on a nontrivial interval $[\alpha,\beta]$. Let p be a boundary point of $[\alpha,\beta]$, i.e., $p=\alpha$ or $p=\beta$. The following two conditions are equivalent.*

(i) *$a(p)\le a(x)$ for all $x\in[\alpha,\beta]$.*

(ii) *For any $\delta\in\{-1,1\}$, there is a unique function ξ on $[\alpha,\beta]$ fulfilling* (3.13), (3.14) *and $\xi(p)=\delta$, $\frac{d}{dx}(a\xi)=0$ on $[\alpha,\beta]$.*

Proof. There is a unique solution $\xi(x)=a(p)\delta/a(x)$ of $\xi(p)=\delta$, $d(a\xi)/dx=0$ on $[\alpha,\beta]$ (satisfying (3.13)). The condition (i) is equivalent to (3.14) so (i) and (ii) are equivalent. Q.E.D.

Proof of Lemma 2. We apply Lemma 1 and Lemma 3 on each interval $[x_i,x_{i+1}]$ $(i=1,2,\dots,m-2)$ (and Lemma 4 on $[x_0,x_1]$ and $[x_{m-1},x_m]$) to get Lemma 2.

The necessity of (C_i) in Remark 2 follows from Lemma 3 and Lemma 1. Similarly the necessity of $a(x)\ge a(x_1)$ on $[x_0,x_1]$, $a(x)\ge a(x_{m-1})$ on $[x_{m-1},x_m]$ in Remark 2 follows from Lemma 4. Q.E.D.

We conclude this subsection by proving that the concavity of a on $[x_i,x_{xi+1}]$ implies (C_i) when a is propotional to b as stated in Remark 3. It follows from the next lemma.

Lemma 5. *Assume that a is concave and positive on a nontrivial interval $[\alpha,\beta]$. Assume that a is proportional to b. Then*

$$\{a(\beta)-a(\alpha)\}\Big/\int_\alpha^\beta b(\tau)\,d\tau\le\{a(x)-a(\alpha)\}\Big/\int_\alpha^x b(\tau)\,d\tau \tag{3.19}$$

for all $x\in(\alpha,\beta]$.

Proof. We may assume that $a\equiv b$ on $[\alpha,\beta]$. We first discuss the case $a(\alpha)\le a(\beta)$. Let I be the interval of form

$$I=\{x\in[\alpha,\beta]\,;\,a(x)\ge a(\beta)\}$$

which may be a singleton (cf. Figure 2). Since $a(x)-a(\alpha)\ge 0$ and $\int_\alpha^x a\le\int_\alpha^\beta a$ for $x\in I$, (3.19) holds for all $x\in I$.

By concavity of a we have

$$\{a(\beta)-a(\alpha)\}/(\beta-\alpha)\le\{a(x)-a(\alpha)\}/(x-\alpha) \tag{3.20}$$

for all $x\in(\alpha,\beta]$. Moreover, if $x\notin I$, then

$$\frac{1}{x-\alpha}\int_\alpha^x a(\tau)\,d\tau\le\frac{1}{\beta-\alpha}\int_\alpha^\beta a(\tau)\,d\tau. \tag{3.21}$$

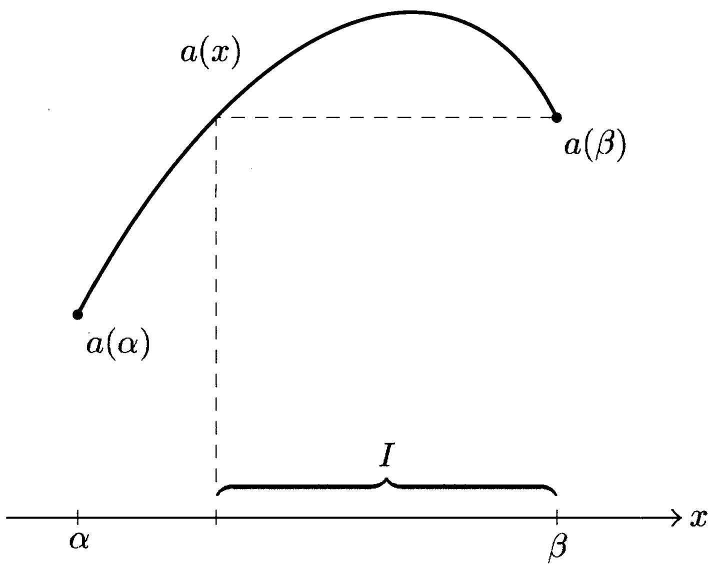


Fig. 2. Interval I defined in the proof of Lemma 5.

From (3.20) and (3.21) it follows that

$$\frac{a(\beta)-a(\alpha)}{\int_\alpha^\beta a(\tau)\,d\tau} = \frac{a(\beta)-a(\alpha)}{\beta-\alpha}\cdot\frac{\beta-\alpha}{\int_\alpha^\beta a(\tau)\,d\tau} \leq \frac{a(x)-a(\alpha)}{x-\alpha}\cdot\frac{x-\alpha}{\int_\alpha^x a(\tau)\,d\tau}$$

for $x \notin I$. Thus (3.19) holds for all $x \in (\alpha, \beta]$.

If $a(\alpha) \geq a(\beta)$ a symmetric argument yields

$$\{a(\beta)-a(\alpha)\}\Big/\int_\alpha^\beta b(\tau)\,d\tau \geq \{a(\beta)-a(x)\}\Big/\int_x^\beta b(\tau)\,d\tau$$

or

$$\{a_b(\tilde\beta)-a_b(0)\}/\tilde\beta \geq \{a_b(\tilde\beta)-a_b(y)\}/(\tilde\beta - y), \quad 0 \leq y < \tilde\beta$$

where $a_b(y) = a(x(y))$, $\tilde\beta = \int_\alpha^\beta b(\tau)\,d\tau$ and $x(y)$ is the inverse function of $y(x) = \int_\alpha^x b(\tau)\,d\tau$; see Remark 1. This evidently implies

$$\{a_b(\tilde\beta)-a_b(0)\}/\tilde\beta \leq \{a_b(y)-a_b(0)\}/y, \quad 0 < y \leq \tilde\beta$$

which is the same as (3.19). Q.E.D.

3.2.2. *Evolution at the time of plateau merging*

We consider the solution u of (3.1) with initial data u_0 of form (3.7) when $h_i^0 \neq h_{i-1}^0$ ($i = 1, 2, \ldots, m-1$). Under the assumptions of Theorem on persistency concerning a and b it follows from Lemma 1 and Lemma 2 that

$$u(t,x) = \nu^i t + h_i^0 \,(= h_i(t)) \quad \text{on } (x_i, x_{i+1}) \ (i = 0, 1, \ldots, m-1)$$

until the first time t_0 at which some plateau merges, i.e., t_0 is the first time t that $h_i(t) = h_{i+1}(t)$ for some i, $0 \le i \le m-1$.

At time t_0 some consecutive plateau merges. We shall discuss whether or not these merged plateaus are forced to split for $t > t_0$ close to t_0. The answer depends on a and b. For this purpose we shall extend Lemma 3 to calculate velocity profile on merged plateaus.

Lemma 6. *Assume that a and b are positive and continuous on $[\alpha, \beta]$ with $\alpha = r_0 < r_1 < r_2 < \cdots < r_k = \beta$, where $k \ge 1$. Assume that (C_i) in Theorem on persistency with x_i replaced by r_i for all $i = 0, 1, \ldots, k-1$.*

(i) (*Monotone velocity profile*) *There is a unique function ξ on $[\alpha, \beta]$ fulfilling* (3.14), (3.15) *with $\delta_1 = \delta_2 \in \{-1, 1\}$ such that the following properties* (a)–(c) *hold.*

(a) *$a\xi$ is continuous on $[\alpha, \beta]$ and C^1 as a function on each (r_i, r_{i+1}) $(i = 0, 1, \ldots, k-1)$.*

(b) *$v(x) = \frac{1}{b(x)} \frac{d}{dx}(a\xi)(x)$ is constant on each interval (r_i, r_{i+1}) $(i = 0, 1, \ldots, k-1)$ and $v(x)$ is nondecreasing for $\delta_1 = 1$ (nonincreasing for $\delta_1 = -1$) as a function of x in (α, β) outside the set of jump discontinuities $\Sigma \subset \{r_1, \ldots, r_{k-1}\}$ of v.*

(c) *$\xi(r_i) = \delta_1$ for $r_i \in \Sigma$ and $i = 1, \ldots, k-1$.*

(ii) (*One peak velocity profile*) *There is a unique function ξ on $[\alpha, \beta]$ fulfilling* (3.14), (3.15) *with $\delta_1 = -\delta_2 \in \{-1, 1\}$ and* (a) *such that the following properties* (d), (e) *hold.*

(d) *$v(x) = \frac{1}{b(x)} \frac{d}{dx}(a\xi)(x)$ is constant on each interval (r_i, r_{i+1}) $(i = 0, 1, \ldots, m-1)$ and $v(x)$ is nondecreasing for $\delta_1 = 1$ (nonincreasing for $\delta_1 = -1$) on $(\alpha, (r_j + r_{j+1})/2) \backslash \Sigma'$ and nonincreasing for $\delta_1 = 1$ (nondecreasing for $\delta_1 = -1$) on $((r_j + r_{j+1})/2, \beta) \backslash \Sigma'$ for some $j \in \{0, 1, \ldots, k-1\}$, where Σ' is the set of jump discontinuities of v.*

(e) *$\xi(r_i) = \delta_1$ for $i = 0, 1, \ldots, j$ and $\xi(r_i) = \delta_2$ for $i = j+1, \ldots, k$ provided that $r_i \in \Sigma'$.*

Remark 5. If ξ is constructed for $\delta_1 = \delta_2 = 1$ (resp. $\delta_1 = -\delta_2 = 1$) then $-\xi$ is the desired ξ for $\delta_1 = \delta_2 = -1$ (resp. $\delta_1 = -\delta_2 = -1$).

Remark 6. To prove Lemma 6 (as well as Lemmas 3 and 4) we may assume that $b \equiv 1$ by changing coordinate and replacing a by a_b as defined in Remark 1.

Remark 7. To prove Lemma 6 it is helpful to give an elementary interpretation of the condition $|\xi| \le 1$. Assume that $(a\xi)_x =: p$ is a constant on $(\alpha_1, \beta_1) \subset [\alpha, \beta]$. Then $|\xi| \le 1$ on $[\alpha_1, \beta_1]$ if and only if $|\eta| \le a$ on $[\alpha_1, \beta_1]$, where η is affine with slope p and $\eta(a_1) = (a\xi)(\alpha)$, $\eta(\beta_1) = (a\xi)(\beta_1)$, i.e., $\eta(x) = p(x - \alpha_1) + (a\xi)(\alpha_1)$.

Proof. (i) By Remark 5 we may assume $\delta_1 = \delta_2 = 1$. By Remark 6 we may assume that $b \equiv 1$. Let ζ be the convex hull of a in $[\alpha, \beta]$. By (C_i) $(i = 0, 1, \dots, k-1)$, ζ is a piecewise linear convex function whose jumps of derivatives are contained in $\{r_i\}_{i=1}^{k-1}$. If we take $\xi = \zeta/a$, then ξ satisfies all desired properties. Indeed, $v = d\zeta/dx$ is piecewise constant and nondecreasing since ζ is convex. Also $|\xi| \le 1$ is fulfilled (by Remark 7), since $0 \le \zeta \le a$. The set Σ should be the set of jump discontinuities of $d\zeta/dx$. For $x \in \Sigma$ and $x = \alpha$, $x = \beta$ we see $\zeta(x) = a(x)$ so that $\xi(x) = 1$.

To see uniqueness for given ξ satisfying (3.14), (3.15) and (a)–(c) we set $\tilde{\zeta} = a\xi$. By (b), $\tilde{\zeta}$ must be convex and piecewise linear. For $x \in \Sigma$ and $x = \alpha, \beta$ we see $\tilde{\zeta}(x) = a(x)$, by (c) and (3.15). By (3.14) $\tilde{\zeta} \le a$ on $[\alpha, \beta]$. Thus $\tilde{\zeta}$ must be the convex hull of a so the uniqueness of ξ has been proved.

(ii) We may assume that $\delta_1 = -\delta_2 = 1$ and $b \equiv 1$. Let ζ^+ be the convex hull of a in $[\alpha, \beta]$ and set $\zeta^- = -\zeta^+$. We would like to construct a new piecewise linear function ζ such that

1° $\zeta^- \le \zeta \le \zeta^+$ on $[\alpha, \beta]$,

2° $\zeta = \zeta^+$ on $[\alpha, r_\ell]$ and $\zeta = \zeta^-$ on $[r_\sigma, \beta]$ for some ℓ, σ satisfying $0 \le \ell < \sigma \le k$,

3° ζ is linear (affine) in $[r_\ell, r_\sigma]$ and its slope $V = d\zeta/dx$ fulfills $V \ge v_+(x)$ on $[\alpha, r_\ell]$ and $V \ge v_-(x)$ on $[r_\sigma, \beta]$ where $v_\pm = d\zeta^\pm/dx$.

Such a function ζ is easy to construct. Indeed, we set

$$v_*(x) = \min(v_+(x), v_-(x)) \quad \text{and} \quad \lambda = \min_{[\alpha,\beta]} \zeta^+ \, (> 0)$$

and find that there is a unique negative number V fulfilling

$$\int_\alpha^\beta \max\{v_*(x) - V, 0\}\, dx = 2\lambda.$$

If we set

$$\zeta(x) = \int_\alpha^x \min(v_*(z), V)\, dz + \zeta^+(\alpha),$$

then ζ satisfies 1°, 2°, 3° by setting

$$r_\ell = \max\{r_i \,;\, v_+(r_i - 0) \le V\}, \quad r_\sigma = \min\{r_i \,;\, v_-(r_i + 0) \le V\}.$$

For notational convenience we set $v_+(r_0 - 0) = -\infty$ and $v_-(r_k + 0) = -\infty$. (The choice of V is important so that $\zeta(r_\sigma) = \zeta^-(r_\sigma)$.) We do not present the proof of 1°, 2°, 3° since it is elementary. Instead, we present an example of the graph of v_* and the value V in Figure 3 (a) and the graph of $\zeta^\pm$ and ζ when $\max_x(\min(v_+(x), 0)) > 0 > \max_x(\min(v_-(x), 0))$ in Figure 3 (b).

We set $\xi = \zeta/a$ and observe that ξ satisfies all desired properties. For example, by Remark 7 $|\xi| \le 1$ is fulfilled since $\zeta^- \le \zeta \le \zeta^+$ implies $|\zeta| \le a$.

For given ξ satisfying (3.14), (3.15) and (d), (e) it is easy to see that $a\xi = \zeta$ fulfills 1°, 2°, 3°. Since ζ satisfying 1°, 2°, 3° is unique, so is ξ. Q.E.D.

There is a boundary version corresponding to Lemma 6 (i) but we do not state it explicitly. By Lemma 1 and Lemma 6 together its boundary version one is able to find an element $f \in \partial\varphi(u_0)$, (which is piecewise constant) such that $u(x,t) = -f(x)t + u_0(x)$ is a solution of (3.7) with $u|_{\partial\Omega} = u_0|_{\partial\Omega}$ for small $t > 0$. By the uniqueness of a solution Theorem on persistency is generalized as follows.

General theorem on persistency. *Assume the same hypotheses of Theorem on persistency concerning a and b. Let u be the solution of* (3.1) *with initial data u_0 given by* (3.7) *with $u|_{\partial\Omega} = u_0|_{\partial\Omega}$. Then for each $i = 0, 1, \ldots, m-1$ the speed $u_t(x,t)$ is independent of $x \in (x_i, x_{i+1})$ and $t \in (0, t_0)$, where t_0 is the first time that some plateau of $u(\,\cdot\,, t)$ merges to another one. Moreover, $u_t(x,t) \equiv 0$ in (x_0, x_1) and (x_{m-1}, x_m) for $0 < t < t_0$.*

Note that some plateau $(x_i, x_{i+k}) \times \{h_i^0\}$, $k \ge 2$ at $t = 0$ may break instantaneously or stay as a plateau. By the above theorem we are able to calculate solution u explicitly by calculating the speed based on Lemma 6 until the time t_0 when some plateau merges. We calculate new speed for $u(x, t_0)$ and use above theorem to find explicit value u until the time t_1 when another plateau merges. We repeat this procedure to calculate u globally in time. Because of monotone nature of velocity profile in Lemma 6 we see that the number of peaks of u does not

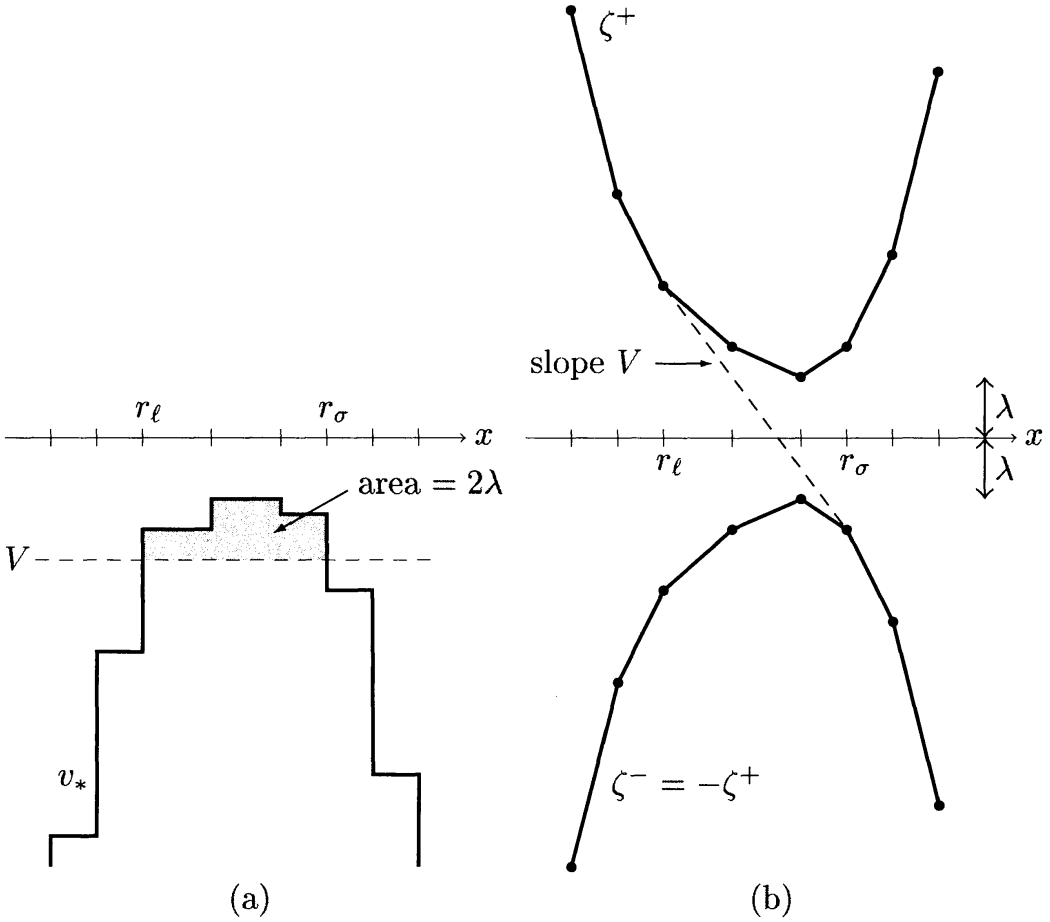


Fig. 3. (a) The graph of v_* and the value V. (b) The graph of $\zeta^{\pm}$. The dotted line indicates a part of the graph of ζ.

increase. For a piecewise constant function u_0 given by (3.7) with $h_i^0 \neq h_{i-1}^0$ $(i = 1, \dots, m-1)$ let $M(u_0)$ (resp. $\mu(u_0)$) denote number of local maximum (resp. minimum) of function $i \mapsto h_i^0$ defined in $\{0, \dots, m-1\}$. (For such a function $i_0 \in \{1, \dots, m-2\}$ is called a local maximum (resp. minimum) if $h_{i_0 \pm 1}^0 < h_{i_0}^0$ (resp. $h_{i_0 \pm 1} > h_{i_0}$).) If $m = 2$, we set $M(u_0) = \mu(u_0) = 0$. If $h_i^0 = h_{i-1}^0$ for some i we renumber x_i's to identify u_0 of form (3.7) with $h_i^0 \neq h_{i-1}^0$ $(i = 1, \dots, m'-1)$ for some $m' < m$. For such identification we define $M(u_0)$ and $\mu(u_0)$.

Theorem on nonincrease of peaks. *Under the same hypotheses and notations of General Theorem on persistency we have*

$$M(u(\cdot\,, t)) \text{ and } \mu(u(\cdot\,, t)) \text{ are constant on } (0, t_0).$$

Moreover, for all $t \geq 0$ *we have*

$$M(u(\,\cdot\,,t)) \leq M(u_0), \quad \mu(u(\,\cdot\,,t)) \leq \mu(u_0).$$

Remark 8. If $b \equiv 1$, it is enough to consider convexification of a on $[x_i, x_{i+k}]$ (instead of original a) to see whether a plateau $(x_i, x_{i+k}) \times \{h_i^0\}$ $(k \geq 2)$ is forced to break instanteneously when $h_{i-1}^0 < h_i^0 < h_{i+k+1}^0$ or $h_{i-1}^0 > h_i^0 > h_{i+k+1}^0$. However, if $b \not\equiv 1$ a simple convexification does not give a right answer. We have to convexify a_b instead of a (cf. Remark 1).

In the following, several numerical simulations will be demonstrated which support our theoretical results. For these calculations, we approximate (3.1) by $u_t = b^{-1}(a\chi_\gamma(u_x)u_x)_x$, $\chi_\gamma(p) = (\tanh \gamma p)/p$ for large γ and adopt the numerical scheme introduced in [KG]. This approximation is justified by the stability theorem. Note that our numerical scheme does not assume any persistency properties of plateaus. Let us give the graph of $a(x)$ by connecting $(0,1)$, $(1/6, a_1)$, $(2/6, 1)$, $(3/6, a_2)$, $(4/6, 1)$, $(5/6, a_3)$ and $(1,1)$ in this order as shown in Figure 4 (a). Here a_1, a_2 and a_3 are assumed to satisfy the relation $0 < a_1 < a_2 < a_3 < 1$. Related to the profile of $a(x)$, the sequence $\{x_i\}$ is given as follows; $x_0 = 0$, $x_1 = 1/6$, $x_2 = 3/6$, $x_3 = 5/6$ and $x_4 = 1$. We take the initial data with $h_1(0) = h_2(0)$ as shown in Figure 4 (b), and the global minimizer of φ defined by (2.5) is indicated in Figure 4 (c). The problem is whether the plateau $(x_1, x_3) \times \{h_1(0)\}$ is broken or not during the transition from the initial state to the final.

We present two simulations in Figure 5 by taking $b(x) \equiv 1$. If a_i's are selected so that the point (x_2, a_2) locates above the line segment connecting (x_1, a_1) and (x_3, a_3), the plateau is kept unbroken as indicated in Figure 5 (c). If it locates below, the plateau splits into the two plateaus as shown in Figure 5 (d). These results assures the validity of the convexity check stated in Remark 8.

However, the convexity check does not always give the right answer if $b(x)$ is not constant. For example, define $a(x)$ by connecting $(0,1)$, $(1/6, 0.2)$, $(2/6, 1)$, $(3/6, 0.401)$, $(4/6, 0.6)$, $(5/6, 0.6)$ and $(1,1)$ (Figure 6 (a) and (b)) and take the same initial data as the one in the previous simulation. According to the convexity check, the plateau should not split as long as $b(x) \equiv 1$ is adopted as is shown in Figure 6 (c). On the other hand, the condition (C_i) is violated on the interval (x_1, x_3) for $x = x_2$ if $b(x) \equiv a(x)$ is assumed. Therefore the plateau must be broken, which is also confirmed in Figure 6 (d).

We present one more example with $b(x) \equiv 1$ and $a(x)$ given in Figure 7 (a), and see what will happen to the big plateau (consists of

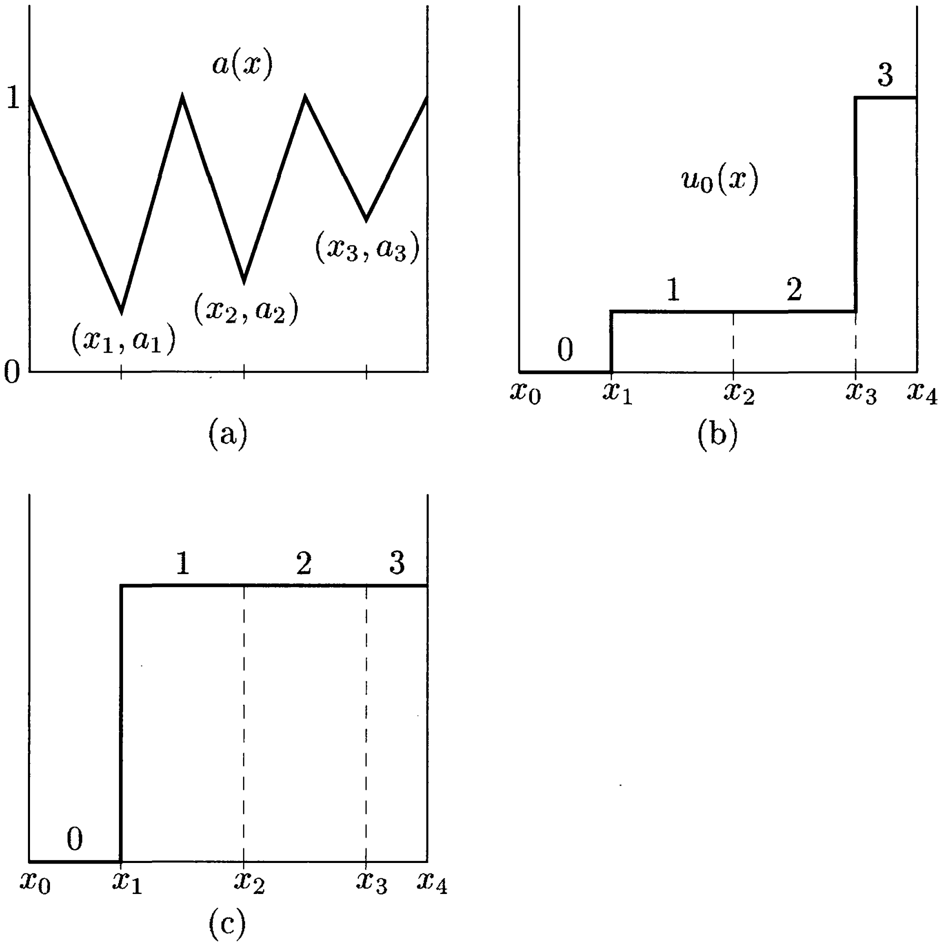


Fig. 4. (a) Piecewise linear profile of $a(x)$ with the three local minima. (b) Initial data $u_0(x)$ given by $h_0(0) = 0.0$, $h_1(0) = h_2(0) = 0.2$ and $h_3(0) = 1.0$. (c) The final state (global minimizer). Small numbers on each segment indicate the index i of $h_i(t)$.

seven segments) of the initial data (Figure 7 (b)). Graphical checking tells us that the plateau will be broken into four pieces, and it is actually observed in the simulation as shown in Figure 7 (c) and (d).

§4. Formation of jumps

We consider (3.1) with $b \equiv 1$, i.e.,

$$u_t = \partial_x(a(x)u_x/|u_x|) \quad \text{on } \Omega \times (0, \infty) \tag{4.1}$$

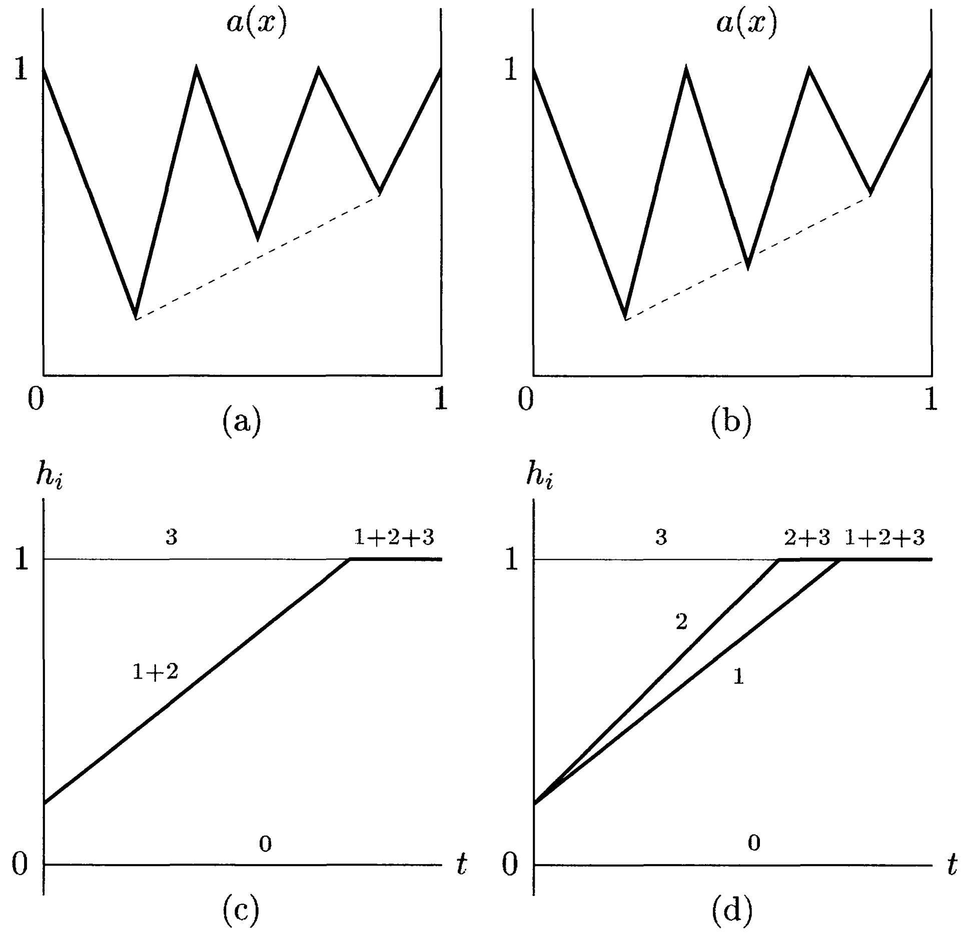


Fig. 5. (a) The function $a(x)$ is given by $a_1 = 0.2$, $a_2 = 0.42$ and $a_3 = 0.6$, and $b(x) \equiv 1$. (b) The function $a(x)$ is given by $a_1 = 0.2$, $a_2 = 0.38$ and $a_3 = 0.6$, and $b(x) \equiv 1$. (c) Time evolution of all the segments for $a(x)$ given in (a) and $b(x) \equiv 1$. (d) Time evolution of all the segments for $a(x)$ given in (b) and $b(x) \equiv 1$.

with initial data u_0 and the boundary condition $u|_{\partial\Omega} = u_0|_{\partial\Omega}$, where Ω is a bounded open interval. The regularity property of solutions of (4.1) is different from that for $u_t = \partial_x(a(x)u_x)$. For example, for the latter equation if $a(x)$ is Hölder continuous on $\overline{\Omega}$ (and $a > 0$ on $\overline{\Omega}$) then solution u is C^2 in x and C^1 in time. For the problem (4.1) a jump discontinuity of solutions may be formed instantaneously. We give such an example and discuss other properties of solutions.

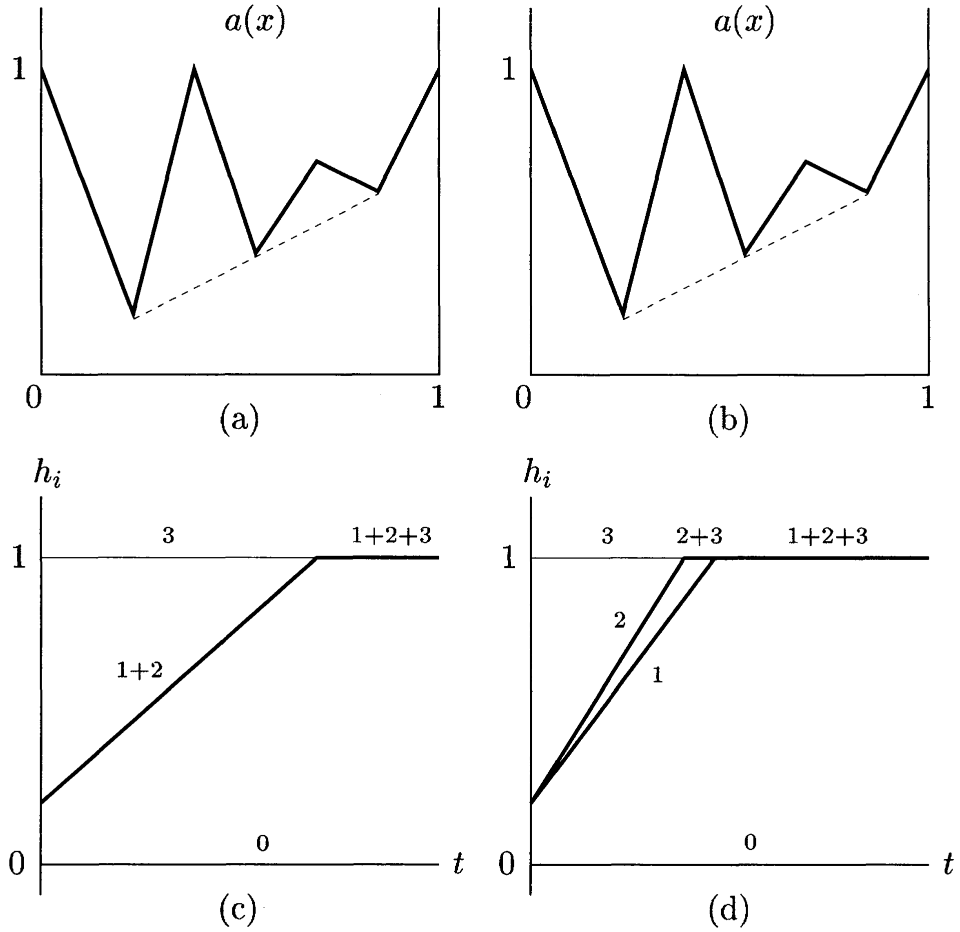


Fig. 6. (a) The point (x_2, a_2) is slightly above the line segment connecting (x_1, a_1) and (x_3, a_3) although it is hard to judge by the figure. (b) Exactly the same graph with (a). (c) Time evolution of all the segments for $a(x)$ given in (a) and $b(x) \equiv 1$. (d) Time evolution of all the segments for $a(x)$ given in (b) and $b(x) \equiv a(x)$.

Example 4.1. Assume that $a(x) = a_0 + |x|$, $x \in \Omega = (-1, 1)$, $a_0 > 0$. If $u_0(x) = x$, then the solution u of the initial-boundary value problem for (4.1) is of form

$$u(x,t) = \begin{cases} \min\{1, x+t\}, & x > 0, \\ \max\{-1, x-t\}, & x < 0. \end{cases}$$

Evidently, $u(x,t)$ has a jump at $x = 0$ for $t > 0$. Moreover, $u(x,t)$ becomes the global minimizer sgn x of energy φ at $t = 1$ and $u(x,t) =$

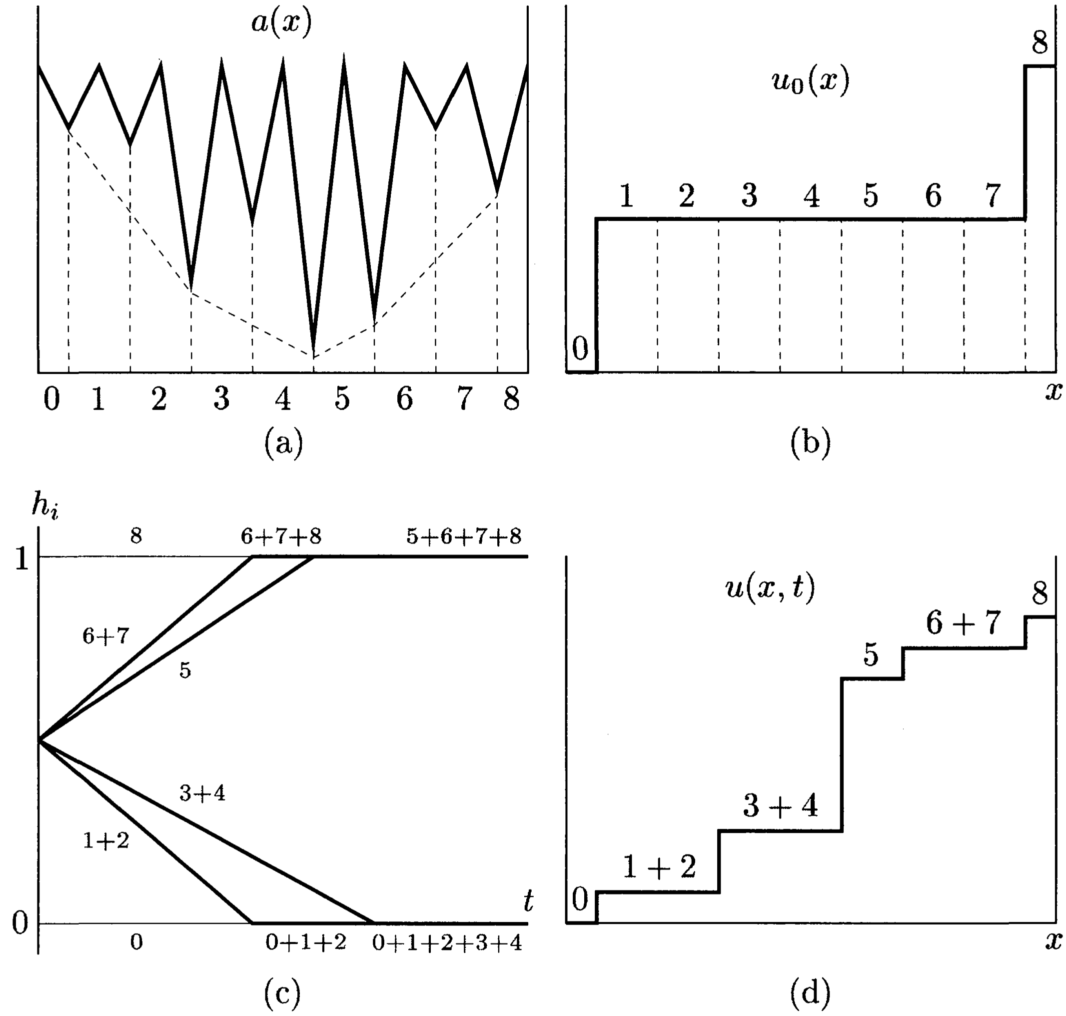


Fig. 7. (a) The sequence $\{x_i\}$ is given by $x_0 = 0$, $x_i = (i - 1/2)/8$ $(i = 1, \dots, 8)$, $x_9 = 1$ and a_i's are appropriately chosen. (b) Initial data given by $h_0(0) = 0$, $h_1(0) = \cdots = h_7(0) = 1/2$, $h_8(0) = 1$. (c) Time evolution of all the segments for $a(x)$ given in (a) and $b(x) \equiv 1$. (d) Snapshot of $u(x,t)$ for $t = 0.21$.

$\operatorname{sgn} x$ for $t \geq 1$. For general initial data u_0 the convergence property to the global minimizer for given initial data holds if $u_0(-1) \neq u_0(+1)$ as stated below.

Theorem. *Assume that* $a(x) = a_0 + |x|$, $x \in \Omega = (-1,1)$, $a_0 > 0$. *Assume that* $u_0 \in BV(\Omega)$ *with* $u_0(-1) \neq u_0(+1)$. *Then the solution* u *of the initial-boundary value problem for* (4.1) *with* $u|_{t=0} = u_0$ *and*

$u|_{\partial\Omega} = u_0|_{\partial\Omega}$ *converges to*

$$U(x) = \begin{cases} u_0(+1), & x > 0, \\ u_0(-1), & x < 0 \end{cases}$$

in a finite time.

Proof. By symmetry we may assume that $u_0(-1) < u_0(+1)$. We may assume that $u_0(-1) = -1$, $u_0(+1) = 1$ so that $U(x) = \operatorname{sgn} x$ by adding and multiplying a constant to u. By assumption u_0 is bounded, i.e., there is a positive constant $M > 1$ such that $|u_0(x)| \leq M$ for all $x \in \Omega$. We note that

$$v(x,t) = \begin{cases} \max\{M - (1+a_0)t, 1\}, & 0 < x < 1, \\ \max\{M - (1+a_0)t + a_0\big(t - \frac{M-1}{1+a_0}\big)_+, -1\}, & -1 < x < 0 \end{cases}$$

is the solution of the initial-boundary value problem for (3.1) with the boundary condition $u(\pm 1, t) = \pm 1$ and initial condition $u|_{t=0} = M$. This can be proved as in §3.2 if we pay attention that v has a jump at $x = \pm 1$. (The solution v is a typical example that the initial data is incompatible with the boundary condition.) Thus we see $u(x,t) \leq U(x)$ in a finite time by the comparison principle. Here $p_+ = \max(p, 0)$. A symmetric argument implies $u(x,t) \geq U(x)$ in a finite time. Q.E.D.

Example 4.2. Assume that $a(x) = a(-x) > 0$, $x \in (-1,1)$ and that a is Lipschitz continuous. Assume that a is C^1 and nondecreasing in x for $x > 0$. If $u_0(x) = x$, then the solution u of the initial-boundary value problem for (4.1) is of the form

$$u(x,t) = \begin{cases} \min\{1, x + \big(\frac{da}{dx}(x)\big)t\}, & x > 0, \\ \max\{-1, x + \big(\frac{da}{dx}(x)\big)t\}, & x < 0 \end{cases}$$

which generalizes Example 4.1. If a is C^1 at the origin, $u(x,t)$ stays continuous for $t \geq 0$ since $da/dx \to 0$ as $x \to 0$. It tends to $U(x)$ in the preceeding theorem as $t \to \infty$, but $u(x,t) \not\equiv U(x)$ for any finite t. It is possible to prove a convergence result to $U(x)$ as $t \to \infty$ for a general initial data with $u_0(-1) \neq u_0(+1)$ under suitable assumptions on a, however, we do not state it here. Instead, we present several numerical calculations.

The simulation of Example 4.1 is indicated in Figure 8, and Figure 9 corresponds to Example 4.2 with $a(x) = a_0 + x^2$. In both cases, the numerical solutions approximate the exact solutions given above quite well.

Note added in the proof. In a recent preprint "The Dirichlet problem for the total variation flow" by F. Andreu, C. Ballester, V. Caselles and J. M. Mazón L^1 framework for (2.7) with $a \equiv b \equiv 1$ is established instead of L^2 framework given in this paper.

The authors are grateful to Professor Naoyuki Ishimura for valuable comments.

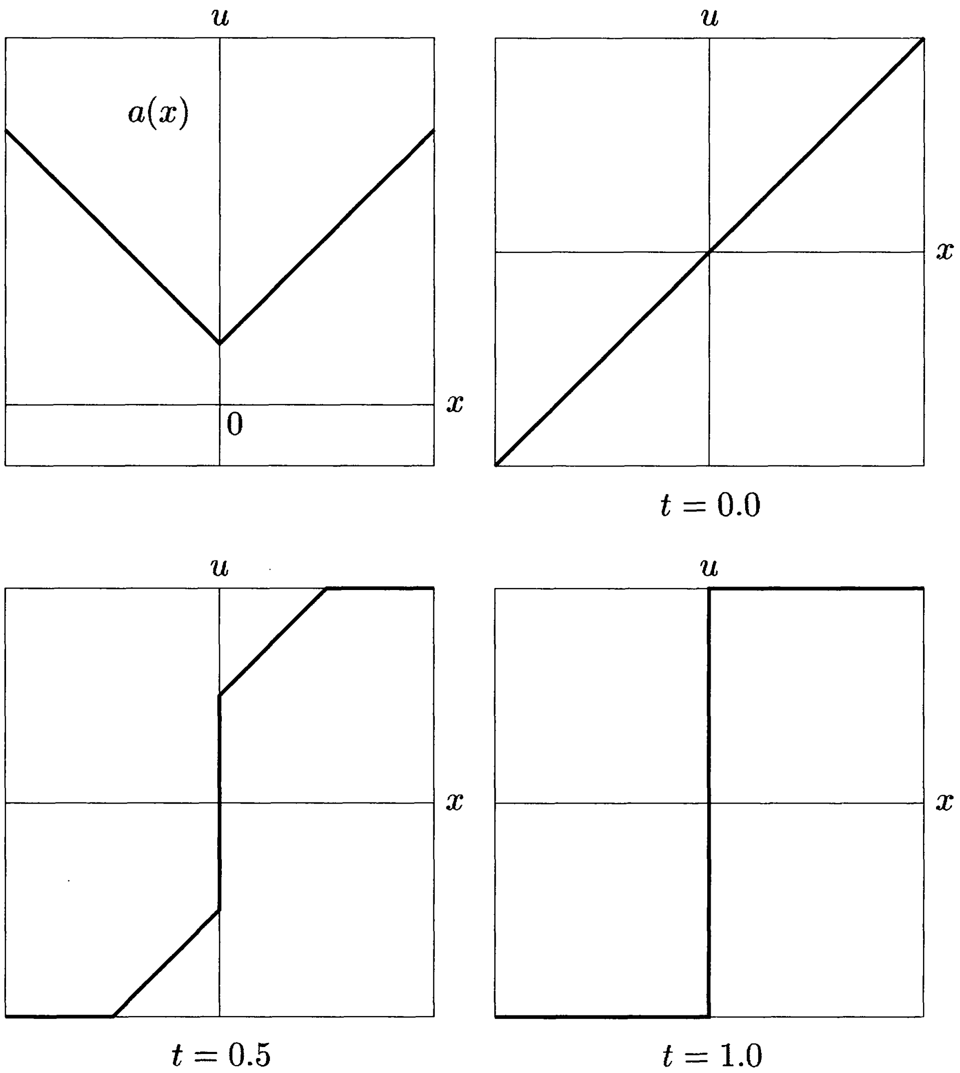


Fig. 8. Simulation of Example 4.1 with $a(x) = 0.2+|x|$, $b(x) \equiv 1$ and $u_0(x) = x$. Discontinuity appears instantaneously and the solution reaches to the final state in a finite time.

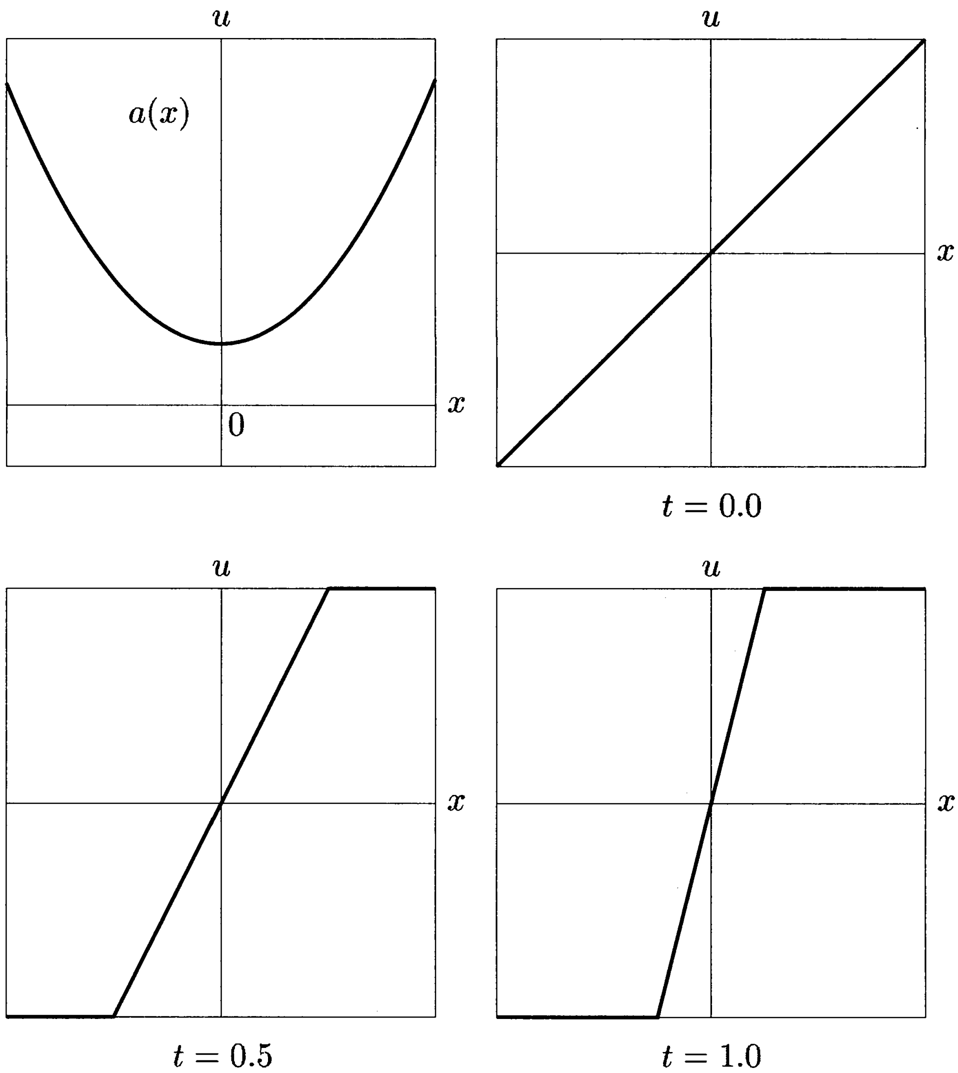


Fig. 9. Simulation of Example 4.2 with $a(x) = 0.2 + x^2$, $b(x) \equiv 1$ and $u_0(x) = x$. Discontinuity never appears and the solution converges to the final state while it doesn't reach the final state in a finite time.

References

[A] S. B. Angenent, The zeroset of a solution of a parabolic equation, J. Reine Angew. Math., **390** (1988), 79–96.

[AG] S. B. Angenent and M. E. Gurtin, Multiphase thermomechanics with interfacial structure 2. Evolution of an isothermal interface, Arch. Rational Mech. Anal., **108** (1989), 323–391.

[AD] H. Attouch and A. Damlamian, Application des methods de convexité et monotonie a l'etude de certaines equations quasi lineares, Proc. Roy. Soc. Edinburgh, **79A** (1977), 107–129.

[Ba] V. Barbu, "Nonlinear Semigroups and Differential Equations in Banach Space", Noordhoff, Groningen, 1976.

[BP] H. Brezis and A. Pazy, Convergence and approximation of semigroups of nonlinear operators in Banach spaces, J. Funct. Anal., **9** (1972), 63–74.

[ESG] C. M. Elliott, A. R. Gardiner and R. Schätzle, Crystalline curvature flow of a graph in a variational setting, Adv. in Math. Sci, Appl., **8** (1998), 425–460.

[CGG] Y.-G. Chen, Y. Giga and S. Goto, Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations, J. Diff. Geom., **33** (1991), 749–786.

[D] A. Damlamian, "Non Linear Evolution Equations with Variable Norms", Thesis, Harvard University, 1974.

[ES] L. C. Evans and J. Spruck, Motion of level sets by mean curvature I, J. Diff. Geom., **33** (1991), 635–681.

[FG] T. Fukui and Y. Giga, Motion of a graph by nonsmooth weighted curvature, in "World Congress of Nonlinear Analysis '92", (V. Lakshmikantham, ed.), de Gruyter, Berlin, vol I, 1996, pp. 47–56.

[G] Y. Giga, A level set method for surface evolution equations, Sugaku Expositions, **10** (1997), 217–241.

[GG1] M.-H. Giga and Y. Giga, Evolving graphs by singular weighted curvature, Arch. Rational Mech. Anal., **141** (1998), 117–198.

[GG2] M.-H. Giga and Y. Giga, A subdifferential interpretation of crystalline motion under nonuniform driving force, Proc. of the International Conference on Dynamical Systems and Differential Equations, Springfield, Missouri (1996); in "Dynamical Systems and Differential Equations" (W.-X. Chen and S.-C. Hu, eds.), Southwest Missouri State Univ. vol. **1**, 1998, pp. 276–287.

[GG3] M.-H. Giga and Y. Giga, Stability for evolving graphs by nonlocal weighted curvature, Commun. in Partial Differential Equations, **24** (1999), 109–184.

[GG4] M.-H. Giga and Y. Giga, Motion driven by nonlocal curvature in the plane, in "Asymptotics in Nonlinear Diffusive Systems" (Nishiura et al, eds.), Tohoku Math. Publ., **8** (1998), pp. 75–83.

[Giu] E. Giusti, "Minimal Surfaces and Functions of Bounded Variations", Birkhauser, Boston, 1984.

[HZ] R. Hardt and X. Zhou, An evolution problem for linear growth functionals, Commun. in Partial Differential Equations, **19** (1994), 1879–1907.

[KG] R. Kobayashi and Y. Giga, Equations with singular diffusivity, J. Stat. Phys., **95** (1999), 1187–1220.

[KWC] R. Kobayashi, A. Warren and Carter, Modeling grain boundaries using a phase field technique, Hokkaido Univ. Preprint Series in Math., **422** (1998).

[Ko] Y. Kōmura, Nonlinear semi-groups in Hilbert space, J. Math. Soc. Japan, **19** (1967), 493–507.

[M] H. Matano, Nonincrease of the lapnumber for a solution of a one-dimensional semilinear parabolic equation, J. Fac. Sci. Univ. Tokyo, Sect. IA **29** (1982), 401–441.

[T] J. Taylor, Constructions and conjectures in crystalline curvature, including triple junctions and boundary points, in "Differential Geometry: Partial Differential Equations on Manifolds", (R. Greene and S. T. Yau, eds.), Proc. Symp. Pure Math., **54**, Part I, 1993, pp. 417–438. Amer. Math. Society, Providence.

[Te] R. Temam, Solutions généralisées de certain équations du type hypersurfaces minimals, Arch. Rational Mech. Anal., **44** (1991), 121–156.

[V] A. Visintin, "Models of Phase Transitions", Birkhäuser, 1996.

[W] J. Watanabe, Approximation of nonlinear problems of a certain type in "Numerical Analysis of Evolution Equations", (H. Fujita and M. Yamaguti, eds.), Lecture Notes in Num. Appl., **1**, Kinokuniya, Tokyo, 1979, pp. 147–163.

Mi-Ho Giga and Yoshikazu Giga
Department of Mathematics
Hokkaido University
Sapporo, 060-0810
Japan

Ryo Kobayashi
Research Institute for Electronic Science
Hokkaido University
Sapporo, 060-0810
Japan

Advanced Studies in Pure Mathematics 31, 2001
Taniguchi Conference on Mathematics Nara '98
pp. 127–145

The Topology of Real and Complex Algebraic Varieties

János Kollár

Contents

§1. Introduction

Definition 1.1. Let $f_i(x_1, \ldots, x_n)$ be polynomials whose coefficients are real or complex numbers. An *affine algebraic variety* is the common zero set of finitely many such polynomials

$$X = X(f_1, \ldots, f_k) := \{\mathbf{x} \mid f_i(\mathbf{x}) = 0, \forall i\}.$$

To be precise, I also have to specify where the variables x_i are. If the f_i have complex coefficients then the only sensible thing is to let the x_i be complex. The resulting topological space is

$$X(\mathbb{C}) := \{\mathbf{x} \in \mathbb{C}^n \mid f_i(\mathbf{x}) = 0, \forall i\},$$

which we always view with its Euclidean topology. If the f_i have real coefficients then we can let the x_i be real or complex. Thus we obtain two "incarnations" of a variety

$$X(\mathbb{R}) := \{\mathbf{x} \in \mathbb{R}^n \mid f_i(\mathbf{x}) = 0, \forall i\}, \quad \text{and}$$
$$X(\mathbb{C}) := \{\mathbf{x} \in \mathbb{C}^n \mid f_i(\mathbf{x}) = 0, \forall i\}.$$

Received April 21, 1999.
Partial financial support was provided by the Taniguch Foundation and by the NSF under grant number DMS-9622394.

Again, both of these are topological spaces where the topology is induced by the Euclidean topology on $\mathbb{R}^n$ or $\mathbb{C}^n$.

It is frequently inconvenient that $X(\mathbb{C})$ is essentially never compact. To remedy this, we introduce projective varieties which are closed subsets of the projective n-space $\mathbb{P}^n$. Since the coordinates of a point in $\mathbb{P}^n$ are defined only up to a scalar multiple, the zero set makes sense only for homogeneous polynomials. Given any number of homogeneous polynomials $F_i(x_0, \dots, x_n)$ we obtain the corresponding *projective variety*

$$X = X(F_1, \dots, F_k) := \{\mathbf{x} \in \mathbb{P}^n \mid F_i(\mathbf{x}) = 0, \forall i\}.$$

As before, we can look at the set of real or complex points

$$\begin{aligned} X(\mathbb{R}) &:= \{\mathbf{x} \in \mathbb{RP}^n \mid F_i(\mathbf{x}) = 0, \forall i\}, \quad \text{and} \\ X(\mathbb{C}) &:= \{\mathbf{x} \in \mathbb{CP}^n \mid F_i(\mathbf{x}) = 0, \forall i\}. \end{aligned}$$

Basic Question 1.2. My main interest is to establish connections between the algebraic properties of X and the topological properties of $X(\mathbb{C})$ and $X(\mathbb{R})$. There are two main direction that one can follow.

Determining which topological spaces can be obtained as $X(\mathbb{C})$ or $X(\mathbb{R})$ is called the *realization problem.* One may also ask for realizations where there is a strong connection between various algebraic and topological properties.

We may also want to know which algebraic properties of X are determined by topological properties of $X(\mathbb{C})$ or $X(\mathbb{R})$. This is the *recognition problem.* Ideally we would like to have a way of computing algebraic invariants from topology.

One can also say that the recognition problem is about obstructions to the realization problem. A recognition result leaves us with less freedom in the realization problem.

The following example illustrates the general features of the recognition problem, which is the main focus of these notes.

Example 1.3. Let $F(x_0, \dots, x_n)$ be a real, homogeneous polynomial of degree d and set $X_F := (F = 0) \subset \mathbb{P}^n$. The basic algebraic invariant of F is its degree $\deg F$. The simplest form of the recognition problem asks if $\deg F$ is determined by $X_F(\mathbb{C})$ or $X_F(\mathbb{R})$.

In this generality the answer is no. Indeed, F and F^2 have the same zero sets but different degrees. Thus it is sensible to assume to start with that F is irreducible. With this assumption the degree is easy to read off from topological data.

$H_{2n-2}(\mathbb{CP}^n, \mathbb{Z}) \cong \mathbb{Z}$ where the generator is given by the hyperplane class $[H]$. It is easy to see that $X_F(\mathbb{C}) \subset \mathbb{CP}^n$ has a homology class $[X_F(\mathbb{C})] \in H_{2n-2}(\mathbb{CP}^n, \mathbb{Z})$ and $[X_F(\mathbb{C})] = \deg F \cdot [H]$.

It is somewhat harder to obtain $\deg F$ from $X_F(\mathbb{C})$ alone. First of all, there are some exceptions. For instance, if $G = x_0^{d-1}x_2 - x_1^d$ then $X_G(\mathbb{C})$ is homeomorphic to S^2. This is caused by the fact that $X_G(\mathbb{C})$ is not a submanifold of $\mathbb{CP}^2$ near the point $(0 : 0 : 1)$.

If we restrict our attention to the case when $X_F(\mathbb{C})$ is a submanifold of $\mathbb{CP}^n$ then we are in a good situation. For instance, it is easy to write down a formula for the Chern classes of $X_F(\mathbb{C})$ in terms of $\deg F$ (cf. [Hirzebruch66, §22]). From this we see that $X_F(\mathbb{C})$ determines $\deg F$ with the sole exception $X_F(\mathbb{C}) \sim S^2$ where $\deg F$ can be 1 or 2.

The real case is trickier. $H_{n-1}(\mathbb{RP}^n, \mathbb{Z}_2) \cong \mathbb{Z}_2$ and the generator is given by the hyperplane class $[H]$. $X(\mathbb{R}) \subset \mathbb{RP}^n$ has a homology class $[X(\mathbb{R})] \in H_{n-1}(\mathbb{RP}^n, \mathbb{Z}_2)$ and $[X(\mathbb{R})] = \deg F \cdot [H]$. Thus the topology determines $\deg F$ mod 2. In fact one can not do better than this. It is not hard to see that if X_F is smooth and F^* is a small homogeneous perturbation of $(x_0^2 + \cdots + x_n^2) \cdot F$ then the pairs $(\mathbb{RP}^n, X_F(\mathbb{R}))$ and $(\mathbb{RP}^n, X_{F^*}(\mathbb{R}))$ are diffeomorphic. This shows that $X(\mathbb{R})$ does not provide an upper bound for the degree.

On the other hand, $X(\mathbb{R})$ does provide a lower bound for $\deg F$. Indeed it is a priori clear that only finitely many topological types can be realized by hypersurfaces of bounded degeree. Thus if $X(\mathbb{R})$ is complicated then $\deg F$ has to be large. [Milnor64] is an explicit result in this direction. I do not even have a conjecture about the precise answer.

Thus we can summarize our results as follows:

Conclusion. *Let $X \subset \mathbb{P}^n$ be a smooth hypersurface. Then*

1. $X(\mathbb{C})$ *determines* $\deg F$.
2. $X(\mathbb{R})$ *determines* $\deg F$ mod 2.
3. *If* $X(\mathbb{R})$ *is complicated then* $\deg F$ *is large.*

The aim of these notes is to collect a series of results and conjectures concerning the recognition problem of real and complex algebraic varieties. As in the hypersurface case, the main idea can be summarized as follows.

Principle 1.4.

1. *$X(\mathbb{C})$ determines the important algebraic invariants of X.*
2. *If $X(\mathbb{R})$ is complicated then X is also complicated.*

§2. Homological Methods

As an intermediate step of our answers, we can study the relationship between $X(\mathbb{C})$ and $X(\mathbb{R})$. The simplest case is when X is the zero set of a real polynomial in one variable. Then $X(\mathbb{C})$ is the set of complex roots and $X(\mathbb{R})$ the set of real roots. We have the following two basic relationships:

1. #(real roots) $\leq$ #(complex roots), and
2. #(real roots) $\equiv$ #(complex roots) mod 2.

It is quite amazing that these elementary assertions can be generalized to arbitrary dimensions.

Theorem 2.1 [Floyd52], [Thom65]. *Let X be a projective variety over $\mathbb{R}$. Then*

$$\sum_i h^i(X(\mathbb{R}), \mathbb{Z}_2) \leq \sum_i h^i(X(\mathbb{C}), \mathbb{Z}_2),$$

where $h^i(X(\mathbb{R}),\mathbb{Z}_2)$ is the dimension of the $\mathbb{Z}_2$-vector space $H^i(X(\mathbb{R}),\mathbb{Z}_2)$.

Theorem 2.2 [Sullivan71]. *Let X be a projective variety over $\mathbb{R}$. Then*

$$\chi(X(\mathbb{R})) \equiv \chi(X(\mathbb{C})) \mod 2,$$

where χ denotes the Euler characteristic. (The choice of the coefficient field does not matter.)

It is slightly disappointing from the algebraic point of view that both of these results are essentially topological. Complex conjugation gives an involution $\tau : X(\mathbb{C}) \to X(\mathbb{C})$ whose fixed point set is precisely $X(\mathbb{R})$. 2.1 and 2.2 hold for the fixed point set of any involution.

Equality frequently holds in 2.1, even in the algebraic case. A generalization is given in [Krasnov83], but the main part is again purely topological. Thus it is possible that the homological aspect of comparing $X(\mathbb{C})$ and $X(\mathbb{R})$ has very little to do with algebraic geometry.

The homological methods give sharp results which are especially useful in dimensions 1 and 2. The reason is that a topological surface is determined by its homology groups. By contrast, the homology groups of a 3-manifold carry very little information. For instance, there are many 3-manifolds M, called *homology spheres* such that

$$H_0(M, \mathbb{Z}) \cong \mathbb{Z}, \quad H_1(M, \mathbb{Z}) \cong H_2(M, \mathbb{Z}) \cong 0, \quad H_3(M, \mathbb{Z}) \cong \mathbb{Z}.$$

For 3-manifolds the crucial information is carried by the fundamental group. In higher dimensions one needs the other homotopy groups as well. One of the main challenges of the theory is to connect the homotopy theoretic properties of $X(\mathbb{C})$ and $X(\mathbb{R})$ with the algebraic nature of X.

§3. The Realization Problem

Over the real numbers, the realization problem has a very nice complete solution. The first results of this type were proved by Seifert. The main contribution to the subject is [Nash52] which was sharpened by [Tognoli73].

Theorem. *For every compact differentiable manifold M there is a real, smooth, projective variety X such that M is diffeomorphic to $X(\mathbb{R})$.*

The case of singular varieties is still not completely solved. For some recent results see [Akbulut-King92].

The realization problem behaves very differently over the complex numbers. There are very few manifolds M which can be written as $X(\mathbb{C})$ for a smooth projective variety X.

First of all, the dimension of M has to be even. The Hodge structure on the cohomology groups of M gives further restrictions. The deepest results in this direction are in [DGMS75].

There has been a lot of recent interest in the fundamental group of complex algebraic varieties. The conclusion is that most finitely presented groups can not be the fundamental group of a smooth projective variety. The simplest such examples are free Abelian groups of odd rank. There is a quite extensive theory of those groups which occur as the fundamental group of a smooth projective or Kähler variety, see for instance [ABCKT96].

[Kapovich-Millson97] found examples of groups which can not be the fundamental group of a smooth quasi-projective variety.

It is quite remarkable that the topological spaces $X(\mathbb{C})$ are special even among compact complex manifolds. As observed by Carlson and Kotschick, the main theorem of [Taubes92] implies that every finitely presented group is the fundamental group of a compact complex 3-manifold.

§4. The Recognition Problem over $\mathbb{C}$

Let X_t be a family of smooth projective varieties depending continuously on a parameter t. Then $X_t(\mathbb{C})$ is a continuously varying family of smooth manifolds, hence locally constant. Thus the topology of $X(\mathbb{C})$ is not able to distinguish the individual varieties X_t from each other. The best we can hope for is that the topology tells us in which family we are.

For a smooth projective curve C the only discrete invariant is the genus $g(C)$. By definition, this is the dimension of the space of holomorphic 1-forms, denoted by $H^0(C, \Omega_C)$. By Serre duality this is dual to the first cohomology group of the structure sheaf $H^1(C, \mathcal{O}_C)$.

The set of complex points $C(\mathbb{C})$ is a compact topological surface. It is also orientable, so it can be obtained from S^2 by attaching handles. The basic invariant is the number of handles.

Theorem 4.1. *Let C be a smooth, projective, algebraic curve. Then*

1. (*Riemann,* 1857) $h^0(C, \Omega_C) =$ *number of handles of* $C(\mathbb{C})$.
2. (*Hurwitz,* 1891) *All curves with the same genus form a connected family.*

Ideally one would like to get similar results in higher dimensions.

Definition 4.2 (Kodaira dimension). Let X be a smooth projective variety of dimension n. In analogy with the curve case it is natural to consider the space of holomorphic n-forms. Unfortunately this is not enough and we have to look at multivalued holomorphic n-forms as well. It is technically easier to work with sections of powers of the line bundle of holomorphic n-forms $H^0(X, (\Omega_X^n)^{\otimes m}) = H^0(X, \mathcal{O}(mK_X))$. The Kodaira dimension, denoted by $\kappa(X)$, essentially measures the growth of these vector spaces.

To be precise, if these groups are always zero then we set $\kappa(X) = -\infty$. Otherwise it turns out that there is a unique integer $0 \leq \kappa(X) \leq \dim X$ and constants $0 < c_1, c_2$ such that

$$c_1 \cdot m^{\kappa(X)} \leq h^0(X, \mathcal{O}(mK_X)) \leq c_2 \cdot m^{\kappa(X)}$$

holds whenever $h^0(X, \mathcal{O}(mK_X)) \neq 0$.

(It is conjectured that there is an integer N and finitely many polynomials $P_i(m)$ such that $h^0(X, \mathcal{O}(mK_X)) = P_i(m)$ if $m \equiv i \bmod N$ and $m \gg 1$, but this is proved only for $\dim X \leq 3$.)

In analogy with 4.1.1 one can ask if the Kodaira dimension of a surface is determined by $X(\mathbb{C})$. It was noticed that the answer is no if we consider $X(\mathbb{C})$ as a topological manifold [Dolgachev66].

As Donaldson theory started to discover the difference between diffeomorphism and homeomorphism in real dimension 4, the hope emerged that this may hold for diffeomorphism. This has been one of the motivating questions of the differential topology of algebraic surfaces. After many contributions, the final step was accomplished by [Pidstrigach95], [Friedman-Qin95]. With the methods of Seiberg-Witten theory, the proof is quite short [Okonek-Teleman95]:

Theorem 4.3. *Let S be a smooth, projective algebraic surface over $\mathbb{C}$. Then $\kappa(S)$ is determined by the differentiable manifold $S(\mathbb{C})$.*

As in 4.1.2 we may also look at the family of all algebraic structures on a given 4-manifold. It is known that they form finitely many connected families. Recent examples of [Manetti98] show that in general there are several families.

Starting with dimension 3, the differentiable stucture of $X(\mathbb{C})$ does not determine the Kodaira dimension of X, as it was first obeserved in [Friedman-Morgan88]. It becomes necessary to find additional topological data. A natural candidate is the symplectic structure of $X(\mathbb{C})$. Hopefully, this provides the right setting in all dimensions.

Definition 4.4 (Symplectic manifolds). A *symplectic* manifold is a pair (M^{2n}, ω) where M is a differentiable manifold of dimension $2n$ and ω is a 2-form $\omega \in \Gamma(M, \wedge^2 T^*)$ which is d-closed and nondegenerate. That is, $d\omega = 0$ and ω^n is nowhere zero.

For a smooth projective variety X the following construction gives a symplectic structure on $X(\mathbb{C})$. On $\mathbb{C}^{n+1}$ consider the *Fubini-Study* 2-form

$$\omega' := \frac{\sqrt{-1}}{2\pi}\left[\frac{\sum dz_i \wedge d\bar{z}_i}{\sum |z_i|^2} - \frac{(\sum \bar{z}_i\, dz_i) \wedge (\sum z_i\, d\bar{z}_i)}{(\sum |z_i|^2)^2}\right].$$

It is closed, nondegenerate on $\mathbb{C}^{n+1} \setminus \{0\}$ and invariant under scalar multiplication. Thus ω' descends to a symplectic 2-form ω on $\mathbb{CP}^n = (\mathbb{C}^{n+1} \setminus \{0\})/\mathbb{C}^*$.

If $X \subset \mathbb{CP}^n$ is any smooth variety, then the restriction $\omega|_X$ makes $X(\mathbb{C})$ into a symplectic manifold.

The resulting symplectic manifold $(X(\mathbb{C}), \omega|_X)$ depends on the embedding $X \hookrightarrow \mathbb{CP}^n$, but the dependence is rather easy to understand:

We say that two symplectic manifolds (M, ω_0) and (M, ω_1) are *symplectic deformation equivalent* if there is a continuous family of symplectic manifolds (M, ω_t) starting with (M, ω_0) and ending with (M, ω_1).

To every smooth projective variety the above construction associates a symplectic manifold $(X(\mathbb{C}), \omega|_X)$ which is unique up to symplectic deformation equivalence.

This allows us to formulate the correct generalization of 4.3.

Conjecture 4.5. *Let X be a smooth, projective variety over $\mathbb{C}$. Then $\kappa(X)$ is determined by the symplectic manifold $(X(\mathbb{C}), \omega|_X)$.*

A few special cases of this conjecture are known. It is conjectured that $\kappa(X) = -\infty$ iff X is uniruled. It was proved in [Kollár98a, 4.2.10] that being uniruled is a property of the symplectic manifold $(X(\mathbb{C}), \omega|_X)$. Mirror symmetry suggests that varieties of Kodaira dimension zero can also be recognized from their symplectic structure.

One can also ask if there are only finitely many connected families of varieties with a given symplectic structure (M, ω). If $b_2(M) = 1$ then this is true even for the differentiable structure by [Kollár98a, 4.2.3] but nothing seems to be known in general.

§5. Rational and Uniruled Varieties

The recognition problem is much less understood for real varieties. As shown by 1.3, we can easily get some mod 2 information about X but it is less clear what else to do. The basic works of [Harnack1876] on curves and [Comessatti14] on surfaces were promising, but it is only recently that some meaningful positive results appeared in higher dimensions.

Very little is known about the connection of $X(\mathbb{R})$ with the Kodaira dimension of X. The special case of varieties with Kodaira dimension $-\infty$ is now becoming clearer, so I mainly concentrate on those.

Partly for historical reasons, I focus on rational and uniruled varieties. Rational varieties are very special but the topology of their real part turns out to be quite interesting. It is conjectured that X is uniruled iff $\kappa(X) = -\infty$, so the study of uniruled varieties fits well within the framework of the recognition probem.

Definition 5.1 (Rational and unirational varieties). In Section 1 we have defined varieties as zero sets of functions. There is another way of associating a geometric object to a function by looking at its graph or its image. This leads to the notions of rational and unirational varieties.

Let $\phi_i(t_0, \dots, t_d)$ be homogeneous polynomials of the same degree. They define a map

$$\Phi : \mathbb{P}^d \dashrightarrow \mathbb{P}^N \quad \text{given as} \quad \mathbf{t} \longmapsto (\phi_0(\mathbf{t}) : \cdots : \phi_N(\mathbf{t})).$$

Over $\mathbb{C}$ the image of such a map is automatically a dense subset of an algebraic variety. A variety which can be written this way is called *unirational*. Note that we do not assume that X has dimension d, but it is not hard to see that once X is unirational we can always choose a parametrization $\Psi : \mathbb{P}^{\dim X} \dashrightarrow X$.

One has to be a little more careful over $\mathbb{R}$. First of all, if we have a real variety X then we are interested only in those parametrizations $\Phi : \mathbb{P}^d \dashrightarrow X$ where the coordinate functions of Φ have real coefficients. Second, the image of $\mathbb{RP}^d$ need not be dense in $X(\mathbb{R})$. For instance the image of

$$(t_0 : t_1) \longmapsto (t_0^4 + t_1^4 : 2t_0^2 t_1^2 : t_0^4 - t_1^4)$$

is only half of the circle $x_1^2 + x_2^2 = x_0^2$.

There are some examples of X where there is no parametrization $\Psi : \mathbb{P}^d \dashrightarrow X$ such that the image of $\mathbb{RP}^d$ is dense in $X(\mathbb{R})$, but for every $x \in X(\mathbb{R})$ there is a parametrization $\Psi_x : \mathbb{P}^d \dashrightarrow X$ such that $\Psi_x(\mathbb{RP}^d)$ contains an open neighborhood of x.

A parametrization of a variety $\Phi : \mathbb{P}^d \dashrightarrow X \subset \mathbb{P}^N$ is especially useful if Φ has an inverse $X \dashrightarrow \mathbb{P}^d$. Over $\mathbb{C}$ this is equivalent to assuming that Φ is injective an a dense open subset of $\mathbb{P}^d$. It is important to note that this fails over $\mathbb{R}$. For instance $\phi : x \mapsto x^3$ gives an injective map $\mathbb{R} \to \mathbb{R}$ but it is a $3:1$ map if viewed as $\phi : \mathbb{C} \to \mathbb{C}$.

The following is an easy example of the recognition problem.

Lemma 5.2. *Let X be a smooth, real, projective variety which is rational. Then $X(\mathbb{R})$ is connected.*

Rationality and unirationality are very useful and strong properties of an algebraic variety but unfortunately they are exceedingly hard to check in practice. A considerable weakeneing of these notions is given next.

Definition 5.3. A variety X of dimension d is called *uniruled* if there is a variety Y of dimension $d-1$ and a map $\Phi : Y \times \mathbb{P}^1 \dashrightarrow X$ which has dense image over $\mathbb{C}$. If in addition Φ has an inverse then we say that X is *ruled.*

As before, one has to be careful with the real versions.

At first sight this notion seems too general. Since Y can be arbitrary, uniruled can be interpreted to mean that X behaves like a rational variety in one direction only. It would be more convincing to have a notion which requires rational like behaviour in every direction. The concept of *rationally connected varieties* was introduced in [KoMiMo92] with exactly this aim in mind.

For our present purposes uniruled is sufficient. The reason is that the current topological methods are not fine enough to detect different type behaviour in different directions. In some sense rational is analogous to positive curvature. At present we have results that distinguish negative curvature from everyting else but we can not handle the mixed curvature case well.

Example 5.4.
1. Nonempty quadrics are rational as shown by the inverse of the stereographic projection from a point of the quadric.

2. Set $S := (x^2 + y^2 + \prod_{i=1}^m (z - a_i) = 0)$ where the a_i are distinct real numbers. Then S is rational over $\mathbb{R}$ iff $m \leq 2$. Indeed, if $m \leq 2$ then this is a quadric so rational. If $m \geq 3$ then $S(\mathbb{R})$ is disconnected, so it can not be rational by 5.2. On the other hand, a surface of the form $x^2 - y^2 + f(z) = 0$ is rational as shown by the substitution

$$(u, v) \longmapsto \left(\frac{f(v) + u^2}{2u}, \frac{f(v) - u^2}{2u}, v \right)$$

So $x^2 + y^2 = f(z)$ is rational over $\mathbb{C}$ but not over $\mathbb{R}$.

3. [Segre51] The cubic surface $z^2 = x^3 + y^3 + c$ is unirational for any c, as shown by the parametrization

$$\begin{aligned} x &= \frac{u^2}{3}, \quad y = \frac{u^6 + 27c - 27v}{9u(6v + u^3)}, \\ z &= \frac{u^6 + 27c - 27v + 54v^2 + 9u^3 v}{9(u^3 + 6v)}. \end{aligned}$$

Uniruled hypersurfaces are easy to characterize.

Theorem 5.5. *Let $X = (f(x_0, \dots, x_n) = 0)$ be a smooth hypersurface of degree d in $\mathbb{P}^n$. Then X is uniruled iff $d \leq n$.*

The question of (uni)rationality of hypersurfaces is more subtle.

Question 5.6. *Let $X = (f(x_0, \dots, x_n) = 0)$ be a smooth hypersurface of degree d in $\mathbb{P}^n$. Is it true that if X is rational then $d \leq 3$?*

I do not know any conceptual reason why the answer should be yes. On the other hand, by now we have many ways of constructing cubic hypersurfaces which are rational [Tregub93], [Hassett99] and there are several unirationality constructions for higher degrees as well. In general X is unirational for $d \leq \Phi(n)$ where Φ is a function that goes to infinity very slowly with n. For instance, a smooth cubic of dimension at least 2 or a smooth quartic of dimension at least 6 is unirational (cf. [Kollár96, V.5.18]).

It is also known that if X is rational and "very general" then $d \leq \frac{2}{3}n + 1$ [Kollár95].

§6. The Nash Conjecture for 3-folds

As we saw in 1.3, in the real version of the recognition problem we can only expect results claiming that if $X(\mathbb{R})$ is complicated then so is X. Among algebraic varieties the rational ones are the simplest. Proving the nonrationality of a variety using only its real points may be the simplest version of the real recognition problem.

A following bold conjecture of Nash asserted that this can not be done:

Conjecture 6.1 [Nash52]. *For every compact differentiable manifold M there is a smooth, real, projective variety X such that X is rational over $\mathbb{R}$ and M is diffeomorphic to $X(\mathbb{R})$.*

As far as I know, this has been the shortest lived conjecture in mathematics since it was disproved 38 years before it was posed. (This seemed not to have been realized for quite some time though.)

Theorem 6.2 [Comessatti14]. *Let S be a smooth, real, projective surface. Assume that S is rational and $S(\mathbb{R})$ is orientable. Then $S(\mathbb{R})$ is either a sphere or a torus.*

The examples $(x^2 + y^2 = z^2 \pm u^2) \subset \mathbb{RP}^3$ show that the sphere and the torus both occur. Also, by blowing up points of $\mathbb{RP}^2$ we see that all nonorientable surfaces do occur. Thus the above theorem gives a necessary and sufficient condition for a topological surface to be representable as the set of real points of a smooth, real, projective surface which is birational to $\mathbb{RP}^2$.

While the negative solution of the surface case suggests that the Nash conjecture may fail in higher dimensions, efforts to use the 2-dimensional case to produce higher dimensional counter examples have failed so far. In fact, most of the higher dimensional results were positive.

The first such result is the solution of the *topological Nash conjecture.* It is generally hoped that a birational map between smooth varieties can be factored as a sequence of blow ups and downs of smooth subvarieties[1]. Blowing up and down makes sense in the topological or differentiable setting. Thus it is reasonable to expect that if X is rational then $X(\mathbb{R})$ can be obtained from $\mathbb{RP}^n$ by blow ups and downs.

[1] This was recently proved by Włodarczyk and by Abramovich, Karu, Matsuki, Włodarczyk.

Theorem 6.3 [Akbulut-King91], [Mikhalkin97]. *Every compact differentiable manifold is obtainable from $\mathbb{RP}^n$ by a sequence of differentiable blow ups and downs.*

One can try to prove the Nash conjecture by trying to make the above sequence of blow ups and downs algebraic. For blow ups one needs to realize certain submanifolds by algebraic suvarieties. This is relatively easy, though not automatic. The algebraic realization of toplogical blow downs is much harder. Assume for instance that $n = 2$ and we want to realize a blow down $\pi : X(\mathbb{R}) \to S$ algebraically. π contracts a simple closed curve $L \subset X(\mathbb{R})$ to a point. In order to make this algebraic, we have to find a smooth rational curve $C \subset X(\mathbb{C})$ of selfintersection -1 such that $C(\mathbb{R})$ is isotopic to L. An algebraic surface frequently contains only finitely many smooth rational curves of selfintersection -1, so it is usually impossible to find such a C. It is easier to find some curve D such that $D(\mathbb{R})$ is isotopic to L. We can then blow up suitable complex conjugate points to achive that D becomes contractible. Contracting D we obtain a surface X' with a very singular point such that $X'(\mathbb{R})$ is a manifold. (Such examples abound in all dimensions, for istance $Y = (x^a + y^b + z^c = t^{2d+1})$ is homeomorphic to $\mathbb{R}^3$ as shown by $\phi(x, y, z) = (x, y, z, \sqrt[2d+1]{x^a + y^b + z^c})$.)

These ideas can be used to solve another weakening of the Nash conjecture:

Theorem 6.4 [Benedetti-Marin92]. *For every compact 3-manifold M there is a singular real algebraic variety X such that X is rational and M is homeomorphic to $X(\mathbb{R})$.*

It turns out that despite these positive partial results, the Nash conjecture fails in dimension 3 as well. Before stating the precise results we need to review the general features of the topology of 3-manifolds.

6.5 (The topology of 3-manifolds).

As a general reference see [Scott83].

Assume for simplicity that we consider only orientable 3-manifolds. By the results of Kneser and Milnor, any such can be written as a connected sum $M_1 \# \cdots \# M_k$ where $\pi_2(M_i) = 0$ and the summands are uniquely determined. Thus one has to concentrate on those 3-manifolds M such that $\pi_2(M) = 0$.

There are 3 known classes of such 3-manifolds.

Seifert fibered: These are 3-manifolds which admit a differentiable map to a surface $M^3 \to F^2$ such that every fiber is a circle. $M^3 \to F^2$ is a fiber bundle outside finitely many points of F and the behaviour of $M^3 \to F^2$ near the exceptional points is fully understood.

Torus bundles: These are 3-manifolds which can be written as a torus bundle over a circle (or are doubly covered by such).

Hyperbolic: These can be written as the quotient of hyperbolic 3-space by a discrete group of motions $\mathbb{H}^3/\Gamma$ for $\Gamma \subset PO(3,1)$. This is the largest class and it is not sufficiently understood.

The geometrization conjecture of Thurston asserts that every 3-manifold M such that $\pi_2(M) = 0$ can be obtained from the above examples. For this to work one has to consider generalizations of these examples to the case of 3-manifolds with boundary and to allow gluing the pieces along boundary components which are tori.

For us the main consequence to keep in mind is that very few 3-manifolds are Seifert fibered.

We are now ready to formulate the 3-dimensional analog of Comessatti's result:

Theorem 6.6 [Kollár98c]. *Let X be a smooth, real, projective 3-fold. Assume that X is uniruled and $X(\mathbb{R})$ is orientable. Then every component of $X(\mathbb{R})$ is among the following*:

1. *Seifert fibered,*
2. *connected sum of several copies of $S^3/\mathbb{Z}_{m_i}$ (called lens spaces),*
3. *torus bundle over S^1 (or doubly covered by a torus bundle),*
4. *finitely many other possible exceptions, or*
5. *obtained from the above by repeatedly taking connected sum with $\mathbb{RP}^3$ and $S^1 \times S^2$.*

I expect the final answer to be even more precise. Unfortunately the current version of the proof falls short of proving these.

Conjecture 6.7. *Notation and assumptions as above.*

1. *Torus bundles over S^1 do not occur (unless they are also Seifert fibered) and the finitely many exceptions in* 6.6.4 *are also not needed.*
2. *All Seifert fibered* 3-*manifolds and all connected sums of lens spaces do occur, at least with X uniruled.*
3. *The list should be much shorter for X rational.*

6.8. I expect that 6.7.2 can be proved using the constructions of [Kollár99b]. The method of [Kollár99c] fails to exclude torus bundles over S^1, but this may not be very hard to achieve eventually.

It is much less clear to me how to deal with the possible finitely many exceptions. There are two related sources of these.

The first part of the proof of 6.6 given in [Kollár99a] shows that $X(\mathbb{R})$ does not change much if we simplify X using the minimal model program (see, for instance, [Kollár-Mori98]). Thus we are reduced to understanding the topology of $X(\mathbb{R})$ where X is in some kind of "standard form". There are 3 types of standard forms and two of these have been dealt with in [Kollár99b], [Kollár99c]. The third class is the so called Fano 3-folds. These are 3-folds such that minus the canonical class is ample. There is a complete list of smooth Fano 3-folds, but unfortunately the reduction method of [Kollár99a] introduces some singularities. There are only finitely many cases by [Kawamata92], but the explicit list is not known. Even if X is some reasonably well known variety, determining $X(\mathbb{R})$ may be quite hard. One of the simplest concrete open problems is the following.

Question. *Let $X \subset \mathbb{P}^4$ be a smooth hypersurface of degree 4. Can $X(\mathbb{R})$ be hyperbolic?*

Standard effective estimates of real and complex algebraic geometry give the following bound for the number of possible cases.

Proposition. *The number of exceptions in 6.6.4 is at most $10^{10^{49}}$.*

Instead of getting bogged down in the minutiae of the precise list, it may be more interesting to consider the following general problem.

Conjecture 6.9. *Let X be a smooth, real, projective, uniruled variety of dimension at least 3. Then none of the connected components of $X(\mathbb{R})$ is hyperbolic. (Even without assuming orientability.)*

A very substantial step towards proving 6.9 is the following:

Theorem 6.10 [Viterbo98]. *Conjecture 6.9 holds if $H_2(X(\mathbb{C}), \mathbb{Z}) \cong \mathbb{Z}$ and X is covered by lines. (A line is a morphism $f : \mathbb{CP}^1 \to X$ such that $f_*[\mathbb{CP}^1]$ generates $H_2(X(\mathbb{C}), \mathbb{Z})$.)*

6.11 (Lagrangian versions). Real algebraic varieties provide the largest known class of Lagrangian and special Lagrangian submanifolds (cf. [McDuff-Salamon95]). It would be quite interesting to know if the above results have their analogs for Lagrangian submanifolds of symplectic varieties. This problem is also related to some of the questions posed by Fukaya during the conference.

§7. The Nonprojective Nash Conjecture

The original conjecture of Nash asked about the existence of a variety X such that $X(\mathbb{R})$ is diffeomorhic to M and X is

1. smooth,
2. projective, and
3. rational.

If we drop the third condition then the asnwer is yes by 3.1. If we drop the smoothness assumption, the answer is again yes by 6.4. Is it possible to drop the projectivity assumption?

Allowing quasi projective varieties instead of projective ones does not help at all. 6.3 suggests to look at compact complex manifolds which can be obtained from $\mathbb{P}^3$ by a sequence of smooth blow ups and downs. The smooth blow up of a projective variety is projective, but it is not clear that the same holds for smooth blow downs, thus we may get something interesting. This class of manifolds was introduced by [Artin68] and [Moishezon67].

Definition–Theorem 7.1. A compact complex manifold Y is called a *Moishezon manifold* or an *Artin algebraic space* if the following equivalent conditions are satisfied:

1. Y is bimeromorphic to a projective variety.
2. Y can be made projective by a sequence of smooth blow ups.

It is not at all clear that there are nonprojective Moishezon manifolds. By a result of Chow and Kodaira, if a smooth compact complex surface is bimeromorphic to a projective variety then it is projective (cf. [BPV84, IV.5]). The first nonalgebraic examples in dimension 3 were found by Hironaka (see [Hartshorne77, App. B.3]).

We of course want to keep the notion of a real structure, thus we look at pairs (Y, τ) where Y is a compact complex manifold and $\tau : Y \to Y$ an antiholomorphic involution. Then $Y(\mathbb{R})$ denotes the fixed point set of τ. Considering such pairs is very natural. The main problem of their theory is that all reasonable names have already been taken. "Real analytic space" is used for something else and "real complex manifold" sounds goofy.

Moishezon manifolds seem quite close to projective varieties. In general, if a property of projective varieties does not obviously involve the existence of an ample line bundle, then it also holds for Moishezon manifolds. It was therefore quite a surpise to me that for the Nash conjecture the nonprojective cases behave very differently.

Theorem 7.2 [Kollár99d]**.** *Let M be a compact, connected 3-manifold. Then there is a sequence of smooth, real blow ups and downs*

$$\mathbb{P}^3 = Y_0 \dashrightarrow Y_1 \dashrightarrow \cdots \dashrightarrow Y_n = X_M$$

such that $X_M(\mathbb{R})$ is diffeomorphic to M.

Acknowledgements. I thank P. Orlik and D. Toledo for many helpful comments.

References

[ABCKT96] J. Amorós, M. Burger, K. Corlette, D. Kotschick and D. Toledo, "Fundamental groups of compact Kähler manifolds", Amer. Math. Soc., 1996.

[Akbulut-King91] S. Akbulut and H. King, Rational structures on 3-manifolds, Pacific J. Math., **150** (1991), 201–204.

[Akbulut-King92] S. Akbulut and H. King, "Topology of Real Algebraic Sets", MSRI Publ. vol. **25**, Springer, 1992.

[Artin68] M. Artin, The implicit function theorem in algebraic geometry, in "Algebraic geometry", Bombay, Oxford Univ. Press, 1968, pp. 13–34.

[BPV84] W. Barth, C. Peters and A. Van de Ven, "Compact Complex Surfaces", Springer, 1984.

[Benedetti-Marin92] R. Benedetti and A. Marin, Déchirures de variétés de dimension trois, Comm. Math. Helv., **67** (1992), 514–545.

[BCR98] J. Bochnak, M. Coste and M-F. Roy, "Real algebraic geometry", Springer, 1998.

[Comessatti14] A. Comessatti, Sulla connessione delle superfizie razionali reali, Annali di Math., **23**(3) (1914), 215–283.

[DGMS75] P. Deligne, Ph. Griffiths, J. Morgan and D. Sullivan, Real homotopy theory of Kähler manifolds, Invent. Math., **29** (1975), 245–274.

[Dolgachev66] I. Dolgachev, On Severi's conjecture concerning simply connected algebraic surfaces, Soviet Math. Dokl., **7** (1966), 1169–1171.

[Floyd52] E. Floyd, On periodic maps and the Euler characteristics of associated spaces, Trans. Amer. Math. Soc., **72** (1952), 138–147.

[Friedman-Morgan88] R. Friedman and J. Morgan, Bull. Amer. Math. Soc., **18** (1988), 1–19.

[Friedman-Qin95] R. Friedman and Z. Qin, On complex surfaces diffeomorphic to rational surfaces, Inv. Math., **120** (1995), 81–117.

[Harnack1876] A. Harnack, Über die Vieltheiligkeit der ebenen algebraischen Kurven, Math. Ann., **10** (1876), 189–198.

[Hartshorne77] R. Hartshorne, "Algebraic Geometry", Springer, 1977.

[Hassett99] B. Hassett, Some rational cubic fourfolds, J. Algebraic Geom., **8** (1999), 103–114.

[Hirzebruch66] F. Hirzebruch, "Topological methods in algebraic geometry", third enlarged ed., Springer, 1966.

[Iskovskikh80] V. A. Iskovskikh, Anticanonical models of three-dimensional algebraic varieties, J. Soviet Math., **13** (1980), 745–814.

[Kapovich-Millson97] M. Kapovich and J. Millson, Artin groups, projective arrangements, and fundamental groups of smooth complex algebraic varieties, C. R. Acad. Sci. Paris Seir. I Math., **325** (1997), 871–876.

[Kawamata92] Y. Kawamata, Boundedness of $\mathbb{Q}$-Fano threefolds, Proc. Int. Conf. Algebra, Contemp. Math., **131** (1992), 439–445.

[Kharlamov76] V. Kharlamov, The topological type of non-singular surfaces in RP^3 of degree four, Funct. Anal. Appl., **10** (1976), 295–305.

[Kollár93] J. Kollár, Effective Base Point Freeness, Math. Ann., **296** (1993), 595–605.

[Kollár95] J. Kollár, Nonrational hypersurfaces, J. Amer. Math. Soc., **8** (1995), 241–249.

[Kollár96] J. Kollár, "Rational Curves on Algebraic Varieties", Springer Verlag, Ergebnisse der Math., vol. **32**, 1996.

[Kollár98a] J. Kollár, Low degree polynomial equations, in "European Congres of Math", Birkhäuser, 1998, pp. 255–288.

[Kollár98b] J. Kollár, Real Algebraic Threefolds I. Terminal Singularities, Collectanea Math. FERRAN SERRANO, 1957-1997, **49** (1998), 335–360.

[Kollár98c] J. Kollár, The Nash conjecture for threefolds, ERA of Amer. Math. Soc., **4** (1998), 63–73.

[Kollár99a] J. Kollár, Real Algebraic Threefolds II. Minimal Model Program, Jour. Amer. Math. Soc., **12** (1999), 33–83.

[Kollár99b] J. Kollár, Real Algebraic Threefolds III. Conic Bundles, to appear.

[Kollár99c] J. Kollár, Real Algebraic Threefolds IV. Del Pezzo Fibrations, in "Complex analysis and algebraic geometry", de Gruyter, 2000, pp. 317–346.

[Kollár99d] J. Kollár, The nonprojective Nash conjecture for threefolds, in preparation.

[KoMiMo92] J. Kollár, Y. Miyaoka and S. Mori, Rationally Connected Varieties, J. Alg. Geom., **1** (1992), 429–448.

[Kollár-Mori98] J. Kollár and S. Mori, "Birational geometry of algebraic varieties", Cambridge Univ. Press, 1998.

[Krasnov83] V. A. Krasnov, Harnack–Thom inequalities for mappings of real algebraic varieties (Russian), Izv. Akad. Nauk SSSR Ser. Mat., **47** (1983), 268–297.

[Manetti98] M. Manetti, On the moduli space of diffeomorphic algebraic surfaces, preprint.

[McDuff-Salamon95] D. McDuff and D. Salamon, "Introduction to symplectic topology", Clarendon, 1995.

[Mikhalkin97] G. Mikhalkin, Blow up equivalence of smooth closed manifolds, Topology, **36** (1997), 287–299.

[Milnor64] J. Milnor, On the Betti numbers of real varieties, Proc. Amer. Math. Soc., **15** (1964), 275–280.

[Moishezon67] B. Moishezon, On n-dimensional compact varieties with n algebraically independent meromorphic functions, Amer. Math. Soc. Transl., **63** (1967), 51–177.

[Nash52] J. Nash, Real algebraic manifolds, Ann. Math., **56** (1952), 405–421.

[Okonek-Teleman95] C. Okonek and A. Teleman, Les invariants de Seiberg-Witten et la conjecture de Van de Ven, C. R. Acad. Sci. Paris, **321** (1995), 457–461.

[Pidstrigach95] V. Pidstrigach, Patching formulas for spin polynomials and a proof of the Van de Ven conjecture, Izvestiya Russ. A.S., **45** (1995), 529–544.

[Scott83] P. Scott, The geometries of 3-manifolds, Bull. London Math. Soc., **15** (1983), 401–487.

[Segre51] B. Segre, The rational solutions of homogeneous cubic equations in four variables, Notae Univ. Rosario, **2** (1951), 1–68.

[Shafarevich72] R. I. Shafarevich, "Basic Algebraic Geometry", (in Russian), Nauka, 1972. English translation: Springer, 1977, 2nd edition, 1994.

[Sullivan71] D. Sullivan, Combinatorial invariants of analytic spaces, Springer Lecture Notes, vol. **192** (1971), 165–168.

[Taubes92] C. Taubes, The existence of anti-self-dual conformal structures, J. Differential Geom., **36** (1992), 163–253.

[Thom65] R. Thom, Sur l'homologie des variétés algébriques réelles, in "Differential and combinatorial topology", Princeton Univ. Press, 1965, pp. 255–265.

[Tognoli73] A. Tognoli, Su una congettura di Nash, Ann. Sci. Norm. Sup. Pisa, **27** (1973), 167–185.

[Tregub93] S. Tregub, Two remarks on four dimensional cubics, Russ. Math. Surv., **48:2** (1993), 206–208.

[Viterbo98] C. Viterbo, Symplectic real algebraic geometry, to appear.

University of Utah
Salt Lake City UT 84112
`kollar@math.utah.edu`

Current address:
Princeton University
Princeton
NJ 08544-1000
U.S.A.
`kollar@math.princeton.edu`

Advanced Studies in Pure Mathematics 31, 2001
Taniguchi Conference on Mathematics Nara '98
pp. 147–165

Approximation of Expectation of Diffusion Process and Mathematical Finance

Shigeo Kusuoka

§1. Introduction

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\{(B^1(t), \ldots, B^d(t)\ ; t \in [0, \infty)\}$ be a d-dimensional Brownian motion. Let $B^0(t) = t$, $t \in [0, \infty)$. Let $V_0, V_1, \ldots, V_d \in C_b^\infty(\mathbf{R}^N; \mathbf{R}^N)$. Here $C_b^\infty(\mathbf{R}^N; \mathbf{R}^n)$ denotes the space of $\mathbf{R}^n$-valued smooth functions defined in $\mathbf{R}^N$ whose devivatives of any order are bounded. We regard elements in $C_b^\infty(\mathbf{R}^N; \mathbf{R}^N)$ as vector fields on $\mathbf{R}^N$.

Now let $X(t, x)$, $t \in [0, \infty)$, $x \in \mathbf{R}^N$, be the solution to the Stratonovich stochastic integral equation

$$X(t, x) = x + \sum_{i=0}^{d} \int_0^t V_i(X(s, x)) \circ dB^i(s). \tag{1}$$

Then there is a unique solution to this equation. Moreover we may assume that with probability one $X(t, x)$ is continuous in t and smooth in x.

In many fields, it is important to compute $E[f(X(T, x))]$ numerically, where f is a function defined in $\mathbf{R}^N$. Let $u(t, x) = E[f(X(t, x))]$, $t > 0$, $x \in \mathbf{R}^N$. Then u satisfies the following PDE:

$$\begin{cases} \dfrac{\partial u}{\partial t}(t, x) = Lu(t, x), \\ u(0, x) = f(x). \end{cases}$$

Here $L = \frac{1}{2} \sum_{i=1}^N V_i^2 + V_0$. So to compute $E[f(X(T, x))]$ is the same to compute the solution $u(T, x)$ to PDE. However, in mathematical finance, if we think of the problem of pricing of Europian options, there are sometimes following difficulties.

(1) L can be degenerate. Moreover, L may not satisfy even the Hörmander condition.

Received April 21, 1999.

(2) f may not be continuously differentiable.

Bally and Talay [1] showed that under the Hörmander condition, Euler-Maruyama approximation gives a good approximation, even if the function f is only bounded measurable. In this paper, we introduce a new method to compute $E[f(X(T,x))]$ numerically. Our method works when the function f is Lipschitz continuous. Our main tools are Malliavin calculus and stochastic Taylor approximation based on Lie algebra. Such stochastic Taylor expansion was initiated by Ben Arous [2], and has been studied by many authors ([3], [8], [10], also see [9]).

§2. Notation and Results

Let $\mathcal{A} = \{\emptyset\} \cup \bigcup_{k=1}^{\infty}\{0,1,\dots,d\}^k$ and for $\alpha \in \mathcal{A}$, let $|\alpha| = 0$ if $\alpha = \emptyset$, let $|\alpha| = k$ if $\alpha = (\alpha^1,\dots,\alpha^k) \in \{0,1,\dots,d\}^k$, and let $\|\alpha\| = |\alpha| + \mathrm{card}\{1 \le i \le |\alpha| \,;\, \alpha^i = 0\}$. For $\alpha, \beta \in \mathcal{A}$, we define $\alpha * \beta \in \mathcal{A}$ by $\alpha * \beta = (\alpha^1,\dots,\alpha^k,\beta^1,\dots,\beta^\ell)$ if $\alpha = (\alpha^1,\dots,\alpha^k) \in \{0,1,\dots,d\}^k$ and $\beta = (\beta^1,\dots,\beta^\ell) \in \{0,1,\dots,d\}^\ell$. Then $\mathcal{A}$ becomes a semigroup with respect the product $*$ with the identity $\emptyset$.

Let $\mathcal{A}_0$ and $\mathcal{A}_1$ denote $\mathcal{A}\backslash\{\emptyset\}$ and $\mathcal{A}\backslash\{\emptyset,(0)\}$, respectively. Also, for each $m \ge 1$, $\mathcal{A}(m)$, let $\mathcal{A}_0(m)$ and $\mathcal{A}_1(m)$ denote $\{\alpha \in \mathcal{A} \,;\, \|\alpha\| \le m\}$, $\{\alpha \in \mathcal{A}_0 \,;\, \|\alpha\| \le m\}$ and $\{\alpha \in \mathcal{A}_1 \,;\, \|\alpha\| \le m\}$ respectively.

Let $B^{\circ\alpha}(t)$, $t \in [0,\infty)$, $\alpha \in \mathcal{A}$, be inductively defined by

$$B^{\circ\emptyset} = 1, \quad B^{\circ(i)} = B^i(t), \quad i = 0,1,\dots,d$$

and

$$B^{\circ\alpha*(i)}(t) = \int_0^t B^{\circ\alpha}(s) \circ dB^i(s), \quad i = 0,1,\dots,d.$$

We define a vector field $V_{[\alpha]}$, $\alpha \in \mathcal{A}$, inductivel by

$$V_{[\emptyset]} = 0, \quad V_{[i]} = V_i, \quad i = 0,1,\dots,d$$
$$V_{[\alpha*(i)]} = [V_\alpha, V_i], \quad i = 0,1,\dots,d.$$

Now we assume the following throughout the paper.

(UFG) There is an integer ℓ such that for any $\alpha \in \mathcal{A}_1$, there are $\varphi_{\alpha,\beta} \in C_b^\infty(\mathbf{R}^N)$, $\alpha \in \mathcal{A}_1$, $\beta \in \mathcal{A}_1(\ell)$, satisfying the following.

$$V_{[\alpha]} = \sum_{\beta \in \mathcal{A}_1(\ell)} \varphi_{\alpha,\beta} V_{[\beta]}.$$

Remark.

(1) Let us think of $C_b^\infty(\mathbf{R}^N)$-module $M = \sum_{\alpha\in\mathcal{A}_0} C_b^\infty(\mathbf{R}^N)V_{[\alpha]}$. Then the assumption (UFG) is equivalent to the assumption that M is finitely generated as a $C_b^\infty(\mathbf{R}^N)$-module.

(2) The following condition (UH) (Uniform Hörmander condition) implies the assumption (UFG).

(UH) There are an integer ℓ and a constant $c > 0$ such that

$$\sum_{\alpha\in\mathcal{A}_1(\ell)} (V_{[\alpha]}, \xi)^2 \geq c|\xi|^2, \quad \text{for all } x, \xi \in \mathbf{R}^N$$

Let V_α, $\alpha \in \mathcal{A}$, be differential operators given by

$$V_\alpha = \text{Identity}, \quad \text{if } \alpha = \emptyset,$$

and

$$V_\alpha = V_{\alpha_1} \cdots V_{\alpha_k}, \quad \text{if } \alpha = (\alpha_1, \ldots, \alpha_k).$$

Let us define a semi-norm $\|\cdot\|_{V,n}$, $n \geq 1$, on $C_0^\infty(\mathbf{R}^N; \mathbf{R})$ by

$$\|f\|_{V,n} = \sum_{k=1}^{n} \sum_{\substack{\alpha_1,\ldots,\alpha_k\in\mathcal{A}_1 \\ \|\alpha_1*\cdots*\alpha_k\|=n}} \|V_{[\alpha_1]}\cdots V_{[\alpha_k]}f\|_\infty.$$

Now let us define a semigroup of linear operators $\{P_t\}_{t\in[0,\infty)}$ by

$$(P_t f)(x) = E[f(X(t,x))], \quad t \in [0,\infty),\ f \in C_b^\infty(\mathbf{R}^N).$$

Then we can prove the following by using a similar argument in Kusuoka-Stroock [7] (also see [5] for the details).

Theorem 1. *For any $n, m \geq 0$ and $\alpha_1, \ldots, \alpha_{n+m} \in \mathcal{A}_1$, there is a constant $C > 0$ such that*

$$\|V_{[\alpha_1]}\cdots V_{[\alpha_n]}P_t V_{[\alpha_{n+1}]}\cdots V_{[\alpha_{n+m}]}f\|_\infty \leq \frac{C}{t^{\|\alpha_1*\cdots*\alpha_{n+m}\|/2}}\|f\|_\infty, \quad f \in C_b^\infty(\mathbf{R}^N).$$

Corollary 2. *For any $n \geq 0$ and $\alpha_1, \ldots, \alpha_n \in \mathcal{A}_1$, there is a constant $C > 0$ such that*

$$\|V_{[\alpha_1]}\cdots V_{[\alpha_n]}P_t f\|_\infty \leq \frac{Ct^{1/2}}{t^{\|\alpha_1*\cdots*\alpha_n\|/2}}\|\nabla f\|_\infty, \quad f \in C_b^\infty(\mathbf{R}^N).$$

Definition 3. We say that a family of random variables $\{Z_\alpha\ ;\ \alpha \in \mathcal{A}_0\}$ is m-moment similar, $m \geq 1$, if $Z_{(0)} = 1$,

$$E[|Z_\alpha|^n] < \infty \quad \text{for any } n \geq 1,\ \alpha \in \mathcal{A}_0,$$

and if

$$E[Z_{\alpha_1} \cdots Z_{\alpha_k}] = E[B^{\circ\alpha_1}(1) \cdots B^{\circ\alpha_k}(1)]$$

for any $k = 1, 2, \dots, m$ and $\alpha_1, \dots, \alpha_k \in \mathcal{A}_0$ with $\|\alpha_1\| + \cdots + \|\alpha_k\| \leq m$.

Let $H : \mathbf{R}^N \to \mathbf{R}^N$ be given by $H(x) = (x_1, x_2, \dots, x_N)$, $x = (x_1, x_2, \dots, x_N) \in \mathbf{R}^N$.

Our main result is the following.

Theorem 4. *Let m be an integer and suppose that a family of random variables $\{Z_\alpha\ ;\ \alpha \in \mathcal{A}_0\}$ is m-moment similar. Let $Q_{(s)}$ be a Markov operator in $C_b(\mathbf{R}^N)$*

$$\begin{aligned} Q_{(s)}f(x) =& E\Big[f\Big(\sum_{k=0}^{m} \frac{1}{k!} \sum_{\substack{\alpha_1, \dots, \alpha_k \in \mathcal{A}_0, \\ \|\alpha_1\| + \cdots + \|\alpha_k\| \leq m}} s^{(\|\alpha_1\| + \cdots + \|\alpha_k\|)/2} \\ & \times (P^0_{\alpha_1} \cdots P^0_{\alpha_k})(V_{[\alpha_1]} \cdots V_{[\alpha_k]} H)(x)\Big)\Big] \end{aligned}$$

for $f \in C_b(\mathbf{R}^N)$ and $x \in \mathbf{R}^N$. Here

$$P^0_\alpha = |\alpha|^{-1} \sum_{k=1}^{|\alpha|} \frac{(-1)^{k-1}}{k} \sum_{\beta_1 * \cdots * \beta_k = \alpha} Z_{\beta_1} \cdots Z_{\beta_k}.$$

Then for any $n \geq 1$ there is a constant $C > 0$ such that

$$\|P_s f - Q_{(s)} f(x)\|_\infty \leq C\Big(\sum_{k=m+1}^{n(m+1)} s^{k/2} \|f\|_{V,k} + s^{(m+1)/2} \|\nabla f\|_\infty\Big),$$
$$s \in (0, 1],\ f \in C_b^\infty(\mathbf{R}^N; \mathbf{R}).$$

Let $T > 0$ and $\gamma > 0$. Let $t_k = t_k^{(n)} = k^\gamma T / n^\gamma$, $n \geq 1$, $k = 0, 1, \dots, n$, and let $s_k = s_k^{(n)} = t_k - t_{k-1}$, $k = 1, \dots, n$. Then we have the following.

Theorem 5. *Let $m \geq 1$ and $Q_{(s)}$, $s > 0$ be as in Theorem 4. Then we have the following.*

For $\gamma \in (0, m-1)$, there is a constant $C > 0$ such that

$$\|P_T f - Q_{(s_n)} Q_{(s_{n-1})} \cdots Q_{(s_1)} f\|_\infty \leq C n^{-\gamma/2} \|\nabla f\|_\infty, \qquad f \in C_b^\infty(\mathbf{R}^N),\ n \geq 1.$$

For $\gamma = m-1$, there is a constant $C > 0$ such that

$$\|P_T f - Q_{(s_n)} Q_{(s_{n-1})} \cdots Q_{(s_1)} f\|_\infty \leq C n^{-(m-1)/2} \log(n+1) \|\nabla f\|_\infty, \qquad f \in C_b^\infty(\mathbf{R}^N),\ n \geq 1.$$

For $\gamma > m-1$, there is a constant $C > 0$ such that

$$\|P_T f - Q_{(s_n)} Q_{(s_{n-1})} \cdots Q_{(s_1)} f\|_\infty \leq C n^{-(m-1)/2} \|\nabla f\|_\infty, \qquad f \in C_b^\infty(\mathbf{R}^N),\ n \geq 1.$$

§3. Example of 5-momemt similar family

Let η_i, $i = 1, \ldots, d$ and η_{ij}, $1 \leq i < j \leq d$, are independent random variables such that

$$P(\eta_i = 0) = \frac{1}{2}, \quad P\left(\eta_i = \pm\sqrt{2 \pm \sqrt{2}}\right) = \frac{1}{8},$$

and

$$P(\eta_{ij} = \pm 1) = \frac{1}{2}.$$

Then we see that

$$E[\eta_i] = E[\eta_i^3] = 0, \quad E[\eta_i^2] = 1, \quad E[\eta_i^4] = 3,$$

and

$$E[\eta_{ij}] = 0, \quad E[\eta_{ij}^2] = 1.$$

Now let us define random variables $\{Z_\alpha\ ;\ \alpha \in \mathcal{A}_0\}$ as follows.

(1) The case where $\|\alpha\| = 1$.

$$Z_i = \eta_i, \quad i = 1, \ldots, d.$$

(2) The case where $\|\alpha\| = 2$.

$$Z_0 = 1,$$
$$Z_{ij} = \begin{cases} \frac{1}{2}(\eta_i\eta_j + \eta_{ij}), & 1 \le i < j \le d, \\ \frac{1}{2}(\eta_i\eta_j - \eta_{ji}), & 1 \le j < i \le d, \\ \frac{1}{2}\eta_i\eta_j, & 1 \le i = j \le d. \end{cases}$$

(3) The case where $\|\alpha\| = 3$.

$$Z_{i0} = Z_{0i} = \frac{1}{2}\eta_i, \quad Z_{iii} = \frac{1}{6}\eta_i^3, \quad 1 \le i \le d,$$
$$Z_{iij} = Z_{jii} = \frac{1}{4}\eta_i, \quad Z_{iji} = 0, \quad 1 \le i \ne j \le d,$$

and $Z_\alpha = 0$ in other cases.

(4) The case where $\|\alpha\| = 4$.

$$Z_\alpha = E[B^{\circ\alpha}],$$

that is

$$Z_{iijj} = \frac{1}{8}, \quad 1 \le i, j \le d,$$
$$Z_{0ii} = Z_{ii0} = \frac{1}{4}, \quad 1 \le i \le d,$$
$$Z_{00} = \frac{1}{2},$$

and $Z_\alpha = 0$ in the other case.

(5) The case where $\|\alpha\| \ge 5$.

$$Z_\alpha = 0.$$

Then the family of random variables $\{Z_\alpha \; ; \; \alpha \in \mathcal{A}_0\}$ is 5-moment similar.

§4. Preparation from Algebra

We say that a polynomial p of $x_\alpha, \alpha \in \mathcal{A}_0$, is m-homogeneous, $m \ge 0$, if

$$p(\varepsilon^{\|\alpha\|}x_\alpha, \alpha \in \mathcal{A}_0) = \varepsilon^m p(x_\alpha, \alpha \in \mathcal{A}_0), \quad \varepsilon > 0.$$

Let $\mathcal{U}$ be the free algebra generated by $\{v_0, v_1, \dots, v_d\}$ over $\mathbf{R}$. Then the algebra $\mathcal{U}$ can be extended to the algebra $\overline{\mathcal{U}}$ of formal power series in $\{v_0, v_1, \dots, v_d\}$. We define $v^\alpha \in \mathcal{U}$, $\alpha \in \mathbf{A}$, by $v^\emptyset = 1$, and by $v^\alpha = v^{\alpha_1} \cdots v^{\alpha_k}$, if $\alpha = (\alpha^1, \dots, \alpha^k)$. Then $\overline{\mathcal{U}}$ is the complete direct sum of the space $\mathbf{R}v^\alpha$, $\alpha \in \mathcal{A}$. We define convergence in $\overline{\mathcal{U}}$ by $\sum_{\alpha\in\mathcal{A}} a_{\alpha,n} v^\alpha \to \sum_{\alpha\in\mathcal{A}} a_\alpha v^\alpha$, $n \to \infty$, if $a_{\alpha,n} \to a_\alpha$ for any $\alpha \in \mathcal{A}$.

For $x, y \in \overline{\mathcal{U}}$, let $[xy] = xy - yx$. For $\alpha \in \mathbf{A}$, let $v^{[\alpha]} \in \mathcal{U}$ denote 0, if $\alpha = \emptyset$, v_i, if $\alpha = i \in \{0, 1, \dots, d\}$, and $[\cdots[[v_{\alpha^1} v_{\alpha^2}] v_{\alpha^3}] \cdots, v_{\alpha^k}]$, if $\alpha = (\alpha^1, \dots, \alpha^k)$ and $k \geq 2$. Let $\overline{\mathcal{U}}^{\mathcal{L}}$ be the closure of $\sum_{\alpha\in\mathcal{A}} \mathbf{R}v^{[\alpha]}$ in $\overline{\mathcal{U}}$. Then $\overline{\mathcal{U}}^{\mathcal{L}}$ is closed under Lie product $[\quad]$ (see Jacobson [4, p.168]).

We use the following two theorems (see Jacobson [J, pp.167–174]).

Theorem 6 (Friedrichs). *Let δ be a continuous homomorphism from $\overline{\mathcal{U}}$ into $\overline{\mathcal{U}} \otimes \overline{\mathcal{U}}$ determined by $\delta(1) = 1 \otimes 1$ and $\delta(v_i) = v_i \otimes 1 + 1 \otimes v_i$, $i = 0, 1, \dots, d$. Then for $x \in \overline{\mathcal{U}}$, $x \in \overline{\mathcal{U}}^{\mathcal{L}}$ if and only if $\delta(x) = x \otimes 1 + 1 \otimes x$.*

Theorem 7. *Let σ be a linear continuous operator from $\overline{\mathcal{U}}$ into $\overline{\mathcal{U}}^{\mathcal{L}}$ given by $\sigma(v^\alpha) = |\alpha|^{-1} v^{[\alpha]}$, $\alpha \in \mathcal{A}$. Then the restriction of σ on $\overline{\mathcal{U}}^{\mathcal{L}}$ is identity.*

Let $\mathcal{B}_{\overline{\mathcal{U}}}$ be a Borel algebra over $\overline{\mathcal{U}}$. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. One can define $\overline{\mathcal{U}}$-valued random variables and their expectaions etc. naturally. Let $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ be a filtration satisfying a usual hypothesis, $(B^1(t), \dots, B^d(t))$, $t \in [0, \infty)$, be a d-dimensional $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-Brownian motion, and $B^0(t) = t$, $t \in [0, \infty)$. We say that $X(t)$ is a $\overline{\mathcal{U}}$-valued continuous semimartingale, if there are continuous semimartingales X_α, $\alpha \in \mathcal{A}$, such that $X(t) = \sum_{\alpha\in\mathbf{A}} X_\alpha(t) v^\alpha$. For $\overline{\mathcal{U}}$-valued continuous semimartingale $X(t)$, $Y(t)$, we can define $\overline{\mathcal{U}}$-valued continuous semimartingales $\int_0^t X(s) \circ dY(s)$ and $\int_0^t \circ dX(s) Y(s)$ by

$$\int_0^t X(s) \circ dY(s) = \sum_{\alpha,\beta\in\mathcal{A}} \Big(\int_0^t X_\alpha(s) \circ dY_\beta(s) \Big) v^\alpha v^\beta,$$

$$\int_0^t \circ dX(s) Y(s) = \sum_{\alpha,\beta\in\mathcal{A}} \Big(\int_0^t Y_\beta(s) \circ dX_\alpha(s) \Big) v^\alpha v^\beta,$$

where

$$X(t) = \sum_{\alpha\in\mathcal{A}} X_\alpha(t) v^\alpha, \quad Y(t) = \sum_{\beta\in\mathcal{A}} Y_\beta(t) v^\beta.$$

Then we have

$$X(t)Y(t) = X(0)Y(0) + \int_0^t X(s) \circ dY(s) + \int_0^t \circ dX(s) Y(s).$$

Since $\mathbf{R}$ is regarded a vector subspace in $\overline{\mathcal{U}}$, we can define $\int_0^t X(s) \circ dB^i(s)$, $i = 0, 1, \ldots, d$, naturally. We can similarly think of $\overline{\mathcal{U}} \otimes \overline{\mathcal{U}}$-valued semimartingales and stochastic calculus for them.

Now let us consider SDE on $\overline{\mathcal{U}}$

$$X(t) = 1 + \sum_{i=0}^{d} \int_0^t X(s) v_i \circ dB^i(s), \quad t \geq 0.$$

One can easily solve this SDE and obtain

$$X(t) = 1 + \sum_{\alpha \in \mathcal{A}_0} B^{\circ\alpha}(t) v^{\alpha}.$$

We also have the following.

Proposition 8. *Let p^0_α, $\alpha \in \mathcal{A}_0$, be $\|\alpha\|$-homogeneous polynomials in x_β, $\beta \in \mathcal{A}_0$, given by*

$$p^0_\alpha(x_\beta, \beta \in \mathcal{A}_0) = |\alpha|^{-1} \sum_{k=1}^{|\alpha|} \frac{(-1)^{k-1}}{k} \sum_{\substack{\beta_1, \ldots, \beta_k \in \mathcal{A}_0 \\ \beta_1 * \cdots * \beta_k = \alpha}} x_{\beta_1} \cdots x_{\beta_k}.$$

Then

$$\log X(t) = \sum_{\alpha \in \mathcal{A}_0} p^0_\alpha(B^{\circ\beta}(t), \beta \in \mathcal{A}_0) v^{[\alpha]}.$$

In other words,

$$X(t) = 1 + \sum_{\alpha \in \mathcal{A}_0} B^{\circ\alpha}(t) v^{\alpha} = \exp\left(\sum_{\alpha \in \mathcal{A}_0} p^0_\alpha(B^{\circ\beta}(t), \beta \in \mathcal{A}_0) v^{[\alpha]} \right).$$

Proof. Note that

$$\delta(X(t)) = 1 \otimes 1 + \sum_{i=0}^{d} \int_0^t \delta(X(s))(v_i \otimes 1 + 1 \otimes v_i) \circ dB^i(s),$$

and

$$\begin{aligned} X(t) \otimes X(t) =& 1 \otimes 1 + \int_0^t \circ d(X(s) \otimes 1)(1 \otimes X(s)) \\ &+ \int_0^t (X(s) \otimes 1) \circ d(1 \otimes X(s)) \\ =& 1 \otimes 1 + \sum_{i=0}^{d} \int_0^t X(s) \otimes X(s)(v_i \otimes 1 + 1 \otimes v_i) \circ dB^i(s). \end{aligned}$$

Since one can easily see the uniqueness of such SDE on $\overline{\mathcal{U}} \otimes \overline{\mathcal{U}}$, we have

$$\delta(X(t)) = X(t) \otimes X(t).$$

For any $u \in \overline{\mathcal{U}}$ with of the form $u = \sum_{\alpha \in \mathcal{A}_0} a_\alpha v^\alpha$, we have

$$\exp(u) \otimes \exp(u) = \exp(u \otimes 1 + 1 \otimes u),$$

which implies

$$\log((1+u) \otimes (1+u)) = \log(1+u) \otimes 1 + 1 \otimes \log(1+u).$$

So we have

$$\delta(\log X(t)) = \log(\delta X(t)) = \log X(t) \otimes 1 + 1 \otimes \log X(t).$$

So by Theorem 6 we see that $\log X(t) \in \overline{\mathcal{U}}^{\mathcal{L}}$ *P-a.s.* On the other hand,

$$\log X(t) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \Bigg(\sum_{\alpha_1, \dots, \alpha_k \in \mathcal{A}_0} B^{\circ \alpha_1}(t) \cdots B^{\circ \alpha_k}(t) v^{\alpha_1 * \cdots * \alpha_k} \Bigg).$$

So acting the linear operator σ in Theorem 7, we have our assertion.
Q.E.D.

Proposition 9. *There are polynomials q_α^0, $\alpha \in \mathcal{A}_0$, in x_β, $\beta \in \mathcal{A}_0$, such that*

$$\log\Bigg(\exp(-x_0 v_0) \exp\Big(\sum_{\alpha \in \mathcal{A}_0} x_\alpha v^{[\alpha]} \Big) \Bigg) = \sum_{\alpha \in \mathcal{A}_1} q_\alpha^0(x_\beta,\, \beta \in \mathcal{A}_0) v^{[\alpha]}$$

for any $x_\beta \in \mathbf{R}$, $\beta \in \mathcal{A}_0$. Moreover, $q_0^0 = 0$ and q_α^0 is $\|\alpha\|$-homogeneous for each $\alpha \in \mathcal{A}_1$.

Proof. Similarly to the proof of Proposition 8, we see that $\log\big(\exp(-x_0 v_0) \exp(\sum_{\alpha \in \mathcal{A}_0} x_\alpha v^{[\alpha]})\big) \in \overline{\mathcal{U}}^{\mathcal{L}}$. Since we have

$$\exp(-x_0 v_0) \exp\Bigg(\sum_{\alpha \in \mathcal{A}_0} x_\alpha v^{[\alpha]} \Bigg) = 1 + \sum_{\alpha \in \mathcal{A}_1} x_\alpha v^{[\alpha]}$$
$$+ \sum_{\ell + k \geq 2} \sum_{\alpha_1, \dots, \alpha_k \in \mathcal{A}_0} \frac{1}{\ell! k!} (-x_0)^\ell x_{\alpha_1} \cdots x_{\alpha_k} v_0^\ell v^{[\alpha_1]} \cdots v^{[\alpha_k]}.$$

Note that $v_0^\ell v^{[\alpha_1]} \cdots v^{[\alpha_k]} \in \mathcal{U}'_{2\ell + \|\alpha_1\| \cdots \|\alpha_k\|}$. So acting the linear operator σ in Theorem 7 again, we have our assertion. Q.E.D.

§5. Basic Estimates

For $n \geq 0$ let φ_n denote a map from $\overline{\mathcal{U}}$ into the space of differential operators in $\mathbf{R}^N$ of order n given by

$$\varphi_n\left(\sum_{\alpha\in\mathcal{A}} a_\alpha v^\alpha\right) = \sum_{\alpha\in\mathcal{A}(n)} a_\alpha V_\alpha, \quad a_\alpha \in \mathbf{R},\, \alpha \in \mathcal{A}.$$

Note that if $u \in \overline{\mathcal{U}}^{\mathcal{L}}$, then $\varphi_n(u)$ is a vector field.

First we observe the following.

Proposition 10. *For any $U \in C_b^\infty(\mathbf{R}^N;\mathbf{R}^N)$,*

$$\left\| f(\exp(U)(\,\cdot\,)) - \sum_{k=0}^{n} \frac{1}{k!} U^k f \right\|_\infty \leq \frac{1}{(n+1)!} \|U^{n+1} f\|_\infty$$

for any $f \in C_b^\infty(\mathbf{R}^N)$ and $n \geq 1$.

Proof. One can prove the following inductively.

$$f(\exp(tU)(x)) = \sum_{k=0}^{n} \frac{t^k}{k!}(U^k f)(x) + \int_0^t \frac{(t-s)^n}{n!}(U^{n+1}f)(\exp(sU)(x))\, ds.$$

Then we have our assertion. Q.E.D.

As corllaries of the above Proposition, we have the following.

Proposition 11. *For any $u = \sum_{\alpha\in\mathcal{A}_1} a_\alpha v^{[\alpha]} \in \overline{\mathcal{U}}^{\mathcal{L}}$, and $n \geq 1$ we have*

$$\begin{aligned}&\|f(\exp(\varphi_n(u))(\,\cdot\,)) - (\varphi_n(\exp(u))f)(\,\cdot\,)\|_\infty \\ &\quad\leq \sum_{k=n+1}^{n(n+1)} \max\{|a_\alpha|^{1/\|\alpha\|}\,;\, \alpha \in \mathcal{A}_1(n)\}^k \|f\|_{V,k}\end{aligned}$$

for any $f \in C_b^\infty(\mathbf{R}^N)$.

Proposition 12. *For any $u = \sum_{\alpha\in\mathcal{A}_0} a_\alpha v^{[\alpha]} \in \overline{\mathcal{U}}^{\mathcal{L}}$, and $n \geq 1$ we have a constant C depending only on d and n such that*

$$\begin{aligned}&\|f(\exp(\varphi_n(u))(\,\cdot\,)) - (\varphi_n(\exp(u))f)(\,\cdot\,)\|_\infty \\ &\quad\leq C \sum_{k=n+1}^{n(n+1)} \max\{|a_\alpha|^{1/\|\alpha\|}\,;\, \alpha \in \mathcal{A}_0(n)\}^k \sum_{\alpha\in\mathcal{A},\, \|\alpha\|=k} \|V_\alpha f\|_\infty\end{aligned}$$

for any $f \in C_b^\infty(\mathbf{R}^N)$.

Also, we have the following.

Proposition 13. *For any* $u^{(i)} = \sum_{\alpha \in \mathcal{A}_0} a_\alpha^{(i)} v^{[\alpha]} \in \overline{\mathcal{U}}^{\mathcal{L}}$, $i = 1, 2$, *and* $n \geq 1$, *we have a constant* C *depending only on* d *and* n *such that*

$$\|f(\exp(\varphi_n(u^{(1)}))(\exp(\varphi_n(u^{(2)}))(\,\cdot\,))) - (\varphi_n(\exp(u^{(2)})\exp(u^{(1)}))f)(\,\cdot\,)\|_\infty$$
$$\leq C \sum_{k=n+1}^{2n(n+1)} \max\{|a_\alpha^{(i)}|^{1/\|\alpha\|}\,;\,\alpha \in \mathcal{A}_0(2n),\, i = 1, 2\}^k \sum_{\alpha \in \mathcal{A},\, \|\alpha\| = k} \|V_\alpha f\|_\infty$$

for any $f \in C_b^\infty(\mathbf{R}^N)$.

Proof. Note that

$$\begin{aligned}
&f(\exp(\varphi_n(u^{(1)}))(\exp(\varphi_n(u^{(2)}))(x))) - (\varphi_n(\exp(u^{(2)})\exp(u^{(1)}))f)(x) \\
&\quad = f(\exp(\varphi_n(u^{(1)}))(\exp(\varphi_n(u^{(2)}))(x))) \\
&\qquad - (\varphi_n(\exp(u^{(1)}))f)(\exp(\varphi_n(u^{(2)}))(x))) \\
&\qquad + (\varphi_n(\exp(u^{(1)}))f)(\exp(\varphi_n(u^{(2)}))(x))) \\
&\qquad - \varphi_n(\exp(u^{(2)}))(\varphi_n(\exp(u^{(1)}))f)(x)) \\
&\qquad + \varphi_n(\exp(u^{(2)}))(\varphi_n(\exp(u^{(1)}))f)(x)) \\
&\qquad - (\varphi_n(\exp(u^{(2)})\exp(u^{(1)}))f)(x).
\end{aligned}$$

Then we have our assertion from previous two propositions. Q.E.D.

§6. Moment Equivalent Families

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

Definition 14. We say that families of random variables $\{Z_\alpha\,;\,\alpha \in \mathcal{A}_0\}$ and $\{Z'_\alpha\,;\,\alpha \in \mathcal{A}_0\}$ are m-moment equivalent, $m \geq 1$, if

$$E[|Z_\alpha|^n] < \infty, \quad E[|Z'_\alpha|^n] < \infty, \quad \text{for any } n \geq 1 \text{ and } \alpha \in \mathcal{A}_0,$$

and

$$E[Z_{\alpha_1} \cdots Z_{\alpha_k}] = E[Z'_{\alpha_1} \cdots Z'_{\alpha_k}]$$

for any $k = 1, 2, \ldots, m$ and $\alpha_1, \ldots, \alpha_k \in \mathcal{A}_0$ with $\|\alpha_1\| + \cdots + \|\alpha_k\| \leq m$.

The main result in this section is the following.

Theorem 15. *Let $m \geq 1$. Let $\{Z_\alpha^{(1)} \,;\, \alpha \in \mathcal{A}_0\}$ and $\{Z_\alpha^{(2)} \,;\, \alpha \in \mathcal{A}_0\}$ are m-moment equivalent families of random variables such that $Z_{(0)}^{(1)} = Z_{(0)}^{(2)} = 1$. Let $Z^{(i)}(\varepsilon)$, $\varepsilon > 0$, be a $\overline{\mathcal{U}}^{\mathcal{L}}$-valued random variable given by $Z^{(i)}(\varepsilon) = \sum_{\alpha \in \mathcal{A}_0} \varepsilon^{\|\alpha\|} Z_\alpha^{(i)} v^{[\alpha]}$.*

Then for any $n \geq 1$, there is a constant $C > 0$ depending only on n and moments of $Z_\alpha^{(i)}$, $i = 1, 2$, $\alpha \in \mathcal{A}_0(n)$, such that

$$\sup_{x \in \mathbf{R}^N} \left| E[f(\exp(\varphi_n(Z^{(1)}(\varepsilon)))(x))] - E[f(\exp(\varphi_n(Z^{(2)}(\varepsilon)))(x))] \right|$$
$$\leq C \Bigl(\sum_{k=m+1}^{n(m+1)} \varepsilon^k \|f\|_{V,k} + \varepsilon^{n+1} \|\nabla f\|_\infty \Bigr), \quad \varepsilon \in (0,1],\ f \in C_b^\infty(\mathbf{R}^N; \mathbf{R}).$$

To prove this theorem we need some preparations.

First we have the following combining Propositions 12 and 13.

Proposition 16. *Let $\{Z_\alpha \,;\, \alpha \in \mathcal{A}_0\}$ is a family of random variables such that $Z_0 = 1$. Let $Z(\varepsilon) = \sum_{\alpha \in \mathcal{A}_0} \varepsilon^{\|\alpha\|} Z_\alpha^{(i)} v^{[\alpha]}$. Then for any $n \geq 1$ and $p \in [1, \infty)$, there is a constant $C > 0$ depending only on n, p, and moments of Z_α, $\alpha \in \mathcal{A}_0(n)$, such that*

$$\sup_{x \in \mathbf{R}^N} E\Biggl[\biggl| f(\exp(\varphi_n(Z(\varepsilon)))(\exp(-\varepsilon^2 V_0)(x)))$$
$$- f\Bigl(\exp\Bigl(\varphi_n\Bigl(\sum_{\alpha \in \mathcal{A}_0} \varepsilon^{\|\alpha\|} q_\alpha^0(Z_\beta,\, \beta \in \mathcal{A}_0) v^{[\alpha]} \Bigr) \Bigr)(x) \Bigr) \biggr|^p \Biggr]^{1/p}$$
$$\leq C \sum_{\substack{\alpha \in \mathcal{A}_0 \\ n+1 \leq \|\alpha\| \leq 2n(n+1)}} \varepsilon^{\|\alpha\|} \|V_\alpha f\|_\infty, \quad \varepsilon \in (0,1],\ f \in C_b^\infty(\mathbf{R}^N; \mathbf{R}).$$

Here polynomials q_α^0, $\alpha \in \mathcal{A}_1$, are as in Proposition 9.

As a corollary we have the following.

Corollary 17. *Let us assume the same as the previous proposition. Then for any $n \geq 1$ and $p \in [1, \infty)$, there is a constant $C > 0$*

depending only on n, p and moments of Z_α, $\alpha \in \mathcal{A}_0(n)$, such that

$$\sup_{x\in\mathbf{R}^N} E\Bigg[\Bigg|\exp(\varphi_n(Z(\varepsilon)))(\exp(-\varepsilon^2 V_0)(x)) - \exp\Bigg(\varphi_n\Big(\sum_{\alpha\in\mathcal{A}_0}\varepsilon^{\|\alpha\|}q^0_\alpha(Z_\beta,\,\beta\in\mathcal{A}_0)v^{[\alpha]}\Big)(x)\Bigg)\Bigg|^p\Bigg]^{1/p}$$
$$\le C\varepsilon^{n+1}\sum_{\substack{\alpha\in\mathcal{A}_0\\ n+1\le\|\alpha\|\le 2n(n+1)}}\|V_\alpha H\|_\infty,\quad \varepsilon\in(0,1].$$

Proof. Let $\psi \in C_b^\infty(\mathbf{R};\mathbf{R})$ such that $\psi(t) = t$, $|t| < 1$, and $0 \le \psi'(t) \le 1$, $t \in \mathbf{R}$. Let $f_{\ell,j} \in C_b^\infty(\mathbf{R}^N;\mathbf{R})$, $\ell \ge 1$, $j = 1,\dots,N$, be given by $f_{\ell,j}(x) = \ell\psi(\ell^{-1}x_j)$. Then we see that

$$\sup_{\ell\ge1,\,j=1,\dots,N}\|\nabla f_{\ell,j}\|_\infty < \infty,$$

and

$$\max_{j=1,\dots,N}\|\nabla^k f_{\ell,j}\|_\infty \to 0,\ \ \ell\to\infty,\quad k\ge 2.$$

So we see that

$$\sup_{\ell\ge1,\,j=1,\dots,N}\|V_\alpha f_{\ell,j}\|_\infty < \infty,\quad \alpha\in\mathcal{A}_0.$$

Therefore applying the previous proposition for $f_{\ell,j}$ and letting $\ell \uparrow \infty$, we have our assetion. Q.E.D.

Similarly by using Proposition 12, we have the following.

Proposition 18. *Let us assume the same as the previous proposition. Then for any $n \ge 1$ and $p \in [1,\infty)$, there is a constant $C > 0$ depending only on n, p and moments of Z_α, $\alpha \in \mathcal{A}_0(n)$, such that*

$$\sup_{x\in\mathbf{R}^N} E[|\exp(\varphi_n(Z(\varepsilon))(x)) - (\varphi_n(\exp(Z(\varepsilon)))H)(x)|^p]^{1/p}$$
$$\le C\varepsilon^{n+1}\sum_{\substack{\alpha\in\mathcal{A}_0\\ n+1\le\|\alpha\|\le 2n(n+1)}}\|V_\alpha H\|_\infty,\quad \varepsilon\in(0,1].$$

Now let us prove Theorem 15. Note that

$$\begin{aligned}
&\sup_{x\in\mathbf{R}^N} \left|E[f(\exp(\varphi_n(Z^{(1)}(\varepsilon)))(x))] - E[f(\exp(\varphi_n(Z^{(2)}(\varepsilon)))(x))]\right| \\
&\quad = \sup_{x\in\mathbf{R}^N} \Big|E[f(\exp(\varphi_n(Z^{(1)}(\varepsilon)))(\exp(-\varepsilon^2 V_0)(x)))] \\
&\qquad\qquad - E[f(\exp(\varphi_n(Z^{(2)}(\varepsilon)))(\exp(-\varepsilon^2 V_0)(x)))]\Big|
\end{aligned}$$

On the other hand, by Corollary 17, we have

$$\begin{aligned}
&\sup_{x\in\mathbf{R}^N} E\Bigg[\Bigg|f(\exp(\varphi_n(Z^{(i)}(\varepsilon))(\exp(-\varepsilon^2 V_0)(x))) \\
&\qquad - f\Bigg(\exp\Bigg(\varphi_n\Bigg(\sum_{\alpha\in\mathcal{A}_0} \varepsilon^{\|\alpha\|} q^0_\alpha(Z^{(i)}_\beta,\ \beta\in\mathcal{A}_0)v^{[\alpha]}\Bigg)(x)\Bigg)\Bigg)\Bigg|\Bigg] \\
&\quad \le C\varepsilon^{n+1}\Bigg(\sum_{\substack{\alpha\in\mathcal{A}_0 \\ n+1\le\|\alpha\|\le 2n(n+1)}} \|V_\alpha H\|_\infty\Bigg)\|\nabla f\|_\infty, \\
&\qquad\qquad\qquad\qquad \varepsilon\in(0,1],\ f\in C_b^\infty(\mathbf{R}^N;\mathbf{R}).
\end{aligned}$$

Also, since $q^0_{(0)} = 0$, by Proposition 11 we have

$$\begin{aligned}
&\sup_{x\in\mathbf{R}^N} E\Bigg[\Bigg|f\Bigg(\exp\Bigg(\varphi_n\Bigg(\sum_{\alpha\in\mathcal{A}_0} \varepsilon^{\|\alpha\|} q^0_\alpha(Z^{(i)}_\beta,\ \beta\in\mathcal{A}_0)v^{[\alpha]}\Bigg)(x)\Bigg)\Bigg) \\
&\qquad - E\Bigg[\Bigg(\varphi_n\Bigg(\exp\Bigg(\sum_{\alpha\in\mathcal{A}_0} \varepsilon^{\|\alpha\|} q^0_\alpha(Z^{(i)}_\beta,\ \beta\in\mathcal{A}_0)v^{[\alpha]}\Bigg)\Bigg)f\Bigg)(x)\Bigg|\Bigg] \\
&\quad \le C\varepsilon^{n+1}\Bigg(\sum_{k=n+1}^{n(n+1)} \|f\|_{V,k}\Bigg), \quad \varepsilon\in(0,1],\ f\in C_b^\infty(\mathbf{R}^N;\mathbf{R}).
\end{aligned}$$

Note that $\{q^0_\alpha(Z^{(i)}_\beta,\ \beta\in\mathcal{A}_0)\ ;\ \alpha\in\mathcal{A}_0\}$, $i=1,2$, are m-moment equivalent to each other, we see that

$$\begin{aligned}
&E\Bigg[\Bigg(\varphi_m\Bigg(\exp\Bigg(\sum_{\alpha\in\mathcal{A}_0} \varepsilon^{\|\alpha\|} q^0_\alpha(Z^{(1)}_\beta,\ \beta\in\mathcal{A}_0)v^{[\alpha]}\Bigg)\Bigg)f\Bigg)(x)\Bigg] \\
&\quad = E\Bigg[\Bigg(\varphi_m\Bigg(\exp\Bigg(\sum_{\alpha\in\mathcal{A}_0} \varepsilon^{\|\alpha\|} q^0_\alpha(Z^{(2)}_\beta,\ \beta\in\mathcal{A}_0)v^{[\alpha]}\Bigg)\Bigg)f\Bigg)(x)\Bigg].
\end{aligned}$$

Therefore we have our theorem.

§7. SDE

Let $X(t,x)$ be the solution of SDE (1). Also, let $\tilde{X}(t)$ be the solution to SDE (2) in $\overline{\mathcal{U}}$. Then we have the following.

Proposition 19. *For any $n \geq 1$, there is a constant C depending only on d and n such that*

$$\sup_{x\in\mathbf{R}^N} E\big[\,\big|f(X(t,x)) - (\varphi_n(\tilde{X}(t))f)(x)\big|^2\,\big]^{1/2}$$
$$\leq Ct^{(n+1)/2} \sum_{\substack{\alpha\in\mathcal{A} \\ \|\alpha\|=n+1,n+2}} \|V_\alpha f\|_\infty, \quad t\in(0,1],\ f\in C_b^\infty(\mathbf{R}^N;\mathbf{R}).$$

Proof. Note that

$$f(X(t,x)) = f(x) + \sum_{i=0}^{d}\int_0^t (V_i f)(X(t,x))\circ dB^i(t).$$

So we have

$$f(X(t,x)) = \sum_{\alpha\in\mathcal{A}(n)} (V_\alpha f)(x)B^{\circ\alpha}(t) + R(t,x).$$

Here

$$R(t,x) = {\sum}' \int_0^t \circ dB^{\alpha^k}(s_k)\int_0^{s_k}\circ dB^{\alpha^{k-1}}\cdots$$
$$\cdots\int_0^{s_1}\circ dB^i(s_0)(V_iV_\alpha f)(X(s_0,x))$$

and $\sum'$ is the summation with respect to $\alpha = (\alpha^1,\ldots,\alpha^k)\in\mathcal{A}(n)$ and $i=0,1,\ldots,d$, with $\|(i)*\alpha\|\geq n+1$. Since

$$\int_0^t (V_iV_\alpha f)(X(s,x))\circ dB^i(s)$$
$$= \int_0^t (V_iV_\alpha f)(X(s,x))\,dB^i(s) + (1-\delta_{0,i})\frac{1}{2}\int_0^t (V_i^2V_\alpha f)(X(s,x))\,ds.$$

we see that there is a constant $C(d,n)$ depending only on d and n such that

$$\sup_{x\in\mathbf{R}^N} E[|R(t,x)|^2]^{1/2}$$
$$\leq C(d,n)t^{(n+1)/2}\max\{\|V_\alpha f\|_\infty\ ;\ \alpha\in\mathcal{A},\ \|\alpha\| = n+1, n+2\}.$$

Since $X(t,\cdot) : \mathbf{R}^N \to \mathbf{R}^N$ is a diffeomorphism, we can think of the push-forward $X(t)^*$. Then we have

$$X(t)^* = \text{Identity} + \sum_{i=0}^{d} \int_0^t X(s)^* V_i \circ dB^i(s)$$

as linear operators in $C^\infty(\mathbf{R}^N)$. So we have

$$\sum_{\alpha \in \mathcal{A}(n)} B^{\circ\alpha}(t) V_\alpha = \varphi_n(\tilde{X}(t)).$$

This proves our assrtion. Q.E.D.

Combining the previous proposition with Propositions 8 and 12, and applying the argument in Corollary 17, we have the following.

Proposition 20. *For any $n \geq 1$, there is a constant $C > 0$ depending only on n and d such that*

$$\sup_{x \in \mathbf{R}^N} E\Bigg[\Bigg| f(X(s,x)) - f\Bigg(\exp\Bigg(\varphi_n\Bigg(\sum_{\alpha \in \mathcal{A}_0} s^{\|\alpha\|/2} p^0_\alpha(B^{\circ\beta}(1),\, \beta \in \mathcal{A}_0) v^{[\alpha]}\Bigg)\Bigg)(x)\Bigg)\Bigg|^2\Bigg]^{1/2}$$
$$\leq C\Bigg(\sum_{\substack{\alpha \in \mathcal{A} \\ n+1 \leq \|\alpha\| \leq n(n+2)}} s^{\|\alpha\|/2} \|V_\alpha f\|_\infty\Bigg),$$
$$s \in (0,1],\ f \in C_b^\infty(\mathbf{R}^N;\mathbf{R}).$$

In particular for any $n \geq 1$, there is a constant $C' > 0$ depending only on n and d such that

$$\sup_{x \in \mathbf{R}^N} E\Bigg[\Bigg| X(s,x) - \exp\Bigg(\varphi_n\Bigg(\sum_{\alpha \in \mathcal{A}_0} s^{\|\alpha\|/2} p^0_\alpha(B^{\circ\beta}(1),\, \beta \in \mathcal{A}_0) v^{\alpha}\Bigg)(x)\Bigg|^2\Bigg]^{1/2}$$
$$\leq C' s^{(n+1)/2} \sum_{\substack{\alpha \in \mathcal{A}_0 \\ n+1 \leq \|\alpha\| \leq 2n(n+1)}} \|V_\alpha H\|_\infty, \quad s \in (0,1].$$

Here p^0_α are polynomials in Proposition 8.

§8. Proof of Theorems

By Theorem 15 and Proposition 20, we have the following.

Theorem 21. *Let $m \geq 1$. Let $\{Z_\alpha\ ;\ \alpha \in \mathcal{A}_0\}$ be m-moment similar family of random variables. Then for any $n \geq 1$, there is a constant $C > 0$ depending only on n and moments of Z_α, $\alpha \in \mathcal{A}_0$ such that*

$$\sup_{x\in\mathbf{R}^N} \Bigg| E[f(X(s,x))] - E\Bigg[f\Bigg(\exp\Bigg(\varphi_n\Bigg(\sum_{\alpha\in\mathcal{A}_0} s^{\|\alpha\|/2} p^0_\alpha(Z_\beta,\, \beta\in\mathcal{A}_0) v^{[\alpha]}\Bigg)\Bigg)(x)\Bigg)\Bigg]\Bigg|$$

$$\leq C\Bigg(\sum_{k=m+1}^{n(m+1)} s^{k/2}\|f\|_{V,k} + s^{(n+1)/2}\Bigg(\sum_{\substack{\alpha\in\mathcal{A}_0 \\ n+1\leq\|\alpha\|\leq 2n(n+1))}} \|V_\alpha H\|_\infty\Bigg)\|\nabla f\|_\infty\Bigg),$$

for $\varepsilon \in (0,1]$, $f \in C_b^\infty(\mathbf{R}^N;\mathbf{R})$.

Now Theorem 4 is an easy consequence of Theorems 15, 21 and Proposition 18.

Now let us prove Theorem 5. By Theorem 4, Corollary 2 and the argument in [6], we have the following.

Proposition 22. *For any $a \geq 1$, there is a constant $C > 0$ such that*

$$\|P_{t+s}f - Q_{(s)}P_t f\|_\infty \leq \frac{Cs^{(m+1)/2}}{t^{m/2}}\|\nabla f\|_\infty$$

for any $s, t \in (0,a]$ and $f \in C_b^\infty(\mathbf{R}^N;\mathbf{R})$ with $s \leq at$.

By this proposition, under the assumption in Theorem 5, we have

$$\begin{aligned}
&\|P_T f - Q_{(s_n)}Q_{(s_{n-1})}\cdots Q_{(s_1)}f\|_\infty \\
&\leq \sum_{k=1}^n \|Q_{(s_n)}\cdots Q_{(s_{k+1})}(P_{s_k} - Q_{(s_k)})P_{t_{k-1}}f\|_\infty \\
&\leq \sum_{k=2}^n \|P_{t_{k-1}+s_k}f - Q_{(s_k)}P_{t_{k-1}}f\|_\infty + \|P_{s_1}f - Q_{(s_1)}f\|_\infty.
\end{aligned}$$

It is easy to see that there is a constant $C > 0$ such that

$$\|P_s f - f\|_\infty \le C s^{1/2} \|\nabla f\|_\infty$$

and

$$\|Q_{(s)} f - f\|_\infty \le C s^{1/2} \|\nabla f\|_\infty$$

for any $s \in (0, 1]$ and $f \in C_b^\infty(\mathbf{R}^N; \mathbf{R})$. So we see that there is a constant $C > 0$ such that

$$\begin{aligned} &\|P_T f - Q_{(s_n)} Q_{(s_{n-1})} \cdots Q_{(s_1)} f\|_\infty \\ &\le C n^{-\gamma/2} \Big(1 + \sum_{k=1}^{n-1} \frac{k^{(m+1)(\gamma-1)/2}}{k^{m\gamma/2}} \Big) \|\nabla f\|_\infty \\ &= C n^{-\gamma/2} \Big(1 + \sum_{k=1}^{n-1} k^{(\gamma-m-1)/2} \Big) \|\nabla f\|_\infty. \end{aligned}$$

This implies our theorem.

References

[1] D. Bally and D. Talay, The law of the Euler scheme for stochastic differential equations I. Convergence rate of the distribution function, Probab. Theory Relat. Fields, **104** (1996), 43–60.

[2] G. BenArous, Flots et sèries de Taylor stochastiques, Probab. Theory Relat. Fields, **81** (1989), 29–77.

[3] F. Castell, Asymptotics expansions of stochastic flows, Probab. Theory Relat. Fields, **96** (1993), 225–239.

[4] N. Jacobson, "Lie algebras", Interscience, New York, 1962.

[5] S. Kusuoka, A remark on Malliavin Calculus, in preparation.

[6] S. Kusuoka and D. W. Stroock, Applications of Malliavin Calculus II, J. Fac. Sci. Univ. Tokyo Sect. IA Math., **32** (1985), 1–76.

[7] S. Kusuoka and D. W. Stroock, Applications of Malliavin Calculus III, J. Fac. Sci. Univ. Tokyo Sect. IA Math., **34** (1987), 391–442.

[8] R. Léandre, Applications quantitatives et géométriques du calcul de Malliavin, in "stochastic integral", (M. Métivier and S. Watanabe, eds.), Lect. Notes in Math., **1322**, Springer, Berlin, 1988, pp. 109–133.

[9] R. S. Strichartz, The Cambell-Baker-Hausdorff-Dynkin formula and solutions of differential equations, J. Funct. Anal., **72** (1987), 320–345.

[10] S. Takanobu, Diagonal short time asyptotics of heat kernels for certain dgenerate second order differential operators of Hörmander type, Publ. Res. Inst. Math. Sci., **24** (1988), 169–203.

Graduate School of Mathematical Sciences
The University of Tokyo
3-8-1 Komaba
Meguro-ku
Tokyo, 153-8914
Japan

Advanced Studies in Pure Mathematics 31, 2001
Taniguchi Conference on Mathematics Nara '98
pp. 167–178

Representation Theory in Characteristic p

George Lusztig

Let k be an algebraically closed field of characteristic p. (Thus, p is either 0 or a prime number.) Let G be a group which is at the same time an affine algebraic variety over k (that is, an algebraic group over k). A *representation* of G is a homomorphism $\rho : G \longrightarrow GL(V)$ of G into the group of automorphisms of a finite dimensional k-vector space V which is at the same time a morphism of algebraic varieties. We also say that V is a *G-module.* We say that V is *irreducible* if $V \neq 0$ and there is no subspace V' of V (other than 0 or V) such that $\rho(g)V' \subset V'$ for all $g \in G$.

Let $\mathfrak{g}$ be the Lie algebra of G. A *representation* of $\mathfrak{g}$ is a k-linear map $\tau : \mathfrak{g} \longrightarrow \mathrm{End}(V)$ where V is a finite dimensional k-vector space such that

$$\tau([\xi, \xi']) = \tau(\xi)\tau(\xi') - \tau(\xi')\tau(\xi)$$

for all $\xi, \xi' \in \mathfrak{g}$. We also say that V is a *$\mathfrak{g}$-module.* The notion of irreducibility of a $\mathfrak{g}$-module is defined in the same way as in the group case.

We will assume that G is connected, almost simple (that is, G has finite centre and G modulo its centre is a simple group) and simply connected (in a suitable sense). Chevalley [C] proved the remarkable result that the classification of such G is the same as the classification of simple complex Lie algebras (achieved by É. Cartan and Killing). Thus, G must be a special linear group, a symplectic group, a spin group or one of five exceptional groups.

The problem that we will discuss in this paper is that of classifying the irreducible G-modules and $\mathfrak{g}$-modules and that of understanding as much as possible the structure of those irreducible modules. Work on these problems have occupied mathematicians throughout much of this century. We will review some of this work. Towards the end of the paper we will engage in speculation about possible future directions.

Received May 29, 1999.

Supported by the Ambrose Monnel Foundation and the National Science Foundation

§1. Algebraic groups

1.1. Characteristic 0

In this subsection we assume that $p = 0$. In this case, taking the differential defines an equivalence between the category of G-modules and the category of $\mathfrak{g}$-modules. É. Cartan [Ca] showed that the isomorphism classes of irreducible $\mathfrak{g}$-modules (hence also those of irreducible G-modules) are naturally indexed by a set X given by the intersection of an open cone $E^>$ in an euclidean vector space E with a lattice in E. Let V_x be the irreducible G-module corresponding to $x \in X$. H. Weyl [W] gave an explicit character formula and a dimension formula for V_x. An algebraic geometric construction for V_x was given by Borel and Weil (see [S]).

For example, for $G = SL_n(k)$, we can take

$$E = \mathbf{R}^n/\mathbf{R}(1, \dots, 1),$$
$$X = \{(a_1, \dots, a_n) \in \mathbf{Z}^n/\mathbf{Z}(1, \dots, 1) \mid a_1 > a_2 > \cdots > a_n\}.$$

If $G = SL_2(k)$ and $(a, 0) \in X$, we can take $V_{a,0}$ to be the space of homogeneous polynomials in two variables t_1, t_2 with coefficients in k of degree $a - 1$. (This is naturally a G-module.)

1.2. Late 1950's to early 1970's

In the rest of this paper we assume that p is > 0 and sufficiently large, unless we specify otherwise.

Taking the differential defines a functor from the category of G-modules to the category of $\mathfrak{g}$-modules but this is by no means an equivalence. A very small part of the characteristic 0 theory survives: the definition of V_x (for each $x \in X$) given by the Borel-Weil construction still makes sense. It provides a G-module which has the same structure (in particular the same dimension) as in characteristic 0, although it is no longer irreducible in general. Chevalley [C] has shown that V_x contains a unique irreducible G-submodule, denoted by L_x and that $\{L_x \mid x \in X\}$ is a set of representatives for the isomorphism classes of irreducible G-modules.

For example, if $G = SL_2(k)$, then $V_{a,0}$ is defined as in 1.1 in terms of our field k and $L_{a,0}$ is the subspace of $V_{a,0}$ spanned by $(a-1)$-th powers of linear polynomials. Note that $L_{a,0}$ is not necessarily equal to $V_{a,0}$. The following table gives (for $p = 3$) the value $\dim V_{a,0} = a$ for $a = 1, 2, \dots, 20$ (in the first row) and the corresponding value of $\dim L_{a,0}$

(in the second row):

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2	3	2	4	6	3	6	9	2	4	6	4	8	12	6	12	18	3	6

Note that $\dim L_{a,0}$ is linear in a, for a in a fixed interval $(lp,(l+1)p)$ (even in $(lp,(l+1)p]$). The ends of these intervals (the points lp) make sense for $SL_n(k)$ as well; they form the family $\mathcal{F}$ of affine hyperplanes in E:

$$x_i - x_j = lp \quad \text{for some } i \neq j \text{ in } [1,n] \text{ and some integer } l.$$

$\mathcal{F}$ is similarly defined for any G (in terms of roots). For each $H \in \mathcal{F}$ let $\sigma_H : E \to E$ be the map which takes any point of E to its mirror image with respect to H. Let W be the group of affine transformations of E generated by the reflections σ_H. We now remove from E the points that lie in at least one hyperplane of $\mathcal{F}$. The resulting set is a disjoint union of open simplices called *alcoves* (the analogues of the open intervals $(lp,(l+1)p)$ for $SL_2(k)$). Let $\mathcal{A}$ be the set of alcoves; let $\mathcal{A}^>$ be the set of alcoves contained in $E^>$.

For any $x \in X$ which does not lie in any hyperplane of $\mathcal{F}$ and any $A \in \mathcal{A}^>$ we denote by x_A the unique point of $X \cap A$ which is in the W-orbit of x.

Let R_G be the free abelian group with basis $\{L_x \mid x \in X\}$. Any G-module M gives rise to an element

$$\sum_x (L_x : M) L_x \in R_G$$

where $(L_x : M)$ is the number of times that L_x appears in a composition series of M. We sometimes write M instead of $\sum_x (L_x : M) L_x \in R_G$. Thus the elements $V_x \in R_G$ are well defined for $x \in X$. They again form a basis of R_G.

The following result has been conjectured by Verma [V] and proved by Jantzen [J1].

There exists a function $\mathcal{A}^> \times \mathcal{A}^> \to \mathbf{Z}$ denoted by $A, B \mapsto (A,B)$ such that for any $B \in \mathcal{A}^>$ and any $x \in X \cap B$ we have

$$L_x = \sum_{A \in \mathcal{A}^>} (A : B) V_{x_A} \in R_G.$$

(In particular, for fixed B we have $(A,B) = 0$ for all but finitely many A.) Since the V_{x_A} can be considered as known, this shows that the dimension of L_x (and other information about L_x) will be known provided the quantities (A, B) are known. Jantzen also showed that even the L_x with $x \in X$ on some hyperplane in $\mathcal{F}$ are explicitly determined by the (A, B).

We define $A^+ \in \mathcal{A}^>$ by the condition that 0 is in the closure of A^+. Let h be the Coxeter number of G. (Thus, $h = n$ for $SL_n(k)$). If the (A, B) are known whenever $A, B \in \mathcal{A}^>$ are contained in $(p-h)A^+$ then, by a tensor product theorem of Steinberg [St], the (A, B) will be determined for general A, B.

Thus, the problem of understanding L_x is reduced to the determination of the unknown quantities (A, B) with A, B contained in $(p-h)A^+$.

1.3. Late 1970's

W together with the reflections σ_H (where $H \in \mathcal{F}$ contains a codimension 1 face of $A^- = -A^+$) is a *Coxeter group* of affine type. Hence by a general definition which applies to any Coxeter group, to any two elements $y, w \in W$ one can attach a polynomial $P_{y,w} \in \mathbf{Z}[v]$ (v an indeterminate) as in [KL1].

For each $w \in W$ we set $A_w = w(A^-)$. Then $w \mapsto A_w$ is a bijection $W \xrightarrow{\sim} \mathcal{A}$. The following was conjectured in [L1].

(a) *Assume that $A, B \in \mathcal{A}$ are contained in $(p-h)A^+$. Define $y, w \in W$ by $A = A_y$, $B = A_w$. Let $m_{A,B}$ be the number of hyperplanes in $\mathcal{F}$ that separate A from B. Then*

$$(A, B) = (-1)^{m_{A,B}} P_{y,w}(1).$$

(See [KL1] for a precursor of this conjecture, which involves only the $P_{y,w}$ where y, w preserve 0.)

1.4. Late 1980's

Around 1985, quantum groups appeared on the scene, due to the work of Drinfeld and Jimbo. These were some strange deformations of complex algebraic groups depending on a parameter $v \in \mathbf{C}^*$. In the original definition v had to be generic, but it turned out that a good definition can be given for arbitrary $v \in \mathbf{C}^*$. The case where v is root of 1 was particularly interesting and in [L3] I found that in this case the representation theory of the quantum group is very similar to that of G in characteristic > 0. For example, in the case of SL_2, in the table in 1.2 one can add a new row (in between the two rows of the original table) giving the value of the dimension of the quantum analogue of $L_{a,0}$ in the case where $v^3 = 1$, $v \neq 1$. One obtains

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2	3	2	4	6	3	6	9	4	8	12	5	10	15	6	12	18	7	14
1	2	3	2	4	6	3	6	9	2	4	6	4	8	12	6	12	18	3	6

Note that the new row (quantum case) has values which are sandwiched between those in the first and third row. This suggests that the quantum group is some kind of stepping stone between the algebraic groups in characteristic 0 and p. This led me in 1988 to formulate four steps in a possible proof of Conjecture 1.3 (a) (at least in types A, D, E).

(I) Show that the process of "reduction modulo p" from representations of the quantum group at a p-th root of 1 to representations of G in characteristic p, takes irreducible modules to irreducible modules (in a suitable range).

(II) Show that the representations of the quantum group at a p-th root of 1 are closely related to certain representations of the corresponding affine Lie algebra with central charge $-p-h$.

(III) Show that the characters of the irreducible highest weight representations of the affine Lie algebra with central charge $-p-h$ can be related to the intersection cohomology of Schubert varieties in an affine flag manifold.

(IV) Show that the intersection cohomology in (III) is computed by the polynomials $P_{y,w}$ of [KL1].

1.5. Early 1990's

Steps (I), (II), (III) were attacked by three teams on three continents. (Step (IV) was already known from [KL2]. A simpler version of step (III) dealing with finite dimensional Lie algebras was also known since the early 1980's.) Thus, (I) was solved (for p in the complement of an unknown finite set) by Andersen, Jantzen and Soergel [AJS]; (II) was solved (for p in the complement of a known finite set) in [KL3]; (III) was solved by Kashiwara and Tanisaki [KT1].

For G of type B, C, F_4, G_2 the four steps had to be modified (see [L4]). In the modified framework, steps (I) and (II) were covered by the existing works [AJS], [KL3]. But the solution of step (II) required a new work of Kashiwara and Tanisaki [KT2]; even step (IV) presented a new problem (solved in [L4]).

The combination of these works provides a solution of the Conjecture 1.3 (a) hence of the problem of describing the structure of the irreducible G-modules L_x (for p in the complement of an unknown finite set).

The last restriction on p is very unsatisfactory. It would be very desirable if somebody will remove the restriction on p from [AJS] and also the (milder) restriction on p from [KL3, IV].

§2. Lie algebras

2.1.

The representation theory of $\mathfrak{g}$ is in a much poorer state than that of G. Here not even a classification of the irreducible modules is available. But a conjectural picture is beginning to emerge, promising a very rich theory.

2.2. Early 1970's

There is a canonical map $\xi \mapsto \xi^{[p]}$ of $\mathfrak{g}$ into itself. For example, if $G = SL_n(k)$ then $\mathfrak{g} = \mathrm{End}(k^n)$ and $\xi^{[p]}$ is the p-th power of ξ as an endomorphism of k^n.

Let $\epsilon : \mathfrak{g} \longrightarrow k$ be a linear form. Following Weisfeiler and Kac [WK], we consider the quotient U_ϵ of the enveloping algebra U of $\mathfrak{g}$ by the two-sided ideal generated by the central elements $\xi^p - \xi^{[p]} - \epsilon(\xi)^p 1$ for various $\xi \in \mathfrak{g}$. Then U_ϵ is a finite dimensional algebra and any simple $\mathfrak{g}$-module can be regarded as a module over U_ϵ for a unique ϵ as above [WK]. Hence the problem of understanding all $\mathfrak{g}$-modules is reduced to the problem of understanding all U_ϵ-modules for any linear function $\epsilon : \mathfrak{g} \rightarrow k$. This last problem for general ϵ can be reduced (see [WK]) to the special case where ϵ is nilpotent. (We identify a linear form on $\mathfrak{g}$ with an element of $\mathfrak{g}$ by the Killing form.)

2.3. Late 1970's to 1997

From now on we assume that $\epsilon \in \mathfrak{g}$ is nilpotent. Let C' be a maximal torus of the centralizer of ϵ in G and let $\bar{C}'$ be the image of C' in the adjoint group of G. Following an idea of Jantzen [J2] (see also [FP], [J4]) one can consider the category $\mathcal{C}_\epsilon$ of U_ϵ-modules which are also $\bar{C}'$-modules, the two module structures being compatible in a natural way; then one studies the simple objects of $\mathcal{C}_\epsilon$ instead of the simple U_ϵ-modules. In the case where ϵ is regular nilpotent inside a Levi subalgebra of some parabolic algebra, the classification of the simple objects of $\mathcal{C}_\epsilon$ has been obtained by Friedlander and Parshall [FP]. A conjecture which would describe much of the structure of these simple objects was given in [L5]. Some examples computed by Jantzen [J3] give support to the conjecture.

2.4. Speculation

Returning to a general nilpotent $\epsilon \in \mathfrak{g}$, we note that $\mathcal{C}_\epsilon$ is a direct sum of indecomposable categories, or blocks. Let us fix a generic block of $\mathcal{C}_\epsilon$. Let $\mathbf{I}$ be an indexing set for the simple objects in this block. For

$\mathbf{i} \in \mathbf{I}$, let $L_{\mathbf{i}}$ be the corresponding simple object of $\mathcal{C}_\epsilon$ and let $Q_{\mathbf{i}}$ be the corresponding indecomposable projective object. For $\mathbf{i} \in \mathbf{I}$ we have

$$Q_{\mathbf{i}} = \sum_{\mathbf{i}' \in \mathbf{I}} n_{\mathbf{i},\mathbf{i}'} L_{\mathbf{i}'}$$

in the appropriate Grothendieck group, where $n_{\mathbf{i},\mathbf{i}'} \in \mathbf{N}$ are zero for all but finitely many $\mathbf{i}'$.

For simplicity we assume that the centralizer of ϵ in the adjoint group of G is connected. In [L6, 14.4, 14.5] I stated a conjecture which gives a parametrization of $\mathbf{I}$ and an interpretation of the matrix $(n_{\mathbf{i},\mathbf{i}'})$ in terms of some apparently totally unrelated K-theoretic objects. (I thank Jantzen for his criticism of that conjecture.)

Here I will state a revised form of the original conjecture. To do this I will review some K-theoretic constructions from [L6].

Let $\mathbf{G}$ be an algebraic group of the same type as G, but over $\mathbf{C}$ instead of k. Let $\mathbf{g}$ be the Lie algebra of $\mathbf{G}$. Let r be the rank of $\mathbf{g}$. Let $\mathcal{B}$ be the variety of all Borel subalgebras of $\mathbf{g}$. Let $e \in \mathbf{g}$ be a nilpotent element of $\mathbf{g}$ of the same type as $\epsilon \in \mathfrak{g}$. We can complete e to an $\mathfrak{sl}_2$-triple (e, h, f) in $\mathbf{g}$. Let $\mathcal{B}_e = \{\mathfrak{b} \in \mathcal{B} \mid e \in \mathfrak{b}\}$.

Following Slodowy, we define Λ_e to be the variety consisting of all pairs $(y, \mathfrak{b})$ where $\mathfrak{b} \in \mathcal{B}$ and y is a nilpotent element of $\mathfrak{b}$ such that $[y - e, f] = 0$. Then $\mathcal{B}_e$ is naturally imbedded in Λ_e by $j : \mathfrak{b} \mapsto (e, \mathfrak{b})$. Moreover, Λ_e is a smooth irreducible variety. Let C be a maximal torus of the simultaneous centralizer of e, h, f in $\mathbf{G}$. Let $H = C \times \mathbf{C}^*$. We regard H as a subgroup of $\mathbf{G} \times \mathbf{C}^*$ as in [L6, 11.1].

Now $\mathbf{G} \times \mathbf{C}^*$ acts on $\mathcal{B}$ by $(g, \lambda) : \mathfrak{b} \mapsto Ad(g)\mathfrak{b}$ and on $\mathbf{g}$ by $(g, \lambda) : y \mapsto \lambda^{-2} Ad(g) y$.

This restricts to an H-action on $\mathcal{B}$ and one on $\mathbf{g}$. The product H-action on $\mathbf{g} \times \mathcal{B}$ leaves the subvariety Λ_e stable and leaves the subvariety $\mathcal{B}_e$ of Λ_e stable. Note also that the H-action on $\mathbf{g}$ restricts to an H-action on the centralizer $\mathfrak{z}$ of f in $\mathbf{g}$.

Hence the equivariant K-groups $K_H(\mathcal{B}_e)$, $K_H(\Lambda_e)$ (based on H-equivariant coherent sheaves on $\mathcal{B}_e$, Λ_e) are well defined (these are naturally R_H-modules where R_H is the representation ring of H.) Moreover, we can form

$$\mathbf{d} = \sum_{t \geq 0} (-1)^t \mathfrak{z}^{(t)} \in R_H$$

where $\mathfrak{z}^{(t)}$ is the t-th exterior power of $\mathfrak{z}$. One can show that

(a) *the R_H-linear map $j_* : K_H(\mathcal{B}_e) \longrightarrow K_H(\Lambda_e)$ induced by the closed imbedding $j : \mathcal{B}_e \longrightarrow \Lambda_e$ induces an isomorphism after tensoring by $R_H[\mathbf{d}^{-1}]$.*

Let $(\ |\ )_{\Lambda_e}$ be the pairing on $K_H(\Lambda_e)$ defined in [L6, 12.16]. By definition, this pairing takes values in the quotient field of $\mathfrak{R}$; but, due to (a), its values are actually in $R_H[\mathbf{d}^{-1}]$. Let

$$^{-}: K_K(\Lambda_e) \to K_H(\Lambda_e)$$

be as in [L6, 12.16]. (The definition involves Serre-Grothendieck duality on Λ_e.)

(The definition of $(\ |\ )_{\Lambda_e}$ and that of $^{-}$ involve an integer $d(e)$. In [L6, 12.9] one should replace $d(e) = (1/2)\dim Ad(L)e$ by $d(e) = (1/2)\dim Ad(G)e$.)

Note that $R_H[\mathbf{d}^{-1}]$ is naturally imbedded in the ring $\mathfrak{U}_H$ of power series in an indeterminate v^{-1} with coefficients in the ring R_C. (Here v is identified with the standard generator of $R_{\mathbf{C}^*}$.) Indeed, $R_H = R_C[v, v^{-1}] \subset \mathfrak{U}_H$ and $\mathbf{d}$ is a product of factors of form $1 - v^c\alpha$ with $c < 0$ and α a character of C; hence $\mathbf{d}^{-1} \in R_C[[v^{-1}]]$. Let

$$\delta : \mathfrak{U}_H \to \mathbf{Z}((v^{-1}))$$

be the group homomorphism defined by $\sum_{n\in\mathbf{Z}} p_n v^n \mapsto \sum_{n\in\mathbf{Z}} \tilde{p}_n v^n$ where $p_n \in R_C$ and $p \mapsto \tilde{p}$ is the group homomorphism which sends a non-trivial representation of C to 0 and sends the unit representation of C to 1.

Let $\bar{C}$ be the image of C into the adjoint group of $\mathbf{G}$ and let $\bar{H} = \bar{C} \times \mathbf{C}^*$.

Following [L6, 12.18] we define $\mathbf{B}^{\pm}_{\Lambda_e}$ to be the set of all elements $\xi \in K_H(\Lambda_e)$ such that

$$\bar{\xi} = \xi \quad \text{and} \quad \delta(\xi\,|\,\xi)_{\Lambda_e} \in 1 + v^{-1}\mathbf{Z}[[v^{-1}]].$$

Following [L6, 12.22] we define $\mathbf{B}^{\pm}_{\Lambda_e,\mathrm{ad}}$ as the intersection of $\mathbf{B}^{\pm}_{\Lambda_e}$ with

$$\mathrm{Im}(K_{\bar{H}}(\Lambda_e) \to K_H(\Lambda_e)) = K_{\bar{H}}(\Lambda_e).$$

Note that $\xi \mapsto -\xi$ is an involution of $\mathbf{B}^{\pm}_{\Lambda_e}$ and of $\mathbf{B}^{\pm}_{\Lambda_e,\mathrm{ad}}$. Let $\mathbf{B}_{\Lambda_e}$, $\mathbf{B}_{\Lambda_e,\mathrm{ad}}$ be the corresponding sets of orbits of this involution. One can show that

(b) $\mu_{\mathbf{b},\mathbf{b}'} := (1-v^2)^{-r}\delta(\mathbf{db}\,|\,\mathbf{b}')_{\Lambda_e} = (1-v^2)^{-r}\delta(\mathbf{b}\,|\,\mathbf{db}')_{\Lambda_e} \in \mathbf{Z}[v, v^{-1}]$

for all $\mathbf{b}, \mathbf{b}' \in \mathbf{B}^{\pm}_{\Lambda_e}$. We conjecture that

(c) *there exists a canonical map* $\zeta : \mathbf{B}^{\pm}_{\Lambda_e,\mathrm{ad}} \to \mathbf{I}$ *which induces a bijection* $\mathbf{B}_{\Lambda_e,\mathrm{ad}} \xrightarrow{\sim} \mathbf{I}$ *such that*

$$\pm\mu_{\mathbf{b},\mathbf{b}'}(1) = n_{\zeta(\mathbf{b}),\zeta(\mathbf{b}')}$$

for all $\mathbf{b}, \mathbf{b}' \in \mathbf{B}^{\pm}_{\Lambda_e,\mathrm{ad}}$.
The sign is taken so that $\pm\mu_{\mathbf{b},\mathbf{b}'}(1) \geq 0$. It is likely that in (b) we have either $\mu_{\mathbf{b},\mathbf{b}'} \in \mathbf{N}[v^{-1}]$ or $-\mu_{\mathbf{b},\mathbf{b}'} \in \mathbf{N}[v^{-1}]$.

Note that both $\mathbf{I}$ and $\mathbf{B}_{\Lambda_e,\mathrm{ad}}$ have natural actions of a free abelian group of rank $\dim C$ and the bijection in (c) should be compatible with these actions.

This conjecture is actually true if $e = 0$ (for those p for which 1.3 (a) holds); this follows from results in [L6].

2.5.

Assume that e is regular (nilpotent). In this case $H = \mathbf{C}^*$ and

$$\mathbf{d} = \prod_{i \in [1,r]} (1 - v^{-2e_i - 2}),$$

where $e_1, \ldots, e_r$ are the exponents of $\mathbf{G}$. Also, $\mathbf{B}^{\pm}_{\Lambda_e} = \mathbf{B}^{\pm}_{\Lambda_e,\mathrm{ad}}$ consists of $\pm\mathbf{b}$ where $\mathbf{b} = \mathcal{O}_{\Lambda_e}$ is the structure sheaf of the point Λ_e. Hence $\mathbf{B}_{\Lambda_e,\mathrm{ad}}$ consists of a single element. We have $\mu_{\mathbf{b},\mathbf{b}} = \mathcal{P}$ where

(a) $$\mathcal{P} = \prod_{i \in [1,r]} \frac{1 - v^{-2e_i - 2}}{1 - v^{-2}} \in \mathbf{N}[v^{-1}].$$

Hence

$$\mu_{\mathbf{b},\mathbf{b}}(1) = |\bar{W}|$$

where $\bar{W}$ is the Weyl group of $\mathbf{G}$. Thus 2.4 (c) holds in this case. (Compare [J4, 10.10].)

2.6.

Assume that e is subregular and $\mathbf{G}$ is of type D or E. In this case $H = \mathbf{C}^*$. One can check that

$$\mathbf{d} = \prod_{i \in [1,r+2]} (1 - v^{-2s_i})$$

where $s_1, s_2, \ldots, s_{r+2}$ is:
$2, 3, 4, 5, 6, 6, 8, 9$ (type E_6),
$2, 4, 6, 6, 8, 9, 10, 12, 14$ (type E_7),
$2, 6, 8, 10, 12, 14, 15, 18, 20, 24$ (type E_8),
$2, 2, 4, 6, \ldots, 2n-4, n-2, n-1, n$ (type D_n).
One can also give a closed formula for $\mathbf{d}$:

(a) $$\mathbf{d} = (1 - v^{-2})^r \mathcal{P} \det(\tilde{A}) \det(A)^{-1}$$

where $\mathcal{P}$ is as in 2.5 (a), $\tilde{A}$ is the square matrix indexed by the vertices of the affine Coxeter diagram J with j, j' entry equal to $1+v^{-2}$ if $j=j'$, equal to $-v^{-1}$ if j, j' are joined in that diagram and equal to 0 in the remaining cases; A is the analogous matrix defined in terms of the ordinary Coxeter diagram.

According to [L7], in this case $\mathbf{B}^{\pm}_{\Lambda_e} = \mathbf{B}^{\pm}_{\Lambda_e,\mathrm{ad}}$ consists of elements $\pm\mathbf{b}_j$ ($j \in J$) where $\mathbf{b}_j$ are certain vector bundles on Λ_e and the matrix $((\mathbf{b}_j \mid \mathbf{b}_{j'})_{\Lambda_e})$ is just the inverse of $\tilde{A}$ above. Now $\tilde{A}^{-1}$ can be computed by the method of [LT]. For $j, j' \in J$, we denote by $[j, j']$ be the subset of J consisting of all vertices that lie on the geodesic joining j, j' in the affine Coxeter graph. Let $\tilde{A}_{j,j'}$ be the submatrix of $\tilde{A}$ obtained by removing all rows and columns indexed by some element of $[j, j']$. Let

$$\tilde{\alpha}_{j,j'} = \det(\tilde{A}_{j,j'}) \in \mathbf{Z}[v^{-1}].$$

Then we have

$$(\mathbf{b}_j \mid \mathbf{b}_{j'})_{\Lambda_e} = \tilde{\alpha}_{j,j'} \det(\tilde{A})^{-1}.$$

Using this and (a), we see that

$$\mu_{\mathbf{b}_j,\mathbf{b}_{j'}} = \mathcal{P}\det(A)^{-1}\tilde{\alpha}_{j,j'} \in \mathbf{Z}[v^{-1}].$$

Hence in this case, 2.4 (c) predicts that $\mathbf{I}$ may be identified with J in such a way that the multiplicities $n_{j,j'}$ are given by

$$n_{j,j'} = |\bar{W}| z_{j,j'} z_{j_0,j_0}^{-1}$$

where $z_{j,j'}$ is the order of the centre of the simply connected group with Coxeter graph $J-[j,j']$ (full subgraph of J) and j_0 is the unique element of J that is not a vertex of the ordinary Coxeter graph.

2.7.

Since $\mathbf{B}^{\pm}_{\Lambda_e,\mathrm{ad}}$ is expected to be a signed basis of the $R_{\mathbf{C}^*}$-module $K_{\bar{H}}(\Lambda_e)$ [L6, 12.23(a)], we see that, in the case where $C=\{1\}$, the Conjecture 2.4 (c) predicts that the number of elements in the block $\mathbf{I}$ is the sum of the Betti numbers of $\mathcal{B}_e$.

2.8.

We expect that the set $\mathbf{B}_{\Lambda_e}$ and the quantities $\mu_{\mathbf{b},\mathbf{b}'}$ are intimately related with the combinatorics of the two-sided cell $\mathbf{c}$ in the extended affine Weyl group attached in [L2, 4.8] to the $\mathbf{G}$-orbit of e.

2.9.

The most urgent task would be to test the Conjecture 2.4 (c) in the example in 2.6. Assuming that the conjecture passes this test, one can try to imagine whether a successsion of steps analogous to (I)–(IV) in 1.4 can be used to prove it. At least step (I) makes sense; one should use the version of quantum group at a root of 1 studied by De Concini and Kac in [DK]. One can also expect that there is a corresponding class of representations of the appropriate affine Lie algebra with negative central charge which are connected with the two-sided cell **c** in 2.9.

Remark added 5.22.1999. After this paper was written, Jantzen (*Subregular nilpotent representations of Lie algebras in prime characteristic*, preprint April 1999) has shown that Conjecture 2.4 (c) holds in the example in 2.6.

References

[AJS] H. H. Andersen, J. C. Jantzen and W. Soergel, Representations of quantum groups at roots of unity and of semisimple groups in characteristic p: independence of p, Astérisque, **220** (1994).

[Ca] É. Cartan, "Oeuvres complètes", Gauthier-Villars, Paris, 1952-54.

[C] C. Chevalley, "Classification des groupes de Lie algébriques", Institut H. Poincaré, 1956-58.

[DK] C. De Concini and V. Kac, Representations of quantum groups at roots of 1, in "Operator Algebras, Unitary Representations, Enveloping Algebras and Invariant Theory (Colloque Dixmier)", (eds. A. Connes et al.), Progress in Math., Birkhäuser, Boston, 1990.

[FP] E. M. Friedlander and B. J. Parshall, Deformations of Lie algebra representations, Amer. J. Math., **112** (1990), 375–395.

[J1] J. C. Jantzen, Zur Characterformel gewisser Darstellungen halbeinfacher Gruppen und Lie Algebren, Math. Z., **140** (1974), 127–149.

[J2] J. C. Jantzen, Über Darstellungen höherer Frobenius Kerne halbeinfacher algebraischer Gruppen, Math. Z., **164** (1979), 271–292.

[J3] J. C. Jantzen, Representations of $\mathfrak{so}_5$ in prime characteristic. preprint Aarhus series 1997:13.

[J4] J. C. Jantzen, Representations of Lie algebras in prime characteristic, in "Representation theories and algebraic geometry", (ed. A. Broer and A. Daigneault), Kluwer, Dordrecht, 1998.

[KT1] M. Kashiwara and T. Tanisaki, Kazhdan-Lusztig conjecture for affine Lie algebras with negative level, Duke Math. J., **77** (1995), 21–62.

[KT2] M. Kashiwara and T. Tanisaki, Kazhdan-Lusztig conjecture for affine Lie algebras with negative level, II: Non-integral case, Duke Math. J., **84** (1996), 771–813.

[KL1] D. Kazhdan and G. Lusztig, Representations of Coxeter groups and Hecke algebras, Invent. Math., **53** (1979), 165–184.

[KL2] D. Kazhdan and G. Lusztig, Schubert varieties and Poincaré duality, Proc. Symp. Pure Math., **36** (1980), 185–203, Amer. Math. Soc.

[KL3] D. Kazhdan and G. Lusztig, Tensor structures arising from affine Lie algebras, I–IV, J. Amer. Math. Soc., **6**, **7** (1993, 1994).

[L1] G. Lusztig, Some problems in the representation theory of finite Chevalley groups, Proc. Symp. Pure Math., **37** (1980), 313–317.

[L2] G. Lusztig, Cells in affine Weyl groups, IV, J. Fac. Sci. U. Tokyo, Sec. IA, **36** (1989), 297–328.

[L3] G. Lusztig, Modular representations and quantum groups, Contemp. Math., **82** (1989), 59–77.

[L4] G. Lusztig, Monodromic systems on affine flag manifolds, Proc. Roy. Soc. Lond. (A), **445** (1994), 231–246; Errata,**450** (1995), 731–732.

[L5] G. Lusztig, Periodic W-graphs, Represent. Th. (electronic), **1** (1997), 207–279.

[L6] G. Lusztig, Bases in equivariant K-theory, Represent. Th. (electronic), **2** (1998), 298–369.

[L7] G. Lusztig, Subregular nilpotent elements and bases in K-theory, Canad. J. Math. (1999), to appear.

[LT] G. Lusztig and J. Tits, The inverse of a Cartan matrix, Analele Univ. Timişoara, **30** (1992), 17–23.

[S] J.-P. Serre, Représentations linéaires et éspaces homogènes Kählériennes des groupes de Lie compacts, Séminaire Bourbaki 100, Mai 1954, Benjamin, New York and Amsterdam.

[St] R. Steinberg, Representations of algebraic groups, Nagoya Math. J., **22** (1963), 33–56.

[V] D.-N. Verma, Role of affine Weyl group in the representation theory of algebraic Chevalley groups and their Lie algebras, in "Lie groups and their representations", Halstead, New York, 1975, pp. 673–705.

[WK] B. Weisfeiler and V. Kac, Irreducible representations of Lie p-algebras, Funkt. Analiz i ego pril., **5** (1971), 28–36.

[W] H. Weyl, Theorie der Darstellung kontinuierlicher halb-einfacher Gruppen durch lineare Transformationen, III, Math. Z., **25** (1926), 377–395.

Department of Mathematics
M.I.T.
Cambridge, MA 02139
U.S.A
and Institute for Advanced Study
Princeton, NJ 08540
U.S.A
`gyuri@math.mit.edu`

Advanced Studies in Pure Mathematics 31, 2001
Taniguchi Conference on Mathematics Nara '98
pp. 179–209

Probability and Geometry

Paul Malliavin

Summary

Fifty years back, Kiyosi Itô started the theory of *Stochastic Differential Equations* (SDE) in a fully geometric framework; in particular he constructed the heat process associated to an elliptic operator on a manifold by patching together Stochastic Differential Equations, the change

Received October 18, 1999.

of charts being mastered by Itô Calculus. We want to present here some aspects of the fastly growing exchange of concepts and methods existing betweeen Probability and Geometry in finite or infinite dimension. The topic is so wide that I am not able to fully cover it. I shall limitate myself to the subjects where I have get some personnal acquaintance in these last twenty five years. I gave to some parts of this paper an autobiographical flavour willing to emphasize here the decisive and unvaluable support that the japanese mathematical community bring to my program all around these years and singularly the Master for all of us: Professor Kiyosi Itô. To be called again to contribute to the Proceedings of a major Japanese Conference is for me a double privilege: it gives me the opportunity to recall the past and the possibility to act in the present.

What are the mathematical facts from which it could be possible to a priori advocate some relations between Probability and Geometry?

A major topic in Global Differential Geometry is the *passage from the infinitesimal to the global.* The elliptic operators constitute a basic tool for this passage: let us mention for instance the computation of the Atiyah-Singer Index by integrating a differential form built from curvature tensors or the cohomology vanishing theorems through positivity in the spirit of Bochner-Kodaira. The advantage of a probabilistic treatement of these elliptic operators is that the processes associated furnishes a *family of curves, geometrically highly significant.* In classical Differential Geometry some theorems are obtained by making *infinitesimal variations of the geodesic flow*: for instance a stricly positive lower bound of the sectional curvature leads to an a priori bound of the diameter of the manifold. It will be possible to proceed in the same way with the trajectories of the process by computing theirs infinitesimal variations by the Stochastic Calculus of Variations and these variations will be expressed in terms of curvatures.

In Probability the study of the joint distribution of a finite number of random variables involves *Geometric Integration* theory. Is it posssible to develop a similar programm directly on the probability space? If so we shall obtain conditional expectations defined for ALL VALUES of the conditioning, significant result from a purely probabilistic point of view.

Many others links between Probability and Geometry will appear below, in particular a full "Geometrization" of the concept of Stochastic Integral.

§1. Passage from the infinitesimal to the global through Com-

parison Lemmas for SDE

In 1972 I was looking to estimate the decreasing at the boundary of the complex Green function of a strictly pseudo-convex domain of C^n; finally I got that this decreasing goes as d^n, where d denotes the distance at the boundary; this result appeared soon as a key step for determining the behaviour of the zeroes of a function in the Nevanlinna class. My original proof, which still do not seem to have an alternative approach, was based on a comparison lemma for SDE [Ma74].

Given an abstract manifold M, an exhaustion function p and an elliptic operator Δ defined on M, the *heat conduction coefficient* and the *projected heat conduction coefficients* are defined respectively by

$$\begin{aligned} &a(m) := \frac{\Delta p}{\|\nabla p\|^2}(m); \\ &a^+(\xi) := \sup_{m\in p^{-1}(\xi)} (a(m)); \quad a^-(\xi) := \inf_{m\in p^{-1}(\xi)} (a(m)). \end{aligned} \tag{1.1}$$

Denote $x_\omega(t)$ the diffusion on M associated to Δ, denote $\xi_\omega^+(t^*)$ (resp. $\xi_\omega^-(t^*)$) the diffusion on R associated to the following ODE $y'' + a^+y'$, (resp. $y'' + a^-y'$). Then after a change of time $t \mapsto t^*$, we have

$$p(x_\omega(t^*)) \in [\xi_\omega^-(t); \xi_\omega^+(t)]; \tag{1.2}$$

the asymptotic behaviour of $x_\omega(t)$ on M is compared to two diffusions on R driven by two ODE. The heat conduction invariant is determined by infinitesimal computations when the behaviour of the Green function is a question of global character.

Comparison lemma quickly became an important tool in Riemmanian Geometry. J. J. Prat (75) showed that a Riemanian manifold with strictly negative sectional curvature has a space of bounded harmonic functions which is of infinite dimension. It was also established comparison between the heat kernel of a general riemannian manifold with the heat kernel of constant curvature spaces ([De-Ga-Ma]). These comparison are stated in terms of radial cordinates in the correspondind exponential charts. In the case of symmetric space of rank $r > 1$ the radial coordinate have to be understood as a point in a cone of R^r. Then it is possible to describe the full asymtotic behaviour of Brownian motions by splitting the probability space into a skew product ([Ma-Ma74]).

§2. Ground state of vector bundle

We give a riemanian manifold M, a vector bundle F above M, an euclidean metric on each fiber F_m and a connection preserving this metric. Then on the vector space of sections $\Gamma(F)$ we have two quadratic forms defined respectively by the L^2 and the H^1 metric. We are interested in $\lambda_0(F)$ which is the infimum of the second quadratic form under the constraint that the first take the value 1. For instance if M is compact and if F together with its connection is trivial then we have $\lambda_0(F) = 0$.

The special case where $F = M \times C$ with a non trivial connection corresponds to the ground state of a Schrödinger operator associated to a magnetic field H; this magnetic field H is the curvature of the given connection, which plays the rôle of the potential vector.

When the curvature is large, the holonomy of the connection is large, and therefore *the ground state* $\lambda_0(F)$ *must be large.* It is difficult to obtain a quantitative statement of this last sentence. Estimation of this ground state is a basic problem in Geometry: it is associated to the deformation of minimal surface or to vanishing theorem needed for realizing the Harish-Chandra discrete series of representations of a semi-simple Lie group. Also a famous conjecture of Selberg on arithmetic group of $\mathrm{Sl}(2, R)$ can be reformulated through inequalities of the form $\lambda_0(F) \geq 1/2$.

Let us state more precisely the problem in the special case of a magnetic field on R^d. The potential vector is 1-differential form $\omega = \sum A_k \, d\xi^k$; the Hamiltonian has the following expression:

$$\mathcal{H} = -\frac{1}{2} \sum_{k=1}^{d} \left(\frac{\partial}{\partial \xi^k} + \sqrt{-1} A_k \right)^2 .$$

Denoting by $x(t)$ the Brownian motion on R^d, then the corrresponding semi-group has the following expression

$$(\exp(-t\mathcal{H})f)(\xi_0) = E_{\xi_0}(\exp(\sqrt{-1} J_t) f(x(t))), \qquad \text{where } J_t := \int_0^t \langle \omega, o\, dx \rangle \tag{2.1}$$

where the last integral is a Stratonovitch integral along the path of a differential form. Gaveau (77) give an exact expression for the ground state in the case of constant magnetic field.

The difficulty in (2.1) comes from the oscillating integral. I was lead in 1974 to the idea that this oscillating term could be estimated through some kind of stationary phase argument. I needed to compute for this reason the "gradient" ∇J_t and this was for me the first occasion to realize some kind of stochastic calculus of variations. The sharp following lower bound of this gradient was founded later by Prat (93):

$$\|\nabla J_t\|^2 \geq \int_0^t \left\| \int_0^s H \wedge o\, dx \right\|^2 ds - \frac{1}{t} \left\| \int_0^t ds \int_0^s H \wedge o\, dx \right\|^2 .$$

From this expression it can be shown that if $\|H\|(\xi) > c > 0$ then there exist $c' > 0$ such that

$$E\Big(\exp\Big(c' \frac{t}{\|\nabla J_t\|}\Big)\Big) < \infty. \tag{2.2}$$

This inequality is appealing for a stationary phase treatement: Taniguchi showed an infinite dimensional holomorphic Cauchy formula which implies an infinite dimensional stationary phase principle; in this context an hypothesis of type of (2.2) could imply an exponential decay of the stationary phase, UNDER the additional assumption of real analyticity for the potential vector.

Transverse analyticity concerns a similar problem; the time is now fixed at t_0 we look for the following exponential decay

$$\lim_{n \to \infty} \log \frac{|E(\exp(n\sqrt{-1}J_{t_0} \mid x(t) = 0)|}{|n|} < 0. \tag{2.3}$$

In 1985 at the autumn meeting of the Japanese Mathematical Society, I showed that (2.3) holds true for any C^3 differential form under the hypothesis that $\|H\|$ has pointwise a uniform lower bound; under the same hypothesis I obtain an estimate of the ground state [Ma85].

Important results related to this circle of idea can be found in [Ik-Ma], [Ma-Sh], [Ma-Ue], [Sh94], [Sh-Ta], [Ue].

§3. Transfer principle from ODE to SDE and Stochastic Calculus of Variations

A consequence of Itô theory of SDE is the possibility to use the probability space of the Brownian motion on R^d for realizing the diffusion process associated to any elliptic operator. The intrinsic construction of the *Itô map* is done in [Ma74'] as follows. Given a manifold M of dimension d and given q C^∞ vector fields $A_1, \dots, A_q$ on M, we associate to any C^1 curve ϕ on R^n the non autonomous vector field defined on M by $Z_t^\phi := \sum \dot{\phi}^k A_k$. Then the *controll map* is the flow of C^∞ diffeomorphisms defined on $M \times R$ by

$$\frac{d}{dt} U^\phi_{t \leftarrow t'} = Z_t; \quad U^\phi_{t' \leftarrow t'}(m) = m.$$

Given an R^q-valued Broxnian motion $x(t)$ we denote by x_n a sequence of piecewise C^1-curves converging towards x and properly constructed. Then

Limit Theorem 3.1. *Almost surely $U^{x_n}_{t\leftarrow 0}$ converges locally uniformly in (m,t) towards $U^{x}_{t\leftarrow 0}$ which is a C^∞-flow of diffeomorphisms on $M\times R$; furthermore the trajectory $t\mapsto U^{x}_{t\leftarrow 0}(m_0)$ is the generic path of the diffusion associated to the operator $\frac{1}{2}\sum\mathcal{L}^2_{A_k}$.*

A big part of classical differential geometry is based on the machinery of ordinary differential equations. As SDE can be reduced *in a canonic way* to a limit of a sequence of ODE, a new geometry appeared: the *Stochastic Differential Geometry.* Furthermore the limit theorem provides a *transfer principle* giving AUTOMATIC PROOF of some statements of Stochastic Differential Geometry in terms of the correponding statements of Ordinary Differential Geometry.

The International Symposium on SDE held in june 1976 at Kyoto was for me an exceptional possibility to present this transfer principle with two of its consequences for SDE: the *Stochastic Flow of Diffeomorphism* and the *Stochastic Calculus of Variations* along a Stochastic Flow.

The limit theorem is a difficult and key result and it is fortunate that there exits now several independent proofs: [Ma76], [Ik-Wa], [Kun], [St-Ta].

The limit theorem gives as byproduct existence theorem of solutions of SDE. But it gives more, it embbed canonically the theory of SDE, as a limiting case, into the theory of ODE.

§4. Mean value formulae for harmonic differential forms

Given a Riemannian manifold M of dimension d, we consider $\pi_t(m_0,dm)$ the fundamental solution of the heat equation $\partial_t-\Delta_0$, where Δ_0 denotes the Laplace-Beltrami operator of M. Then for any bounded harmonic function h the following mean value formula holds true:

$$h(m_0)=\int h(m)\pi_t(m_0,dm)=E_{m_0}(h(p(t))),\quad \forall t>0$$

where $p(*)$ denotes the Brownian motion on M: the diffusion process associated to Δ_0.

The De Rham Hodge operator of degree r is defined on exterior differential forms of degree r by the formula $\Delta_r:=-\frac{1}{2}(\delta d+d\delta)$ where $\delta=d^*$ is the adjoint of the coboudary operator d. Given an harmonic differential form ω, is it possible to show that ω satisfies some kind mean value formula? Is it possible to construct a probabilistic representation of the heat semi-group $\exp(t\Delta_r)$?

The notion of mean values assumes implicitely that it is possible to add, in an intrinsic way, values of the differential form ω at two different points $m, m' \in M$, which is a strange assertion for a differential geometer.

The *procedure of scalarization* is a way to avoid this difficulty. A *frame* on M, let be r, is by definition an euclidean isomorphism of R^d onto $T_m(M)$; we denote $m = \pi(r)$. The collection $O(M)$ of all frames of M has a natural structure of C^∞ manifold; as the orthogonal group $SO(d)$ operates on $O(M)$ we get that $O(M)$ is a principal bundle over M. To a differential form ω on M we can associate its component in a frame; in this way is defined a function $f_\omega : O(M) \mapsto R^d$; then $f_\omega(rog) = gf_\omega(r), \forall\, g \in SO(d)$. Furthermore the correspondance $\omega \mapsto f_\omega$ identifies the differential forms on M to the functions on $O(M)$ satisfying the previous equivariance property.

Given a frame r_0, we denote $\gamma(\tau)$ the geodesic of M tangent to the first coordinate vector of r_0. Then the Levi-Civita parallel transport of r_0 along $\gamma(*)$ defines a curve $t \mapsto \tilde{\gamma}(\tau)$ on $O(M)$. The tangent vector to $\tilde{\gamma}(*)$ at $t = 0$ defines a tangent vector at r_0 which will be denoted $A_1(r_0)$. We obtain in this way d canonical tangent vector fields A_k on $O(M)$, which will be called the canonic *horizontal vector fields.* Finally the *horizontal laplacian* is defined by

$$\Delta_{O(M)} := \frac{1}{2}\sum_{k=1}^{d} \mathcal{L}^2_{A_k}$$

and the following Weitzenböck formula realizes the scalarization procedure at the level of differential operators

$$f_{\Delta_1(\omega)} = \Delta_{O(M)} f_\omega - \frac{1}{2}\operatorname{Ricc}(f_\omega),$$

where $\operatorname{Ricc}(r)$ is the $d \times d$ matrix obtained by expressing the Ricci tensor in the frame r.

The Stratanovitch SDE

$$dr_x = \sum_k A_k o\, dx^k$$

gives a global canonic parametrization by the Brownian motion x of R^d of the diffusion associated to $\Delta_{O(M)}$. The *Itô map* $\mathcal{I}$ is defined by

$$\mathcal{I}(x)(\tau) := \pi(r_x(\tau)); \tag{4.1}$$

if $r_0 \in \pi^{-1}(m_0)$ is fixed then $\mathcal{I}$ realizes a 1 to 1 parametrization of the Brownian motion of M starting from m_0 by the brownian motion in R^d.

The Itô parallel displacement, presented by Itô at ICM 1962, can be constructed by the transfer principle from the Levi-Civita parallel displacement along smooth curve. Within the formalism of the frame bundle it has the following expression

$$t^p_{\tau\leftarrow 0} = r_x(\tau)o[r_x(0)]^{-1} \quad \text{where } p(*) = (\mathcal{I}(x))(*). \tag{4.2}$$

The intrinsic character of (4.1) makes possible to get the following asymptotic expansion of the holonomy of the Stochastic parallel transport:

Theorem ([Ma75], [Be]). *Denote by* $\exp_{m_0}$ *the riemannian normal chart on* M *at* m_0*, then when* $\tau \to 0$ *we have*

$$t^{p_x}_{0\leftarrow\tau}o\big(\exp_{m_0}\big)_* = \text{Identity} + R_{m_0}\left(\int_0^\tau x \wedge dx\right) + o(t^{3/2-\epsilon}) \tag{4.3}$$

where R_{m_0} *denote the curvature tensor of* M *at* m_0*.*

Theorem ([Ma74']). *The Heat semi-group on* 1*-differential form has the following scalarised expression*

$$f_{\exp(\tau\Delta_1)(\omega)}(r_0) = E_{r_0}(\mathcal{R}_{0\leftarrow\tau}(f_\omega(r_x(\tau)))) \tag{4.4}$$
$$\textit{where } \frac{d}{d\tau}\mathcal{R}_{0\leftarrow\tau} = \frac{-1}{2}\mathcal{R}_{0\leftarrow\tau}\,\mathrm{Ricc}(r_x(\tau)),$$

and where $\mathcal{R}_{0\leftarrow 0}$ = *Identity.*

The Atiyah-Singer Theorem gives an expression of the index of an elliptic operator above M where the function $\lambda/\sinh\lambda$ mysteriously appears. Bismut [Bi84] starting from (4.3) and (4.4) gives a probalistic proof of Atiyah-Singer Theorem where this mysterious function arrive quite naturaly as the characteristic function of the random variable $\tau^{-1}\int_0^\tau x_1\,dx_2 - x_2\,dx_1$, expression which is the $(1,2)$ component of $\int_0^\tau x \wedge dx$.

§5. Path Space as a parallelized manifold; Tangent Process

Begining 1993, I received an invitation to teach at the Spring Quarter 1994 at Kyoto University. In the perpective of this course I start with Professor Cruzeiro a systematic study of the differential geometry on $P_{m_0}(M)$, the space of paths on a riemannian manifold, starting from

$m_0 \in M$. After numerous discussions with ours Kyoto colleagues and subsequent work at the Institute Mittag-Leffler, we produced [Cr-Ma96].

The elements of the tangent space $T_{p_0}(P_{m_0}(M))$ at the point p_0 are identified with the continuous maps $Z : [0,1] \mapsto T(M)$ such that $Z_\tau \in T_{p_0(\tau)}(M)$. The choice of some regularity for the map $(\tau, p) \mapsto Z_\tau(p)$ is a main issue which will be discussed below. We denote X' the Banach space $P_0(R^d)$ of continuous paths on R^d and by X the Wiener space that is the same space enriched with its structure of filtered probability space. We denote

$$\text{(5.1)} \qquad \begin{aligned} &\Omega_p : T_p^0(P_{m_0}(M) \longmapsto X' \\ &\text{defined by } \Omega_p(Z) = z, \;\; z(\tau) := t^p_{0\leftarrow\tau}(Z_\tau). \end{aligned}$$

Then Ω can be looked upon as an X'-valued 1-differential form defined on $P_{m_0}(M)$ which $\forall\, p$ realizes an isomorphism of the tangent space at p onto a fixed Banach space: we say that Ω defines the *canonic parallelism of the Path Space.*

We have now the concept *constant vector field*, which mean a vector field Z defined on $P_{m_0}(M)$ such that there exists a fixed $z \in X'$ verifying

$$\text{(5.2)} \qquad \begin{aligned} &\langle \Omega_p, Z(p) \rangle = z, \quad \forall\, p \in P_{m_0}(M); \\ &\text{or equivalently } Z_\tau(p) = t^p_{\tau\leftarrow 0}(z_\tau). \end{aligned}$$

A similar situation in a finite dimension setting is a Lie group parallelized throug its Maurer-Cartan differential form; then the constant vector fields correspond to left invariant vector fields; the bracket of left invariant vector fields defines the *structural constants* of the asssociated Lie algebra.

We denote by $\mathcal{T}$ the vector space of smooth cylindrical functions on $P_{m_0}(M)$, that is such that there exits a finite set $\{\tau_i\}_{1\le i\le j}$ and a smooth function $F : M^j \mapsto R$ for which $f(p) = F(\ldots, p(\tau_i), \ldots)$; we define

$$(D_Z(f))(p) := \sum_{i=1}^{i=j} \langle d_i F, Z_{\tau_i}(p) \rangle .$$

Now for Z constant vector field the fact $f \in \mathcal{T}$ do not imply that $D_Z f \in \mathcal{I}$; this technical difficulty is eliminated by extending by closure the domain of D_Z as it will be seen in 5.8. It is then possible to define the bracket $Z^3 = [Z^1, Z^2]$ of two vector fields Z^1, Z^2 by the relation

$$D_{Z^3} f = (D_{Z^1} D_{Z^2} - D_{Z^2} D_{Z^1}) f.$$

The *structural equation of the Path Space* is defined by

$$[z^1, z^2]_p = \langle \Omega_p, [Z^1, Z^2] \rangle \tag{5.2}$$

where Z^s, $(s = 1, 2)$ are the constant vector fields associated to z^s. Then $\forall\, p_0 \in P_{m_0}(M)$ the structural equation defines a bilinear antisymmetric map $[*, *]_{p_0}$ of $X' \times X' \mapsto X'$.

Theorem ([Cruzeiro-Ma96]).

$$[z^1, z^2]_p(\tau) = Q_{z^1}(\tau) z^2_\tau - Q_{z^2}(\tau) z^1_\tau \tag{5.3}$$
$$\textit{where } Q_z(\tau) = \int_0^\tau R_\lambda(z_\lambda, o\, dx(\lambda))$$

where $p = \mathcal{I}(x)$ and where R_λ denotes the curvature tensor of M expressed in the frame $r_x(\lambda)$.

Remark. The curvature tensor depends upon four indices, the above integral saturates two indices, therefore $Q_{z^1}(\tau)$ is an antisymmetric matrix, which by operating on the vector z^2_τ gives a vector.

Corollary. *Denote $d_\tau z$ the differential in τ of z, then the Stratanovitch differential of the structural eqation is*:

$$d_\tau([z^1, z^2]) = R(z^1, z^2)(o\, dx) + Q_{z^1}(d_\tau z^2) - Q_{z^2}(d_\tau z^1). \tag{5.4}$$

Remark. For $z^1, z^2 \in H^1$ then $[z^1, z^2] \in H^1$ if and only if $R_*(z^1, z^2) = 0$.

We call a *tangent process on the Wiener space X* ([Cr-Ma96]) the data of a semimartingale ζ such that its Itô differential satisfies

$$d_\tau \zeta^j = a^j_i\, dx^i + c^j\, d\tau \quad \text{where } a^j_i + a^i_j = 0. \tag{5.5}$$

Example. For $z^1, z^2 \in H^1$ it results from (5.4) that $[z^1, z^2]$ is a tangent process on X.

Theorem ([Bi84'], [Dr92], [Fa-Ma93], [En-St95], [Cr-Ma96]). *The differerential operator D_ζ associated to a tangent process satisfies the following formula of integration by part*:

$$E(D_\zeta f) = E\left(f \int_0^1 c^j\, dx^j\right) \tag{5.6}$$

where the last integral is an Itô Stochastic Integral.

We call Z^* a *Tangent Process on the Path Space* if $\zeta^* := \langle \Omega, Z^* \rangle$ is a Tangent Process on the Wiener space.

Main Theorem. *Define the differential of the Itô map by*

$$\mathcal{I}'(x).\zeta := t^p_{0\leftarrow *}\left(\frac{d}{d\epsilon}_{|\epsilon=0}\mathcal{I}(x+\epsilon\zeta)\right) \quad \textit{where } p=\mathcal{I}(x);$$

then $\mathcal{I}'$ realizes an isomorphism of the space of tangent process on X onto the space of tangent process on $P_{m_0(M)}$; this isomorphism can be computed through the system of Stratonovitch SDE

$$d\zeta = d\zeta^* + \rho o\, dx; \quad d\rho = -R(o\, dx, \zeta^*) \tag{5.7}$$

all the initial conditions at $\tau = 0$ being equal to 0.

The combinaison of (5.7) and of (5.6) gives the

Corollary. *Tangent Processes on $P_{m_0}(M)$ satisfy formulae of integration by part; therefore the derivation operator along a Tangent Process defined on smooth cylindrical functions is closable.*

In order to explicit this formula we write the first equation of (5.7) in Itô form in the case where $\zeta^* = z \in H^1$, then

$$d\zeta = \left(\dot{z} + \frac{1}{2}\operatorname{Ricc}(z)\right)d\tau + \rho\, dx.$$

As the application of Itô preserves the probability measure the formula of integration by part can be pull back to X: introduce $\Phi := fo\mathcal{I}$; Φ is defined on X and

$$E(D_Z f) = E(D_\zeta \Phi) = E\left(\Phi \int_0^1 \left(\dot{z} + \frac{1}{2}\operatorname{Ricc}(z)\right)dx\right). \tag{5.8}$$

The origin of the introduction of tangent process has been the necessity to handle properly the structural equation. It is therefore interesting that this procedure stabilizes itself and that it not necessary to introduce new objects to compute braket of tangent processes:

Theorem 5.9 ([Cr-Ma96], [Dr99]). *Given two tangent process on $P_{m_0}(M)$, with differentiable coefficients, then their bracket is a tangent process.*

§6. Intertwinning formulae; Integration by part formulae

Intertwining at the level of PDE.

On the Riemannian manifold M we have the following classical intertwining formulae for Laplace-DeRham-Hodge operators:

$$d(d\delta + \delta d) = (d\delta + \delta d)d \quad \text{or} \quad d\Delta_r = \Delta_{r+1}d.$$

We write this intertwinning formula for $r = 0$ and we exponentiate we get

$$d\exp(\tau\Delta_0) = \exp(\tau\Delta_1)d. \tag{6.1}$$

Consider the final value problem for backward heat equation:

$$\Big(\frac{\partial}{\partial\tau} + \Delta_0\Big)\phi_\tau = 0, \quad \phi_1 = F \text{ given}$$

then this equation has for solution

$$\phi_\tau(m) = E_{p(\tau)=m}(F(p(1)) = (\exp(1-\tau)\Delta_0)F)(m).$$

In Mathematical Finance F will be the liquidative value at maturity $\tau = 1$ of a portfolio. Denote $\pi_\tau = d(\phi_\tau)$, then (6.1) and (4.4) give

$$\pi_\tau = E_{p(\tau)=m}(\tilde{t}^p_{\tau\leftarrow 1}(dF)_{p(1)})$$

where $\tilde{t}$ is the *damped parallel transport* [Fang-Ma93] defined by

$$\begin{aligned} \tilde{t}^p_{\tau\leftarrow\tau'} &= Q(\tau,\tau')t^p_{\tau\leftarrow\tau'} \\ &\text{where } d_\tau Q(\tau,\tau') = \frac{1}{2}(\mathrm{Ricc})Q(\tau,\tau')\,d\tau, \;\; Q(\tau',\tau') = I. \end{aligned} \tag{6.2}$$

The word "damped" has the following origin: if we suppose that Ricc $=$ $2I$ then $Q(\tau,1) = e^{\tau-1}$.

The function $\phi(\tau, m)$ is a parabolic function, therefore its variation in the space time is given by the following Itô stochastic integral

$$\begin{aligned} F(p(1) - \phi_\tau(p(\tau))) &= \int_\tau^1 \langle \pi_\lambda, dx(\lambda)\rangle \\ &= \int_\tau^1 E^{\mathcal{N}_\lambda}(\tilde{t}^p_{\lambda\leftarrow 1}(dF)_{p(1)})\,dx(\lambda), \end{aligned} \tag{6.3}$$

where $\mathcal{N}_t$ denotes the filtration on X ([Ai-Ma95]).

Given a smooth cylindrical function $f(p) := F(\ldots, p(\tau_i), \ldots)$ we defined its *damped gradient* $\tilde{D}_\tau f$, $\tau \in [0,1]$, by the formula

$$\tilde{D}_\tau f = \sum_{1\le i\le j} 1(\tau < \tau_i)\tilde{t}_{\tau\leftarrow\tau_i}(d_i F)$$

We get the *Clark-Bismut-Ocone formula*

$$f - E(f) = \int_0^1 E^{\mathcal{N}_\tau}(\tilde{D}_\lambda f)\, dx(\lambda) \tag{6.4}$$

firstly by specializing (6.3) to the case $\tau = 0$ which give the wanted formula for cylindrical functions of the form $f(p) = F(p(1))$ and finally by generalizing to arbitrarly cylindrical functions through Markov property applied at the sequence of times τ_i.

Given a constant vector field Z on $P_{m_0}(M)$ which has for image in the parallelism $z \in H^1$ we denote

$$\tilde{D}_Z f := \int_0^1 \tilde{D}_\tau f \dot{z}(\tau)\, d\tau;$$

using the energy identity for Itô stochastic integral we get

$$E(\tilde{D}_Z f) = E\Big(\Big(\int_0^1 \dot{z}\, dx\Big)\Big(\int_0^1 E^{\mathcal{N}_\tau}(\tilde{D}_\tau f)\, dx(\tau)\Big)\Big);$$

using Clark-Bismut-Ocone formula we get finally the following *formula of integration by part*

$$= E\Big(f \int \dot{z}\, dx\Big). \tag{6.5}$$

It can be remarked the proof of (6.5) given here depends only upon Itô formula and re-presentation of the heat semi-group on 1-differential forms. As it is possible by elementary computations to deduce (5.8) from (6.5), we could have used this approach to give an elementary proof of (5.9); nevertheless the use of Geometry on $P_{m_0}(M)$ is essential for the following intertwinning.

Intertwining by transference at the level of Path space.

Denote $\pi_\tau(m_0, dm)$ the fundamental solution of the heat with pole at m_0. We call *transference* a formula which replace a derivation relatively to the starting point m_0 by a derivation relatively to the ending point m:

Transference Theorem ([Cruzeiro-Ma98]). *Given a vector $a \in T_{m_0}(M)$, consider its damped parallel transport along generic trajectory of the Brownian motion of M starting from m_0, then for every $F \in C_b^1(M)$ we have*

$$\frac{d}{d\epsilon}_{|\epsilon=0} \int_M \pi_\tau(m_0 + \epsilon a, dm) F(m) = E_{m_0}(\langle \tilde{t}^p_{\tau\leftarrow 0}(a), (dF)_{p(\tau)}\rangle). \tag{6.6}$$

The proof is based on the machinery of tangent process. It is possible to deduce from this transference theorem an intertwinning result for first order PDE.

Corollary. *Define a vector field Y on M by the formula*

$$Y_m := E^{p(\tau)=m}_{p(0)=m_0}(\tilde{t}^p_{\tau\leftarrow 0}(a))$$

then

$$(\partial_a(\exp(\tau\Delta_0)F))(m_0) = (\exp(\tau\Delta_0)(\partial_Y F))(m_0). \tag{6.7}$$

Remark. This formula differs from (6.1) in the sense that the tangent vector a is arbitrarly choosen when from the other hand the point m_0 where the intertwinnig is computed is fixed.

Theorem (Bismut-Harnack identity) ([Bi84'], [En-St], [El-Li], [El-Le-Li], [Fa-Ma]).

$$\begin{aligned}\frac{d}{d\epsilon}_{|\epsilon=0}\int_M \pi_\lambda(m_0+\epsilon a, dm)F(m) \\ = E_{m_0}\left(F(p(\lambda))\frac{1}{\lambda}\int_0^\lambda \langle \tilde{t}_{\tau\leftarrow 0}(a), dx(\tau)\rangle\right)\end{aligned} \tag{6.8}$$

where the stochastic integral is an Itô stochastic integral.

We can deduce a proof by applying to (6.6) Girsanov type theorem.

§7. Construction of measures in infinite dimension through infinitesimal geometry

Our point of view is to define a probability measure μ as the invariant mesure of a second order elliptic operator. This point of view illustrate the paradigm of *going from the infinitesimal to the global* as it produces a global object, the measure μ, from infinitesimal constructions. Our methodology will be presented in the framework of finite dimensional manifold but with *dimension free constants.*

We consider an elliptic operator $\tilde{\Delta}'$ on a manifold M such that $\Delta'(1) = 0$. Then the second order terms define a riemannian metric and there exits a unique vector field Z on M such that

$$\Delta' f = \frac{1}{2}\Delta_0 f - \langle Z, df\rangle.$$

The scalarization procedure associates to Z an R^d-valued function f_Z defined on $O(M)$. We denote $\partial^h_* f_Z$ the $(d \times d)$ matrix which has for k^{th} column $\partial_{A_k} f_Z$ where $\ldots, A_k, \ldots$ are the canonic horizontal vector fields. We define a $(d \times d)$ matrix valued function Ricc' on $O(M)$ by

$$\mathrm{Ricc}' = \mathrm{Ricc} + 2\partial_* f_Z.$$

Lemma 7.1. *Replacing* Ricc *by* Ricc′ *then the heat kernel* π'_τ *associated to* $\exp(\tau\Delta')$ *satisfies* (6.6).

The proof follows the same line as in Section 6 and the Stochastic Calculus of Variations along the diffusion generated by Δ' brings into the computation the matrix $\partial^h_* f_Z$.

Theorem 7.2 ([Cr-Ma98]). *Assume that* $\exists\, \delta > 0$ *such that*

$$\mathrm{Ricc}' + (\mathrm{Ricc}')^* \geq 2\delta \times \mathrm{Identity}$$

then Δ' *has a unique invariant probability measure* μ.

The proof is based on the extension of the intertwining formula (6.7) to derivative of second order, extension which is obtained by differentiating the identity (6.6) and computing its r.h.s. by the Stochastic Calculus of Variations. The derivation of the semi group $\frac{d}{d\tau}\exp(\tau\Delta') = \Delta' \exp(\tau\Delta')$ gives then

$$\frac{d}{d\tau}\int \pi'_\tau(m_0, dm) f(m) = \Delta'_{m_0} \int_M \pi_\tau(m_0, dm) f(m).$$

We transfert the first (by (6.7)) and second order derivatives appearing in Δ'_{m_0} from the starting point to the end point: we get a second order elliptic operator $\mathcal{Q}_\tau$, called the *desintegrated adjoint operator* such that

$$\frac{d}{d\tau}\int_M \pi'_\tau(m_0, dm) f(m) = \int_M \pi'_\tau(m_0, dm)(\mathcal{Q}_\tau f)(m) \tag{7.3}$$

where the norm of $\mathcal{Q}_\tau$ relatively to the underlying riemannian metric of M satisfies

$$\|\mathcal{Q}_\tau\| \leq c \exp(-\delta\tau). \tag{7.4}$$

We call the *desintegrated adjoint process* the process $\hat{p}(\tau)$ associted to $\exp(\tau\mathcal{Q}_\tau)$ and starting at $\tau = 0$ from m_0. Then the law at time τ of the desintegrated adjoint diffusion is equal to $\pi'_\tau(m_0, *)$; furthermore (7.4) implies that

$$\hat{p}(\infty) := \lim_{\tau\to\infty} \hat{p}(\tau)$$

exists.

We construct μ as the law of $\hat{p}(\infty)$.

Corollary. *Assume furthermore that* Ricc$'$ *is symmetric, then the measure μ is reversible and the following formula of integration by part holds true:*

$$\int_M \partial_V \Phi \, d\mu = \int_M \Phi(2(Z|V) - \mathrm{trace}(\partial_* f_V)) \, d\mu. \tag{7.5}$$

§8. Vector fields on Probability spaces, their divergence and their flow

An *admissible vector field* on a probability space is a derivation operator defined on an algebra of cylindrical functions and satisfying a formula of integration by part part relatively to the probability measure μ:

$$E(D_Z\phi) = E(\phi\delta_\mu(Z));$$

the function $\delta_\mu(Z)$ is called the *divergence of Z relatively to μ*.

On the Wiener space X of the Brownian motion on R^d, classically is considered the Cameron-Martin space

$$H^1 = \Big\{ h \in X \; ; \; \int \|\dot{h}\|^2_{R^d} \, d\tau := \|h\|^2_{H^1} < \infty \Big\}.$$

Given $h \in H^1$, we have the following Cameron-Martin integration by part formula:

$$E(D_h\phi) = E\Big(\phi \int_0^1 \dot{h} \, dx\Big)$$

where the last integral is an Itô Wiener Stochastic Integral.

The existence of a formula of integration by part implies that the derivation operator D_h is closable in L^r, $r \in]1, \infty[$. Then the intersection $\forall\, h \in H^1$ of the domains these closures defines a Banach space, the *Gross-Stroock Sobolev space* $D_1^r(X)$. In the same way the Gross-Stroock Sobolev space $D_s^r(X)$ of s-times differentiable functions can be defined. Given an abstract Hilbert space $\mathcal{H}$, the Gross-Stroock Sobolev space $D_1^r(X; \mathcal{H})$ of $\mathcal{H}$-valued functions is similarly defined. Then

Theorem (Watanabe 84). *For all $r > 1$, $\exists\, c_r < \infty$ such that*

$$\|\delta_\mu(Z)\|_{L^r} \leq c_r \|Z\|_{D_1^r(X; H^1)} \tag{8.1}$$

and the finiteness of the r.h.s. implies the existence of the divergence appearing in the l.h.s.

We have then the following infinite dimensional analog of the finite dimensional fact that smooth vector field generates a flow:

Theorem (Cruzeiro 83). *Given $Z \in D_1^r(X; H^1)$, $\forall\, r < \infty$ such that $\exists\, c > 0$ for which*

$$E(\exp c|\delta(Z)|) < \infty, \quad E(\exp(c\|Z\|_{H^1}) < \infty, \qquad (8.2)$$
$$E(\exp(c\|DZ\|_{H^1 \otimes H^1}) < \infty,$$

then, $\forall\, \tau \in R$, there exists a map $U_\tau : X \mapsto X$, defined μ-a.e., preserving the class of the Wiener measure, satisfying the group property $U_\tau o U_\sigma = U_{\tau+\sigma}$ such that denoting $\mu_\tau := (U_\tau)_* \mu$ the image of the Wiener measure, we have

$$\frac{d\mu_\tau}{d\mu}(x_0) = \exp\Big(\int_0^\tau (\delta Z)(U_{-\lambda}(x_0))\, d\lambda\Big); \qquad (8.3)$$
$$\frac{d}{d\tau}\int f\, d\mu_\tau = \int f\delta(Z)\, d\mu_\tau.$$

In the case of a constant vector field we get back a Theorem of Cameron-Martin. Several papers amplify this theorem see [Bo-Mw].

We have the following extension to the Riemannian Path Space.

Theorem 8.4 (Driver 92, Hsu 95). *Fix $z \in H^1(R^d)$ then define a constant vector field Z on $P_{m_0}(M)$ by $Z(p) = \Omega_p^{-1}(z)$ then the flow associated to Z exists.*

This theorem cannot be reduced by the Itô map $\mathcal{I}$ to Cruzeiro Theorem: in fact the inverse image of a constant vector field is a tangent process and no more an H^1 vector field. The original proof of Driver is a "tour de force": at the same time he constructed the flow and produced a new approach to the Bismut formula of integration by part (5.6). From the Bismut book (1984) to the Driver paper (1992) the study of the Riemannian Path pace stayed essentially quiet. After the Driver breakthrough, the subject became very active.

§9. Geometrization of the Anticipative Stochastic Calculus by Divergences

On a finite dimensional Riemannian manifold the *energy identity for*

1-*differential form* ω has the following expression

$$\|d\omega\|^2_{L^2_\mu} + \|\delta_\mu(\omega)\|^2_{L^2_\mu} = \int_M \|\nabla\omega\|^2 \, d\mu + \int_M (\mathrm{Ricc}''(\omega)|\omega) \, d\mu \tag{9.1}$$

where $d\mu = \exp(-\phi)\, dm$, dm being the riemannian volume measure, where δ_μ is the *divergence* that is the ajoint in L^2_μ of d which means:

$$\int_M (\omega|df) \, d\mu = \int_M f\delta_\mu(\omega) \, d\mu,$$

where ∇ denotes the Levi-Civita connection on M and where $\mathrm{Ricc}'' := \mathrm{Ricc} + \nabla^2\phi$.

Take for instance $M = R^d$, $\phi = \frac{1}{2}\|\xi\|^2$, then μ is the canonic gaussian measure on R^d and $\mathrm{Ricci}'' = \mathrm{Identity}$.

Theorem 9.2 (Shigekawa 86). *On the Wiener space X the identity* (9.1) *holds true with* $\mathrm{Ricc}'' =$ *Identity.*

A far reaching discovery of Gaveau-Trauber is that, on the Wiener space, $\delta_\mu(\omega)$ is equal to the Skorokhod anticipative integral; their proof is based on an L^2 chaos expansion and therefore is of **global** nature. Zakai-Nualart-Pardoux ([Nu-Za] and [Nu-Pa]) define Anticipative Integral as the divergence and produce a constructive scheme of approximation by a sequence of finite sums producing in this way a **local** construction of the Anticipative Integral. As a differential form ω is an H^1 valued functional we introduce $a := d_\tau\omega$ and then a takes its value in $L^2([0,1])$; we assume that $a \in D^2_1(X; L^2([0,1]))$. Then Zakai-Nualart-Pardoux prove the following energy identity

$$\begin{aligned} E\Big(\Big(\int_0^1 a_\tau \, dx(\tau)\Big)^2\Big) \\ = E\Big(\int_0^1 |a_\tau|^2 \, d\tau + \int_0^1 \int_0^1 D_\tau a_\lambda D_\lambda a_\tau \, d\lambda \, d\tau\Big). \end{aligned} \tag{9.3}$$

For a is adapted, $D_\tau a_\lambda = 0$ for $\tau > \lambda$; the second term of the r.h.s. vanishes and the Itô energy identity appears.

A very short proof of (9.3) can be obtained by a direct application of (9.1), and (9.2).

Using Schwarz inequality we deduce from (9.3) the inequality

$$E\Big(\Big(\int_0^1 a_\tau \, dx(\tau)\Big)^2\Big) \le E\Big(\int_0^1 |a_\tau|^2 \, d\tau + \int_0^1 \int_0^1 |D_\tau a_\lambda|^2 \, d\tau \, d\lambda\Big). \tag{9.4}$$

On a riemanian manifold for a constant vector field, the formula (6.5) proves that the Itô type Stochastic Integral is the transposed of the **damped** gradient $\tilde{D}$. It could be natural to define the anticipative integral of a non adapted process as the operator $(\tilde{D})^*$. We shall not develop this point of view and will look for majorations of $\delta := D^*$. As the passage from the gradient to the damped gradient is an operator which is uniformly bounded with its inverse, a majoration for D^* will imply majoration of the stochastic integral.

Manageable generalization to $P_{m_0}(M)$ of (9.1) seems presently not available. The Levi-Civita connection appearing in (9.1) can be explicitely computed on $P_{m_0}(M)$ but leads quickly to exploding formulas ([Cr-Ma96]). These two unfortunate facts make difficult the approach to the following inequality (9.6).

We define a bounded inversible map $H^1 \mapsto H^1$ let $z \mapsto \hat{z}$ through the relation

$$\dot{\hat{z}} = \dot{z} + \frac{1}{2}\,\mathrm{Ricc}(z)$$

For constant vector fields z, the divergence is given by the following Itô Stochastic integral:

$$\delta(z) = \int_0^1 \dot{\hat{z}}\, dx$$

A connection ∇_u on $P_{m_0}(M)$ will be given by its action on constant vector fields $\nabla_u z$, defined by the following formula

$$d_\tau \widehat{\nabla_u z} := \Gamma_u(\tau)\dot{\hat{z}}(\tau) \qquad \text{where } \Gamma_u(\tau) := \int_0^\tau R_\lambda(u(\lambda), o\, dx(\lambda)).$$

A *new riemannian metric* on $P_{m_0}(M)$ is introduced

$$\|z\|^2 = \int_0^1 |\dot{\hat{z}}|^2 d\tau,$$

and for this metric we have

Theorem (Cruzeiro-Fang 97).

$$\|\delta(z)\|^2 + \|\tilde{d}z\|^2 = \|\nabla z\|^2 + \|z\|^2 \tag{9.6}$$

where $\tilde{d}$ is the antisymmetrized covariant derivative.

This formula is a perfect analog of (9.1). It implies L^2 inequality analog to (9.4). Later Fang 98 has obtained the analogous of (9.4) for L^r-norms.

§10. Geometric measure theory. Principle of descent

I was invited to a Taniguchi Conference at Katata, on lake Biwa, in 1982. After the push that the Stochastic Calculus of Variations received from its presentation at the International Symposium on SDE, RIMS 1976, some participants of this 1982 Conference kindly discussed some analytical aspects of this subject. I present an *implicit function theorem* valid for $D^\infty(X) := \bigcap_{r,s<\infty} D^r_s(X)$ functions.

In the finite dimensional case, we have the classical Sobolev embedding $D^r_s(R^d) \subset C(R^d)$ the space of continuous funtions. In the case of the Wiener space $D^\infty(X)$ is not contained in $C(X)$ and this was the difficulty to overcome to build a local inversion needed to construct the implicit function. This isolated result develop into new subject: the *quasi-sure analysis.*

The dual space of $D^\infty(X)$ can be written as

$$D^{-\infty}(X) = \bigcup_{r>1,\, s<\infty} D^r_{-2s}(X),$$

$D^r_{-2s}(X)$ being characterized as the image by $(-\Delta_X + 1)^s$ of the space $L^r(X)$, where Δ_X denotes the Ornstein-Uhlenbeck elliptic operator on X.

Theorem 10.1 (Sugita 88). *Given $\Psi \in D^{-\infty}(X)$ we say that $\Psi > 0$ if $\Psi(f) \geq 0$ for all $f \geq 0$, $f \in D^\infty(X)$. Then given $\Psi > 0$ there exists on X a positive Radon measure ν_Ψ of finite total mass such that*

$$\Psi(u) = \int_X u(x)\nu_\Psi(dx)$$

for all u smooth cylindrical functions.

We says that a borelian $A \subset X$ is *slim* if $\forall\, \Psi > 0$, $\Psi \in D^{-\infty}(X)$, $\nu_\Psi(A) = 0$.

Redefinition Theorem 10.2 (in the version of K. Itô 90). *Given $f \in D^\infty(X)$ we can construct a borelian function f^*, defined everywhere and almost everywhere equal to f, and a decreasing sequence of open subsets $O_n \subset O_{n-1} \subset \cdots X$ such that $\bigcap O_n$ is slim and such that the restriction of f^* to O_n^c the complement of O_n is continuous. Two such functions f^* are equal ouside a slim set; identifying two functions which differ on a slim set we call f^* the redefinition of f.*

Given a map $g \in D^\infty(X; R^d)$, we defined its *covariance matrix* $\mathcal{C}(g)$ as the $(d \times d)$ matrix $(\nabla g^k | \nabla g^l)_{H^1}$. We say [Ma76] that

$$(10.3) \qquad g \text{ is } \textit{non degenerated} \text{ if } \det(\mathcal{C}^{-1}) \in L^r, \quad \forall\, r < \infty.$$

In finite dimension the hypothesis $\mathcal{C}(x_0)$ invertible implies the existence of an implicit in a neighberhood of x_0; therefore (10.3) appears as global quantitative version of this finite dimensional hypothesis (see Lescot [Le] for a qualitative version).

In abstract theory of conditioning, the conditional expectation or the conditional law can be defined only **almost surely**. The following coarea theorem provides conditional laws defined for all resonnable conditioning.

Coarea Theorem (Airault-Ma 88). *Under the hypothesis* (10.3), *the law of the random variable g has a C^∞ density $k(\xi)$ relatively to the Lebesgue measure $d\xi$ of R^d. Denoting Q the interior of the support of k, there exists a continuous map $Q \mapsto D^{-\infty}$, which associates to $\xi \in Q$ a probability measure ν_ξ such that for all $f \in D^\infty(X)$*

$$(10.4) \qquad E(f) = \int_Q k(\xi)\, d\xi \left[\int_{(g^*)^{-1}(\xi)} f^*(x) \nu_\xi(dx) \right].$$

Fixing $\xi \in Q$ it is possible to write $(g^*)^{-1}(\xi) = \bigcup V_n$ where V_n is an increasing sequence of manifolds; it is possible to define an gaussian area measure λ on V_n such that

$$(10.5) \qquad \nu_\xi(dx) = \frac{1}{k(\xi)} \frac{1}{\sqrt{\det(\mathcal{C})}} \lambda(dx).$$

10.6. Principle of Descent.

A property true outside a slim set is said to be true *quasi-surely.* For instance given a sequence of functions f_n converging in $D^\infty(X)$ then theirs redefinitions f_n^* converge pointwise quasi-surely.

The principle of descent states that a property true quasi-surely remains true almost surely under the conditioning by $g^*(x) = \xi_0$, $\xi_0 \in Q$. Its proof results from the fact that $\nu_{\xi_0} \in D^{-\infty}$ and therefore do not charge slim sets.

Kusuoka has developed a *quasi-sure theory of differential forms* on the Wiener space.

§11. From Path Group to Loop Group

The reading of Pressley-Segal book "Loop Groups" engaged me in 1986 to try to work on probabilistic questions linked to this framework. The first obvious question was to find *quasi-invariant measures.* The French-Japanese joint Seminar run in Paris in the spring 1987 in the context of the centenary of the birth of Paul Lévy, push us to this work which finally appeared in [Ma-Ma90].

We denote by G a compact simply connected Lie group; we choose on its Lie algebra $\mathcal{G}$ an euclidean metric which is invariant under the adjoint action; for simplicity in the following we shall restrict ourselfs to matrix group, as for instance $SU(2)$; picking an orthonormal basis e_k of $\mathcal{G}$, we consider the matrix Stratonovitch SDE:

$$dg_x = g_x \sum_k e_k o\, dx^k \tag{11.1}$$

where $x \in X$ the Wiener space of the Brownian motion on R^d. The infinitesimal generator associated to this process is $\Delta^r := \frac{1}{2}\sum_k (\partial^r)_k^2$ where $\partial_k^r \phi(g) = \frac{d}{d\epsilon_0}\phi(g\exp(\epsilon e_k))$. We can exchange in all the previous constructions right action by left action, consider left derivatives ∂_k^l and the left operator $\Delta^l := \frac{1}{2}\sum_k(\partial_k^l)^2$. The fact that the adjoint action of G acts on $\mathcal{G}$ by orthogonal transformations implies that $\Delta^r = \Delta^l$. The metric on $\mathcal{G}$ induces a riemannian metric on G; denoting Δ_0 its Laplace-Beltrami operator, we have also $\Delta_0 = \Delta^r$.

We denote $P_e(G)$ the space of paths on G starting from the unit element e. This is a special case of a Riemannian Path Space to which we could apply the results of Section 5. But the Stochastic Calculus of Variations depends upon the type of variations we use along a path of the diffusion. We shall use variations fitted to the SDE (11.1): writting

$$\frac{d}{d\epsilon}_{|\epsilon=0} g_{x+\epsilon h}(\tau) = u(\tau) g_x(\tau)$$

and differentiating (11.1) we get

$$\dot{u} g_x = g_x \sum e_k \dot{h}^k \quad \text{or} \quad \dot{u}(\tau) = \mathrm{Ad}(g_x^{-1}(\tau))(\dot{h}(\tau)). \tag{11.2}$$

As Ad is an orthogonal transformation the map $h \mapsto k$ is a unitary automorphism of H^1: the situation is much more simpler than in (5.7) where we need to introduce tangent processes in order to find an isomorphism. Given $u \in P_0(\mathcal{G})$ and given a cylindrical function $f(x) =$

$F(\dots, g_x(\tau_i), \dots)$ we define the derivative

$$(D_u f)(x) := \frac{d}{d\epsilon}_{|\epsilon=0} F(\dots, \exp(\epsilon u(\tau_i)) g_x(\tau_i), \dots).$$

Then the relation (11.2) implies the following formula of integration by part:

$$(11.3) \qquad E(D_u f) = E(q(x) f(g_x))$$
$$\text{where } q(x) := \int_0^1 \langle \mathrm{Ad}(g_x(\tau)) u_x(\tau), dx(\tau) \rangle$$

We denote by $L_e(G)$ the *loop group* constituted by the paths satisfying $p(0) = p(1) = e$. We take on $L_e(G)$ the measure of the Brownian bridge $\mu_{L_e(G)}$. We want to get on $L_e(g)$ a formula of integration by part on $L_e(G)$ for admissible variations constituted by $u \in H^1$ such that $u(0) = u(1) = 0$. The two approachs that we consider below are valid in the more general context of $L_{m_0}(M)$, the Loop Space of a Riemannian manifold M.

The first approach [Ma-Ma90] is based on *quasi-sure analysis.* We denote by $\Phi(x) := g_x(1)$; then $\Phi \in D^\infty(X; G)$. Fixing x the map $h \mapsto u$ defined in (11.2) is unitary: the matrix $\mathcal{C}$ = Identity; therefore Φ is a non degenerated map.

The function q defined in (11.3) satisfies $q \in D^\infty(X)$. We denote by q^* its redefinition, by the descent principle q^* is defined $\mu_{L_e(G)}$ almost everywhere. As $D_u \Phi = 0$, we have $\forall\, \theta \in C^\infty(G)$

$$E((\theta o \Phi) D_u f) = E(D_u(\theta(\Phi) f)) = E((\theta o \Phi) q f);$$

we apply the coarea formula along Φ to the first and the third term of the above identity; we obtain an identity valid for all the conditionning $(\Phi^*)^{-1}(g_0)$. Taking $g_0 = e$ we obtain the localization of (11.3) to $L_e(G)$.

The second approach is based on the *Doob h-theory.* We denote $\pi_\tau(g)\, dg$ the law of $g_x(\tau)$. Then the measure $\mu_{L_e(G)}$ is parametrized by the following Stratonovitch SDE, driven by a new Brownian motion y:

$$dl_y(\tau) = l_y(\tau) \left(\sum_k e_k o\, dy^k + \nabla^r \log(\pi_{1-\tau}(l_y(\tau)))\, d\tau \right).$$

The Stochastic Calculus of Variations associated to this new parametrization is computed by considering $y \mapsto y + \epsilon h$. This computation involve the second derivative:

$$\nabla^l \nabla^r \log \pi_\lambda, \text{ where } \nabla^l \text{ denotes the left gradient.}$$

For bounding these derivatives the following thorem is usefull

Theorem (Stroock-Ma 96). *Given a Riemannian manifold M, fix $m_0 \in M$ and consider $\pi_\lambda(m_0, m)\, dm$ the fundamental solution of the heat. The Levi-Civita covariant derivative ∇ defines a canonic Hessian ∇^2. Then*

$$\lim_{\lambda \to 0} \lambda \nabla^2 \log \pi_\lambda = -\frac{1}{2} \nabla^2 d^2$$

where d denotes the Riemannian distance to m_0, this limit being uniform $m \in K$ where K is a compact which do not intersect the cutt locus of m_0.

Some others theorems linked to differential analysis on Loop group

Ergodicity Theorem (Gross 93). *Assume G simply connected. A function $f \in D_1^r(L_e(G))$ satisfying $D_u f = 0$, $\forall u \in H^1$, such that $u(0) = u(1) = 0$ is constant.*

Research of Logarihmic Sobolev inequalities on Loop spaces was started by Gross and after has been very much discussed; this important topic is at the frontier of our subject, therefore we quote only few recent results [Ai-El], [Fa99'], [Ma-Se].

§12. Heat equation on some infinite dimensional groups

Invited to the Hashibara Forum, a satellite Conference of ICM Kyoto 1990, the Group Γ of diffeomeorphism of the circle, in reason of its relation to Mathematical Physics, seems to me a possible subject for this pluridisciplinar Conference under the general thema of "Special Functions". We constructed [Ma-Ma91] a quasi-invariant measure, which was used afterwards to construct some unitary representation of Γ.

Mathematical Physic is mainly interested in a central extension $\tilde{\Gamma}$ of Γ. Identifying the Lie algebra of Γ with the C^∞ functions on the circle T, the central extension is then defined by the cocycle

$$\gamma(u, v) := \int_T \dot{u}(v - \ddot{v})\, d\theta.$$

Denoting by $H(v)$ the Hilbert transform on the crcle of v, and by $\hat{v}(n)$ its n^{th} Fourier coefficient, we have the identity

$$\gamma(H(v), v) = \sum_n (|n| + |n|^3) |\hat{v}(n)|^2.$$

The r.h.s. correspond to the Sobolev norm $H_{3/2}$.

Theorem 12.1 [Ma99]. *The Sobolev norm 3/2 defines on Γ a "Riemannian metric" for which the law π_T of the associated Brownian is carried by a group of homemorphisms.*

The question of quasi-invariance of this measure is open.

Given G as in the preceeding section the *Free Loop Group* is defined as

$$L(G) := \bigcup_{g\in G} gL_e(G).$$

We can identify $L(G)$ with the space of maps $T \mapsto G$; its Lie algebra is identified $L(\mathcal{G})$ the space of free loops on $\mathcal{G}$. The rotation group operates on these two spaces. The $L(\mathcal{G})$-valued Brownian motion is defined by

$$\psi_t(\theta) := \sum_{n=1}^{+\infty} \frac{1}{n}(B_n(t)\cos n\theta + B'_n(t)\sin n\theta),$$

where B_n, B'_n are independent Brownian motions.

The brownian motion on $L(G)$ is defined by solving the following family of G-valued SDE depending upon the parameter θ

$$d_t g_{x,\theta}(t) = g_{(x,\theta)}(t) o\, d_t\psi_t(\theta).$$

It is possible to construct a continuous version relatively to (t,θ) of these equations: denoting $l_t = g_{x,*}(t)$ this version we get an $L(G)$-valued process with continuous trajectories.

There exists on $L(G)$ a left invariant "elliptic operator" $\Delta_{L(G)}$ which is the infinitesimal generator of the process l_t.

For each $t>0$, consider the Brownian bridge measure $\mu^t_{L_e(G)}$ constructed in Section 11 for on $L_e(G)$. Define $\mu^t := \int g\mu^t_{L_e(G)}\, dg$ where dg is the Haar measure of G.

Theorem 12.2 (Airault-Ma 92). *Given f a cylindrical functions on $L(G)$, then we have*

$$\frac{d}{dt}\int f\, d\mu^t = \int \Big(\Delta + \frac{1}{t}a\Big) f\, d\mu^t$$

where

$$a = b - E(b), \quad b = \left\| \int_0^{2\pi} odl(\theta)l^{-1}(\theta)\right\|^2_{\mathcal{G}}.$$

Another point of view is to denote π_t the law at time t of the Brownian motion on $L(G)$, starting for $t = 0$ from the constant loop equal to e. A pending problem for about five years was the question of quasi-invariance of π^t:

Theorem 12.3 (Driver 97'). *Under the left action of $L^1(\mathcal{G})$, the measure π^t is quasi-invariant.*

The difficulty to prove such kind of theorem was the fact that the adjoint action $\mathrm{Ad}(l_x(t))$ is an unbouded operator on $L^1(\mathcal{G})$.

Driver takes the alternative approach to replace left invariant connection on $L(G)$ by the Levi-Civita connection for the underlying Hilbert-Riemann metric of $L(G)$; then it become possible to transfer to $L(G)$ the Bismut-Harnack estimate proved for a compact manifold M. The same strategy as before is used that is study of the geometry of the path space sitting above: $P(L(G))$. As in finite dimension this study depends upon a computation of the Ricci tensor of $L(G)$. Developing this point of view, Fang obtained the logarihmic Sobolev inequality on $P(L(G))$.

For all present and prospective developpements the proof of the existence of Ricci tensor (i.e., the summability of the trace defining Ricci) and its effective computation are questions of paramount importance: see [Sh-Ta], [Fa99] and also all the litterature in Mathematical Physic concerning this point.

Heat operator acts as an homotopy operator realizing a continuous deformation between the considered family of measures μ^t towards the initial measure, the Dirac mass at the starting point. In the case of gaussian measure of variance t on a vector space, for instance in the case of $L(\mathcal{G})$ considered in the previous paragraph, this homotopy can be reealized by the linear operator $\psi \mapsto e^{\lambda}\psi$ which is generated by the *dilatation vector field*: $\psi \mapsto -\psi$. An interesting fact is that the compression vector field exists also in a finite dimensional non linear setting (Airault-Ma96) and also on $L(G)$ (Mancino).

Conclusion. In these exchange of ideas between Geometry and Probability, what is the benefit for each discipline considered from its own sake?

In the case of finite dimensional riemannian geometry, stochastic comparison equation of Section 1 started to raise natural questions which have been subsequently considerably developed nowadays mainly by purely analytic tools do not involving probability. The advantage of probabilistic approach is to be often more conceptual. Probabilistic methods seems to be more difficult to replace in the case of study of boundaries at infinity, in particular for symmetric space of rank > 1.

The mean values formulas for harmonic forms leaded to a fully intrinsic proof of Atiyah-Singer Theorem. After the publication of this proof, purely analytic approach appears using on the frame bundle probabilistic ideas in a disguised analytic form. Here again probability methods have a conceptual impact upon some analytical subsequent developments.

Estimate of ground state of vector bundles stays a very intriguing problem which is not covered untill now by analytic methods. Nevertheless the interesting results avalaible by probabilistic methods do not answer simple geometrical questions as the stability of minimal surfaces.

Some cohomology vanishing theorems obtained by probabilistic approach have not get untill now an alternative proof by classical analytical methods see [El-Li-Ro] as a recent exemple.

The geometry of groups of infinite dimension appearing in Mathematical Physics seems to rely heavily on probabilistic arguments.

From the point of view of probability it can be sayed that the geometric point of view has fully revolutionized the field: Anticipative calculus, Quasi-sure Analysis, Stochastic Calculus of Variations, Stochastic Flows of Diffeomorphisms, Harnack inequalities, representation of martingales through stochastic integral, conditionning, our vision of all these purely probabilistic objects have been fully changed by the geometric point of view.

References

[Ai-El] S. Aida and D. Elworthy, Logarithmic Sobolev inequality on Path Spaces, C. R. Acad. Sci. Paris, **321** (1995), 97–102.

[Ai-Ma88] H. Airault et P. Malliavin, Intération géométrique sur l'espace de Wiener, Bul. Sci. Math., **112** (1988), 3–52.

[Ai-Ma92] H. Airault and P. Malliavin, Integration on Loop Group II: Heat Equation for the Wiener measure, J. Funct. Anal., **104** (1992), 71–109.

[Ai-Ma95] H. Airault and P. Malliavin, Semi-martingales with values in an euclidean vector bundle and Ocone's formula on a Riemannian manifold, Proc. Symp. Pure and Appl. Math., **57**, (M. Cranston and M. Pinsky ed.), Cornell (1993), Amer. Math. Soc. (1995), 173–193.

[Ai-Ma96] H. Airault and P. Malliavin, Integration by part formula and dilatation vector fields on elliptic probability spaces, Prob. Th. Rel. Fields, **106** (1996), 447–494.

[An-Dr] L. Anderson and B. Driver, Finite dimensional Approximations to Wiener Measure and Path space integral formulas on Manifolds, J. Funct. Anal., **165** (1999), 431–499.

[Be] G. Ben Arous, Flots et séries de Taylor stochastiques, Prob. Th. Rel. Fields, **81** (1989), 73–99.

[Bi84] J. M. Bismut, The Atiyah-Singer theorems, a probabilistic approach, J. Funct. Analysis, **57** (1984), 56–99.

[Bi84'] J. M. Bismut, "Large deviations and the Malliavin Calculus", Birkhäuser Progress in Math., **45**, 1984, pp. 216.

[Bo-Mw] V. Bogachev and E. Mayer-Wolf, Absolutely Continuous Flows Generated by Sobolev Class Vector Fields in finite or Infinite Dimensions, J. Funct. Analysis, **167** (1999), 1–68.

[Cl] J. M. C. Clark, The representation of functionals of the Brownian motion by stochastic integrals, Ann. Math. Statistics, **41** (1970), 1282–1295.

[Cr] A. B. Cruzeiro, Equations diffentielles sur l'espace de Wiener, J. Funct. Anal., **54** (1983), 206–227.

[Cr-Fa] A. B. Cruzeiro and S. Fang, L^2-estimate for Riemannian Anticipative Stochastic integrals, J. Funct. Anal., **143** (1997), 400–414.

[Cr-Ma96] A. B. Cruzeiro and P. Malliavin, Renormalized Differential Geometry on path spaces, I – Structural equation, curvature, J. Funct. Anal., **139** (1996), 119–181.

[Cr-Ma98] A. B. Cruzeiro and P. Malliavin, Non pertubative construction of invariant mesaures through confinement by curvature, J. Math. pures et appl., **77** (1998), 527–537.

[De-Ga-Ma] A. Debiard, B. Gaveau et E. Mazet, Théorèms de comparaison en géométrie, J. Res. Inst Kyoto Univ., **12** (1976), 391–429.

[Dr92] B. Driver, A Cameron-Martin type quasi-invariance theorem for Brownian motion on a Riemannian manifold, J. Funct. Anal., **110** (1992), 272–376.

[Dr97] B. Driver, Integration by part for heat kernel measure revisited, J. Math. pures et appl., **76** (1997), 703–737.

[Dr97'] B. Driver, Integration by prt and quasi-invariance for heat kernel measures on loop groups, J. Funct. Anal., **149** (1997), 470–547.

[Dr99] B. Driver, The Lie bracket of adapted vector fields on Wiener spaces, Appl. Math. Optim., **39,** No.2 (1999), 179–210.

[Dr-Lo] B. Driver and T. Lohrentz, Logarihmmic Sobolev inequalities for pinned loop groups, J. Funct. Anal., **140** (1996), 381–448.

[El-Le-Li] D. Elworthy, Y. Le Jan and X. M. Li, Integration by part for degenerated diffusions on path space and diffeomorphism group, C. R. Acad. Sci. Paris, **325** (1996), 921–926.

[El-Li] D. Elworthy and X. M. Li, Formula for derived heat semi-group, J. Funct. Anal, **125** (1994), 252–286.

[El-Li-Ro] D. Elworthy, X. M. Li and S. Rosenberg, Bounded and L^2 harmonic forms on universal cover, Geom. Funct. Anal., **8** (1998), 283–305.

[En-St] O. Enchev and D. Stroock, Towards a Riemannian Geometry on the path space over a riemannian manifold, J. Funct. Anal., **134** (1996), 392–416.

[Fa98] S. Fang, Stochastic Anticipative Calculus over the Path space over a compact riemannian manifold, J. Math. pures et appl., **77** (1998), 249–282.

[Fa99] S. Fang, Ricci tensors on some infinite dimensional Lie Algebras, J. Funct. Anal., **161** (1999), 132–161.

[Fa99'] S. Fang, Integration by part and logarithmic Sobolev inequality on the path space over loop groups, The Annals of Probability, **27** (1999), 664–683.

[Fa-Ma] S. Fang and P. Malliavin, Stochastic Analysis on the Path space of a riemannian manifold I. Markovian Stochastic Calculus, J. Funct. Anal., **118** (1993), 249–274.

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

[Ga-Tr] B. Gaveau et P. H. Trauber, L'intégrale stochastique comme opérateur de divergence dans l'espace fonctionnel, J. Funct. Anal., **46** (1982), 230–238.

[Gr] L. Gross, Uniqueness of ground state for Schrödinger operators over Looop Groups, J. Funct. Anal., **112** (1993), 373–441.

[Hs] E. P. Hsu, Quasi invariance of the Wiener measure on the path space over a compact Riemann manifold, J. Funct. Anal., **134** (1995), 417–450.

[Ik-Ma] N. Ikeda and H. Matsumoto, Brownian motion on the hyperbolic plane and Selberg trace formula, J. Funct. Anal., **163** (1998), 63–110.

[Ik-Wa] N. Ikeda and S. Watanabe, "Stochastic differential equations and diffusion processes", North Holland, 1989, pp. 554.

[It50] K. Itô, Stochastic differential equation on a manifold, Nagoya Math. J., **1** (1950), 35–47.

[It62] K. Itô, The Brownian motion and tensor fields on Riemannian manifolds, Proc. ICM-62 Stockholm, 536–539.

[It90] K. Itô, "Positive generalized functionals on R^∞", Proc. Coll. White Noise analysis, Bielefeld 1989, World Scientific 1990, pp. 166–179.

[Kun] H. Kunita, "Stochastic flows and stochastic differential equations", Cambridge Univ. Press, 1988.

[Kus] Kusuoka, Analysis on Wiener Space II. Differential forms, J. Funct. Anal., **103** (1992), 229–274.

[Le] P. Lescot, Un théorème de désintégration en Analyse quasi-sure, Sém. Proba. XXVII, Springer Lecture notes, **1557** (1993), 256–275.

[Ma74] P. Malliavin, Asymptotic of the Green function of a riemannian manifold and Its stochastic integral, Proc. Nat. Acad. Sci., **71** (1974), 381–383.

[Ma74'] P. Malliavin, Formules de la moyenne, calcul des pertubations et théorémes d'annulation pour les formes harmoniques, J. Funct. Anal., **17** (1974), 274–291.

[Ma75] P. Malliavin, Paramétrix trajectorielle, C. R. Acad. Sciences Paris, **281** (1975), 241–245.

[Ma76] P. Malliavin, Stochastic Calculus of Variations and hypoelliptic operators, in "Proc. Int. Symp. on SDE", (K. Itô, ed.), Kyoto, 1976, pp. 195–263.

[Ma77] P. Malliavin, Un principe de transfert et son application au calcul stochastique, C. R. Acad. Sci. Paris, **284** (1977), 187–189.

[Ma78] P. Malliavin, "Géométrie Différentielle stochastique", Sém. Univ. Montréal **64**, 1978, p. 132.

[Ma82] P. Malliavin, Implicit functions in finite corank on the Wiener Space, Proc. the Taniguchi Symp. Stoch. Anal., Katata/Kyoto K. Itô ed., 369–386.

[Ma85] P. Malliavin, Intégrales Stochastiques oscillantes, C. R. Acad. Sciences Paris, **301** (1985), 141–143; 767–770 * **302** (1986), 73–78; 359–362; 481–486.

[Ma97] P. Malliavin, "Stochastic Analysis", Grundlehren der mathematischen Wissenschaften, Springer, 1997, pp. 343.

[Ma99] P. Malliavin, The canonic diffusion above the diffeomorphism group of the circle, C. R. Acad. Sci. Paris, **329** (1999), 325–330.

[Ma-Ma74] M. P. Malliavin et P. Malliavin, Factorisation et lois limites de la diffusion horizontale au-dessus d'un espace riemannien symétrique, Springer Lecture Notes in Math., **404** (1974), 155–210.

[Ma-Ma90] M. P. Malliavin and P. Malliavin, Integration on loop groups I: quasi invariant measures, J. Funct. Anal., **93** (1990), 207–237.

[Ma-Ma91] M. P. Malliavin and P. Malliavin, An infinitesimally quasi-invariant measure on the group of diffeomorphisms of the circle, in "Special functions", (M. Kashiwara and T. Miwa, eds.), ICM 90 Satellite Conference, Springer, Tokyo, 1991, pp. 234–244.

[Ma-Se] Z. Ma and Sen, Logarithmic Sobolev inequalities on Loop spaces, J. Funct. Anal. (1999).

[Ma-Sh] S. Manabe and I. Shigekawa, A Comparison Theorem for Eigenvalue of the covariant Laplacian, J. Funct. Anal., **94** (1990), 349–357.

[Ma-St] P. Malliavin and D. Stroock, Short time behaviour of the heat kernel and of its logarihmic derivatives, J. Differential Geometry, **44** (1996), 550–570.

[Man] M. E. Mancino, Dilatation vector field on the Loop group, J. Funct. Anal., **166** (1999), 130–147.

[Ma-Ue] H. Matsumoto and N. Ueki, Spectral analysis of Schrödinger operators with magnetic field, J. Funct. Anal., **140** (1996), 218–255.

[Nu-Pa] D. Nualart and E. Pardoux, Stochastic Calculus with anticipating integrand, Prob. Th. Rel. Fields, **78** (1988), 535–581.

[Nu-Za] D. Nualart and M. Zakai, Generalized stochastic integrals and the Malliavin calculus, Prob. Th. Rel. Fields, **73** (1986), 255–280.

[Pr75] J. J. Prat, Convergence angulaire de la diffusion sur une variété à courbure négative, Comp. Rend. Acad. Sci. Paris, **286** (1975), 687–690.

[Pr93] J. J. Prat, Résolution de l'équation de Schrödinger par intégration stochastique du champ magnétique, Prob. Th. Rel. Fields, **95** (1993), 217–237.

[Sh86] I. Sigekawa, The de Rham-Hodge-Kodaira decomposition on an Abstract Wiener space, J. Math. Kyoto Univ., **26** (1986), 191–202.

[Sh94] I. Shigekawa, Eigenvalue problem for Schrödinger operator with magnetic field, J. Funct. Anal., **120** (1994), 344–379.

[Sh-Ta] I. Shigekawa and S. Taniguchi, A Kähler metric on based Loop Group and a covariant differentiation, in "Its Stochstic calculus and Probability Theory", (Ikeda, Watanabe, Fukushima, Kunita, eds.), Springer, Tokyo, 1996, pp. 327–347.

[St-Ta] D. W. Stroock and S. Taniguchi, Diffusions as integral curves of vector fields. Stratonovitch diffusion without Integration, in "Vol. in honour of E. Dynkin", (Freidlin, ed.), 1994, Birkhäuser Prog. in Prob., **34**, pp. 333–369.

[Su] H. Sugita, Positive generalized functions and potential theory over an abstract Wiener space, Osaka Math. J., **25** (1988), 665–696.

[Tan] S. Taniguchi, On the exponential decay of oscilatory Integrals on an abstract Wiener space, J. Funct. Anal., **154** (1998), 424–443.

[Ue] N. Ueki, Lower Bounds for the spectra of Schrödinger operators with magnetic fields, J. Funct. Anal., **120** (1994), 344–379.

[Wa] S. Watanabe, Lectures on stochastic differential equations and Malliavin Calculus, in "Lecture Tata Institute for fundamental research **73**", 1984, p. 111.

10 rue Saint Louis en l'Isle, 75004
Paris
`sli@ccr.jussieu.fr`

Advanced Studies in Pure Mathematics 31, 2001
Taniguchi Conference on Mathematics Nara '98
pp. 211–233

The Intrinsic Hodge Theory of p-adic Hyperbolic Curves

Shinichi Mochizuki

Contents

§1. Uniformization Theory as a Hodge Theory at Arithmetic Primes

(A) Uniformization as a Catalogue of Rational Points

We begin our discussion by posing the following fundamental problem concerning algebraic varieties over the complex numbers (where, roughly speaking, an "algebraic variety over the complex numbers" is a geometric object defined by polynomial equations with coefficients which are complex numbers):

> **Problem**: Given an *algebraic variety* Z over $\mathbf{C}$, it is possible to give some sort of *natural explicit catalogue* of the rational points $Z(\mathbf{C})$ of Z?

Received March 5, 1999.

To gain a sense of what is meant by the expression "a natural explicit catalogue", it is useful to begin by thinking about some basic examples. Perhaps the simplest nontrivial examples of algebraic varieties are *plane curves*, i.e., subvarieties of $\mathbf{A}^2_{\mathbf{C}}$ (two-dimensional affine space over $\mathbf{C}$) defined by a single polynomial equation

$$f(X,Y) = 0$$

in two variables. In this case, the set of rational points $Z(\mathbf{C})$ of the corresponding variety Z is given by

$$Z(\mathbf{C}) = \{(x,y) \in \mathbf{C}^2 \mid f(x,y) = 0\}.$$

Moreover, we can classify plane curves by the degree of the defining equation $f(X,Y)$. We then see that the resulting sets $Z(\mathbf{C})$ may be explicitly described as follows:

(1) <u>The Linear Case</u> ($\deg(f) = 1$): Up to coordinate transformations, this is the case given by the equation $f(X,Y) = X$. In this case, we then obtain an explicit catalogue of the rational points by:

$$(0,?) : \mathbf{C} \xrightarrow{\sim} Z(\mathbf{C})$$

(i.e., mapping $z \in \mathbf{C}$ to $(0,z) \in Z(\mathbf{C})$).

(2) <u>The Quadratic Case</u> ($\deg(f) = 2$): Up to coordinate transformations (and ruling out degenerate cases), we see that this is essentially the case where the equation $f(X,Y) = X \cdot Y - 1$. In this case, an explicit catalogue is given by the *exponential map*:

$$\exp : \mathbf{C} \longrightarrow Z(\mathbf{C}) = \mathbf{C}^{\times}$$

(In fact, the map may be defined intrinsically, without using coordinate transformations to render the defining equation in the "standard form" $X \cdot Y = 1$).

(3) <u>The Cubic Case</u> ($\deg(f) = 3$): Up to adding the point(s) at infinity, this is essentially the case where we are dealing with an *elliptic curve* E. In this case, as well, we have a natural exponential map:

$$\exp_E : T_E \longrightarrow E(\mathbf{C})$$

(where T_E is the one-dimensional complex vector space given by the tangent space to some fixed point – "the origin" – of E). This map allows us to think of E as being of the form "$\mathbf{C}/\Lambda$" (where $\Lambda \cong \mathbf{Z}^2$ is a lattice in $\mathbf{C}$).

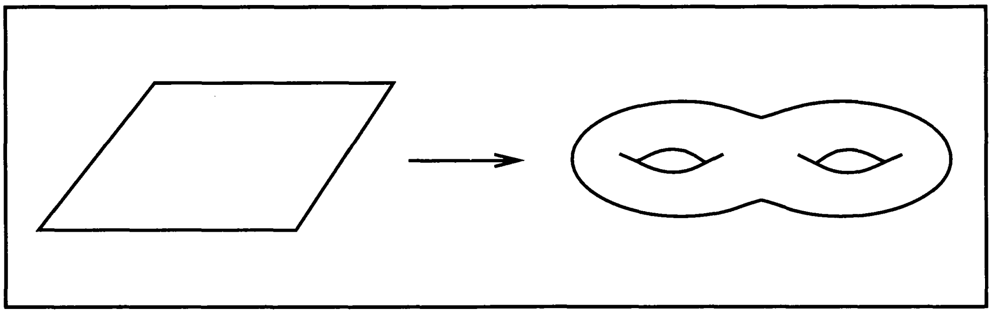
(4) <u>*Higher Degree*</u>: If $\deg(f) \geq 4$, or we wish to consider lower degree cases with lots of points removed, then we are led naturally to the following notion:

A *hyperbolic curve* Z is a smooth, proper connected algebraic curve of genus g with r points removed, where we assume that $2g - 2 + r > 0$.

According to the *uniformization theorem of Köbe*, hyperbolic curves may be uniformized by the upper half-plane $\mathfrak{H} \stackrel{\text{def}}{=} \{z \in \mathbf{C} \mid \mathrm{Im}(z) > 0\}$, i.e., we have a surjective (holomorphic) covering map:

$$\mathfrak{H} \longrightarrow Z(\mathbf{C})$$

which allows us to think of $Z(\mathbf{C})$ as being of the form $\mathfrak{H}/\Gamma$, where Γ is some discrete group acting on $\mathfrak{H}$.



(B) "Intrinsic" Hodge Theories

In the preceding section, we posed the problem of *explicitly cataloguing the rational points of a variety* (over $\mathbf{C}$). By looking at various examples, we saw that this problem may also be worded – in perhaps more familiar terms – as the problem of finding natural *uniformizations* of varieties. In the present section, we would like to further refine our understanding of the problem of finding natural uniformizations/explicit catalogues of rational points by rewording this problem in terms of the language of "Hodge theory".

First, let us discuss what we mean in general by the notion of a "*Hodge theory*". By a Hodge theory, we shall mean an *equivalence* of the following form:

$$\begin{pmatrix} \text{algebraic} \\ \text{geometry} \\ \text{(e.g., rational} \\ \text{points)} \end{pmatrix} \Longleftrightarrow \begin{pmatrix} \underline{\text{Over } \mathbf{C}}\text{: topology +} \\ \text{differential geometry} \\ \\ \underline{\text{Over the } p\text{-adics}}\text{:} \\ \text{pro-}p \text{ étale topology +} \\ \text{Galois action} \end{pmatrix}$$

The most familiar example of such an equivalence is the *Hodge theory of cohomologies.* Over $\mathbf{C}$, this amounts to "classical Hodge theory", i.e., the well-known isomorphism between the de Rham cohomology of an algebraic variety (which is well-known to be an *algebro-geometric* invariant of the variety) and the singular cohomology of the variety (which is a *topological* invariant). More recently, the *p-adic Hodge theory of cohomologies* has been developed by Fontaine et al. (cf., e.g., [Falt2], [Falt3]). This theory asserts an equivalence between the (algebraic) de Rham cohomology of an algebraic variety over a finite extension K of $\mathbf{Q}_p$ and the p-adic étale cohomology of the variety, equipped with its natural Galois action (i.e., action of $\mathrm{Gal}(K)$).

This "Hodge theory of cohomologies" is the most basic example of a "Hodge theory" as defined above. In the present manuscript, however, we would like to consider a different kind of Hodge theory which we shall call an *intrinsic Hodge theory*, or IHT, for short. By an intrinsic Hodge theory, we mean a Hodge theory – i.e., an equivalence of the form discussed above – in which the invariant which appears on the "algebraic geometry" side is the "*variety itself*".

There are several different ways to interpret the phrase "the variety itself". In the present manuscript, we shall consider the following two interpretations:

(1) <u>The Physical Interpretation</u>: In this interpretation, one takes the phrase the "variety itself" to mean the "rational points of the variety".

(2) <u>The Modular Interpretation</u>: In this interpretation, one takes the phrase the "variety itself" to mean the "moduli of the variety".

Note that (it is a tautology of terminology that) a physical IHT essentially amounts to some sort of explicit description of the rational points of the variety in terms of topology and geometry/Galois theory. Thus, one may summarize the above discussion as follows:

$$\begin{aligned} \text{physical IHT} &= \text{uniformization of the variety,} \\ \text{modular IHT} &= \text{uniformization of the moduli space of} \\ &\qquad\qquad \text{deformations of the variety.} \end{aligned}$$

Before concluding this section, we make some remarks on the relationship between the "IHT's" that we wish to discuss here and the more well-known "Hodge theories of cohomologies". First of all, although in general, IHT's are not the same as Hodge theories of cohomologies, typically in proving theorems which realize IHT's, the technical tools of the corresponding Hodge theory of cohomologies (e.g., in the p-adic case, the techniques of so-called "p-adic Hodge theory" as in [Falt2]) play an important role. Secondly, in the case of $\mathbf{G}_{\mathrm{m}} = V(X \cdot Y - 1)$ (i.e., Example (2) in §1, (A)), as well as in the case of elliptic curves (i.e., Example (3) in §1, (A)), it just so happens that the first cohomology module of the curve "is" the curve itself, i.e., in more sophisticated language, in these cases the curve in question is a *1-motive.* Thus, in these cases, it turns out that the notions of IHT and Hodge theory of H^1 *coincide.* In particular, in these cases, the well-known Hodge theory of H^1 already provides a uniformization of the curve. Note that this differs quite substantially from the case of higher genus (Example (4) in §1, (A)).

(C) Completion at Arithmetic Primes

So far in our discussion, we have ignored the important issue of what sort of *ground field* should be considered in our discussion of uniformizations/explicit catalogues of rational points/IHT's. In the examples of §1, (A), we considered the situation over *the complex number field,* since this is the most elementary and well-known example of a ground field over which IHT's may be realized.

Of course, ideally, one would like to realize IHT's over any field, for instance, over a *number field* (i.e., a finite extension of $\mathbf{Q}$) – a case for which the problem of determining the set of rational points explicitly is of prime interest. Unfortunately, however, typically, in order to realize an IHT (or, indeed, any sort of "Hodge theory"), one must work over a base which is complete with respect to some sort of "arithmetic prime". The three main examples of this sort of base are the following:

(i) <u>the complex number field $\mathbf{C}$</u> (this also covers the case of $\mathbf{R}$ by working with objects over $\mathbf{C}$ equipped with an action of complex conjugation, i.e., of $\mathrm{Gal}(\mathbf{C}/\mathbf{R})$)

(ii) <u>a p-adic field K</u> (i.e., a finite extension of $\mathbf{Q}_p$)

(iii) <u>power series over $\mathbf{Z}$</u> – typically arising as the completion of some sort of moduli space at a $\mathbf{Z}$-valued point corresponding to a *degenerate object* parametrized by the moduli space.

Indeed, all completions of number fields fall under cases (i) and (ii). Thus, if one is ultimately interested in rational points of number fields, IHT's over bases as in (i) and (ii) are of natural interest. Also, since the coefficients of the powers series appearing in (iii) are integers, case (iii) is also of substantial arithmetic interest.

In the following discussion, the following principle will serve as an important guide:

Guiding Principle: For every type of arithmetic prime (i.e., cases (i), (ii), and (iii) above), one expects that there should be a *canonical uniformization theory* at that prime.

In general, however, one does not expect that the canonical uniformization theories at different primes should be compatible with one another. We will return to this point in more detail later.

In terms of types of varieties, the main cases in which one has well-developed physical and modular IHT's are the following:

(1) abelian varieties.

(2) hyperbolic curves.

The physical and modular IHT's in these cases may be roughly summarized in the following charts:

(1) Abelian Varieties

$\mathbf{C}$	p-adic	Degenerate Object
exponential map of abelian varieties/ Siegel upper half-plane uniformization	Tate's theorem/ Serre-Tate theory	Schottky uniformizations of Tate/Mumford/ Faltings/Chai

(2) Hyperbolic Curves

$\mathbf{C}$	p-adic	Degenerate Object
Fuchsian uniformization/ uniformizations of Teichmüller and Bers	§2 (anabelian conjecture)/ §3 (p-adic Teichmüller theory)	formal algebraic Schottky uniformization of Mumford

Of these two examples, undoubtedly the example of abelian varieties is the more well-known. *Over* $\mathbf{C}$, the exponential map of an abelian variety gives a uniformization of the abelian variety by a complex linear space. This generalizes Example (3) of §1, (A). Moreover, by using the periods that one obtains from this uniformization, one obtains a uniformization of the moduli space of abelian varieties by the Siegel upper half-plane. Thus, we see both the *physical* and *modular* aspects of the IHT of abelian varieties in evidence.

In the *p-adic case*, the IHT of an abelian variety essentially amounts to the p-adic Hodge theory of H^1 of the abelian variety. Although the p-adic Hodge theory of H^1 of an abelian variety has many different aspects, most of these may be traced to the fundamental paper of Tate

([Tate]) in the late 1960's. In this paper, the main theorem ("Tate's theorem" in the chart) states that homomorphisms between formal groups (e.g., the formal groups arising from abelian varieties) over p-adic fields are essentially equivalent to homomorphisms between the corresponding Tate modules. In some sense, this result is the analogue for abelian varieties of the main theorem (Theorem 1) of §2 below, and may be regarded as a sort of physical IHT for abelian varieties. On the other hand, the modular aspect of the IHT of abelian varieties may be seen most easily in Serre-Tate theory, which gives rise to p-adic parameters on the moduli space of abelian varieties that are very much analogous to the Siegel upper half-plane uniformization in the complex case.

Finally, in a neighborhood of a point in the moduli space corresponding to a *degenerating abelian variety*, one has the theory of Tate curves, generalized by Mumford and Faltings/Chai ([FC]). Moreover, it turns out that in the case of abelian varieties, the complex, p-adic, and degenerating object theories are all compatible with one another. For instance, if one specializes the uniformizing parameters that one obtains on the moduli space in a neighborhood of a point corresponding to a degenerating abelian variety to a p-adic (respectively, complex) base, one obtains parameters compatible with the Serre-Tate parameters (respectively, the Siegel upper half-plane uniformization).

Next, we consider the case of *hyperbolic curves*. *Over* $\mathbf{C}$, the physical aspect of the IHT of hyperbolic curves (cf. Example (4) of §1, (A)) essentially amounts to the Fuchsian uniformization. Then just as the exponential map uniformization of an abelian variety "induces" the Siegel upper half-plane uniformization of the moduli space of abelian varieties, the Fuchsian uniformization of a hyperbolic curve "induces" the Bers uniformization of the moduli space of hyperbolic curves (cf. §3, (A) below, as well as the Introduction of [Mzk4]).

In a neighborhood of a point in the moduli space corresponding to a totally degenerate (proper) hyperbolic curve, one has the theory of [Mumf]. Note, however, that this theory is *not* compatible with the theory of the Fuchsian uniformization in the following sense: If one specializes the ($\mathbf{Z}$-coefficient) power series in the base ring to some $\mathbf{C}$-valued point in a neighborhood of a point in the moduli space corresponding to a totally degenerate curve, the resulting uniformization over $\mathbf{C}$ that one obtains is the so-called *Schottky uniformization* of the curve. This uniformization is completely different from the Fuchsian uniformization.

Finally, we come to the *p-adic case*. It seems that the IHT of p-adic hyperbolic curves has not been studied extensively until relatively recently ([Mzk1-5]). Thus, *It is the goal of this manuscript to report on recent developments concerning the intrinsic Hodge theory* (IHT) *of p-adic hyperbolic curves.*

The physical aspect, which concerns a theorem (Theorem 1) that gives a strong solution to *Grothendieck's anabelian conjecture*, will be the topic of §2. The modular aspect, which concerns a theory – which we call *p-adic Teichmüller theory* – which may be regarded as either the hyperbolic curve analogue of Serre-Tate theory or the p-adic analogue of the theory of the Fuchsian and Bers uniformizations, will be discussed in §3. We remark here that this p-adic Teichmüller theory is *not* compatible with the specialization of the theory of [Mumf] to the p-adic case. This may disappoint some readers, but is, in fact, natural in view of the fact that even over $\mathbf{C}$, the specialization of the theory of [Mumf] to the complex numbers is not compatible (as remarked above) with the theory of the Fuchsian uniformization. Moreover, it is in line with the general Guiding Principle discussed above that to each sort of arithmetic prime there should correspond a natural uniformization theory of hyperbolic curves. Thus, it seems to the author that it is meaningless to argue as to whether it is the specialization of the [Mumf] to the p-adic case or the theory of [Mzk1], [Mzk2], [Mzk3], [Mzk4] which is the "true" p-adic analogue of the Fuchsian uniformization. That is to say, it seems more natural to the author to regard the theory of [Mumf] as the "true" analogue of the Fuchsian uniformization at the "degenerating object prime", and the theory of [Mzk1], [Mzk2], [Mzk3], [Mzk4] as the "true" analogue of the Fuchsian uniformization at "the prime p".

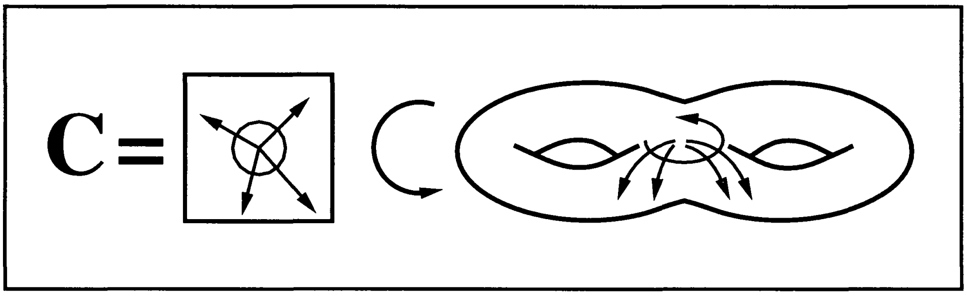
§2. The Physical Aspect: Embedding by Automorphic Forms

(A) The Complex Case

We begin our discussion by considering the complex case. The complex case is important to understand not only for reasons of philosophical analogy, but also because it provides the motivation for the proof of the main result in the p-adic case.

In the complex case, the physical IHT aspect of the Fuchsian uniformization may be summarized schematically as follows:

$$\begin{aligned}
&\underline{\text{algebraic}}\text{ hyperbolic curve } X\\
&\quad\Longleftrightarrow SO(2)\backslash PSL_2(\mathbf{R})/\Gamma \quad \text{(physical/analytic obj.)}\\
&\quad\Longleftrightarrow \pi_1(\mathcal{X}) + \text{action of } \pi_1(\mathcal{X}) \text{ on } \mathfrak{H}\\
&\quad\Longleftrightarrow \underline{\text{topology}} + \underline{\text{arithmetic structure}} \text{ (geometry)}
\end{aligned}$$



Here, the illustration is of an *algebraic* hyperbolic curve thought of as a topological surface equipped with an additional *arithmetic structure*, namely the geometry arising from the Poincaré metric on the upper half-plane $\mathfrak{H}$. This geometry is depicted as an "action of $\mathbf{C}$" on the topogical surface, given by the flows along geodesics defined by the geometry.

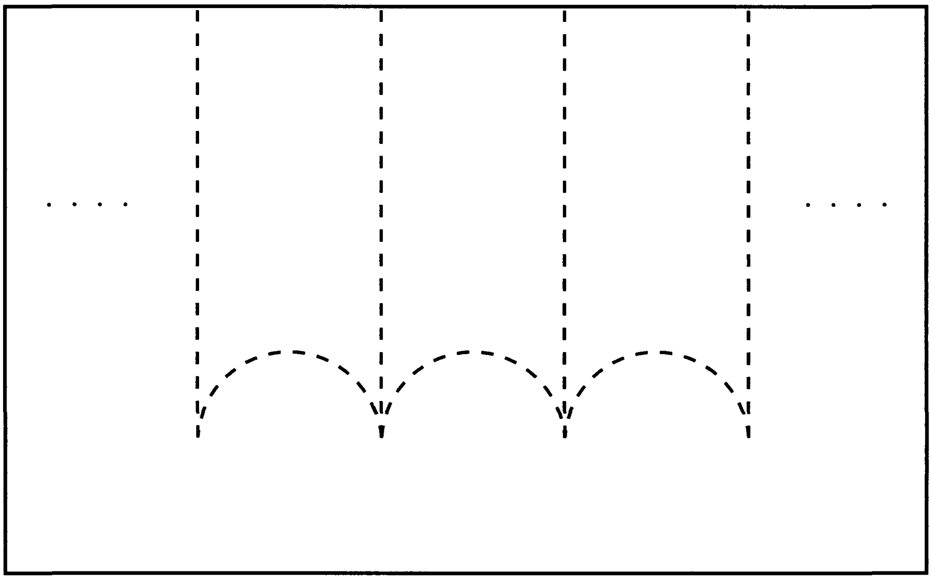
In order to generalize this picture to the p-adic case, it is necessary to recall the conceptual machinery that gives rise to the first "$\Longleftrightarrow$" in the chain of equivalences appearing above. If one starts from the right-hand side, i.e., the upper half-plane equipped with some action of $\pi_1(\mathcal{X})$, then the *algebraic* structure (i.e., the left-hand side) may be recovered by considering *automorphic forms* on $\mathfrak{H}$ which are invariant with the respect to the action of $\pi_1(\mathcal{X})$. These automorphic forms define a morphism from $\mathfrak{H}$ to projective space whose image is necessarily (by Chow's Theorem!) *algebraic* and, in fact, equal to the original algebraic curve X. That is to say, one has a commutative diagram:

$$\begin{array}{ccc} \text{Upper half-plane } \mathfrak{H} & \longrightarrow & \text{Projective Space} \\ \downarrow & & \Vert \\ \text{Algebraic Curve } X & \hookrightarrow & \text{Projective Space} \end{array}$$

Put another way, the main point is the following: Although ultimately $\pi_1(\mathcal{X})$-invariant differential forms on $\mathfrak{H}$ define algebraic forms on X, such forms may be defined *directly* from the data of the action $\{\pi_1(\mathcal{X}) \curvearrowright \mathfrak{H}\}$ and, moreover, by the above diagram, allow one to recover the *algebraic structure* of X from the analytic data $\{\pi_1(\mathcal{X}) \curvearrowright \mathfrak{H}\}$. That is, to put the matter more succinctly, the key idea is the following:

> Key Idea: *Consider analytic reprepresentations of algebraic differential forms.*

It turns out that the proof of the main p-adic result (Theorem 1 below) of this section consists precisely of implementing this key idea in the p-adic case by using the technical machinery of p-adic Hodge theory.



The Case of $SL_2(\mathbf{Z})$

(B) The Arithmetic Fundamental Group

In this section, we prepare for the statement of Theorem 1 in the following section by introducing various notations and terminology.

Let K be a field of characteristic zero. Let us denote by $\overline{K}$ an algebraic closure of K. Let $\Gamma_K \stackrel{\text{def}}{=} \text{Gal}(\overline{K}/K)$. Suppose that X_K is a *variety over* K. Then we will denote by

$$\pi_1(X_K)$$

the *algebraic fundamental group of* X_K (cf. [SGA1]). This group is a compact, profinite topological group, well-defined up to inner automorphism (since we did not specify a "base-point"), and which has the following property: The category of finite sets with a continuous $\pi_1(X_K)$-action is naturally equivalent to the category of finite étale coverings of X_K. Moreover, if K is, for instance, an algebraically closed subfield of $\mathbf{C}$, then $\pi_1(X_K)$ may be identified with the *profinite completion* (= inverse limit of all finite quotients) *of the usual topological fundamental group* $\pi_1^{\text{top}}(X_{\mathbf{C}})$ (where $X_{\mathbf{C}} \stackrel{\text{def}}{=} X_K \otimes_K \mathbf{C}$).

Now let X_K be a *hyperbolic curve over* K; write $X_{\overline{K}} \stackrel{\text{def}}{=} X \times_K \overline{K}$. Then one has an exact sequence

$$1 \longrightarrow \pi_1(X_{\overline{K}}) \longrightarrow \pi_1(X_K) \longrightarrow \Gamma_K \longrightarrow 1$$

of *algebraic fundamental groups.* We shall refer to $\pi_1(X_{\overline{K}})$ as the *geometric fundamental group* of X_K. Note that, by the above discussion of the case where $\overline{K} \subseteq \mathbf{C}$, it follows that the structure of $\pi_1(X_{\overline{K}})$ is determined entirely by (g, r) (i.e., the "type" of the hyperbolic curve X_K). In particular, $\pi_1(X_{\overline{K}})$ *does not depend on the moduli of* X_K. Of course, this results from the fact that K is of *characteristic zero.* In positive characteristic, on the other hand, preliminary evidence ([Tama2]) suggests that the fundamental group of a hyperbolic curve over an algebraically closed field (far from being independent of the moduli of the curve!) may in fact *completely determine* the moduli of the curve.

We shall refer to $\pi_1(X_K)$ (equipped with its augmentation to Γ_K) as the *arithmetic fundamental group* of X_K. Although it is made up of two "parts" – i.e., $\pi_1(X_{\overline{K}})$ and Γ_K – which do not depend on the moduli of X_K, it is not unreasonable to expect that the extension class defined by the above exact sequence, i.e., the structure of $\pi_1(X_K)$ as a group equipped with augmentation to Γ_K, may in fact depend quite strongly on the moduli of X_K. Indeed, according to the *anabelian philosophy* of Grothendieck (cf. [LS]), for "sufficiently arithmetic" K, one expects that *the structure of the arithmetic fundamental group* $\pi_1(X_K)$ *should be enough to determine the moduli of* X_K. Although many important versions of Grothendieck's anabelian conjectures remain unsolved (most notably the so-called *Section Conjecture* (cf., e.g., [LS], p. 289, 2)), in the remainder of this section, we shall discuss various versions that have been resolved in the affirmative. For instance, such a version of these conjectures which will be discussed in (B) below (Theorem 1) states roughly that (*nonconstant*) *morphisms from a smooth* K*-variety to* X_K *should be in bijective correspondence with* (*open*) *homomorphisms* (*over* Γ_K) *between the corresponding arithmetic fundamental groups.* Thus, there is an obvious analogy between this (form of Grothendieck's) conjecture and the Tate conjecture on abelian varieties, which states roughly that morphisms between abelian varieties are equivalent to morphisms between their arithmetic fundamental groups.

Note that this anabelian philosophy is a *special case of the notion of "intrinsic Hodge theory"* discussed above: indeed, on the algebraic geometry side, one has "the curve itself", whereas on the topology plus arithmetic side, one has the arithmetic fundamental group, i.e., the purely (étale) topological $\pi_1(X_{\overline{K}})$, equipped with the structure of extension given by the above exact sequence.

In fact, it is interesting to note – especially relative to the discussion at the beginning of §1, (C) – that Grothendieck's anabelian philosophy arose as an *approach to diophantine geometry.* It is primarily for this reason that it was originally thought that the most natural sort of "arithmetic" base field K over which one should expect Grothendieck's anabelian conjectures to hold was a *number field.* Another reason for the idea that the base field in these conjectures should be a number field was the analogy with Tate's conjecture on homomorphisms between abelian varieties (cf., e.g., [Falt1]). Indeed, in discussions of Grothendieck's anabelian philosophy, it was common to refer to statements such as that of Theorem 1 below as the "anabelian Tate conjecture", or the "Tate conjecture for hyperbolic curves". In fact, however, there is an important difference between Theorem 1 and the "Tate conjecture" of, say, [Falt1]: Namely, whereas Theorem 1 below holds over *local fields* (i.e., finite extensions of $\mathbf{Q}_p$), *the Tate conjecture for abelian varieties is false over local fields.* Moreover, until the proof of Theorem 1, it was generally thought that, just like its abelian cousin, the "anabelian Tate conjecture" was essentially global in nature. That is to say, it appears that the point of view of the author, i.e., that Theorem 1 should be regarded as a p-adic version of the "physical aspect" of the Fuchsian uniformization of a hyperbolic curve, does not exist in the literature (prior to the work of the author).

Finally, we remark, relative to the issue of locating an analogue of Theorem 1 in the theory of abelian varieties, that it seems that it is more natural to think of "Tate's theorem" (cf. the discussion in §1, (C)) as the proper analogue of Theorem 1 for abelian varieties. Indeed, not only does Tate's theorem hold over local fields, but it plays an important technical role in the proof of Theorem 1 below (cf. [Mzk5]).

(C) The Main Theorem

Building on earlier work of H. Nakamura and A. Tamagawa (see, especially, [Tama1]), the author applied the p-adic Hodge theory of [Falt2] and [BK] to prove the following result (cf. Theorem A of [Mzk5]):

Theorem 1. *Let p be a prime number. Let K be a subfield of a finitely generated field extension of $\mathbf{Q}_p$. Let X_K be a hyperbolic curve over K. Then for any smooth variety S_K over K, the natural map*

$$X_K(S_K)^{\mathrm{dom}} \longrightarrow \mathrm{Hom}^{\mathrm{open}}_{\Gamma_K}(\pi_1(S_K), \pi_1(X_K))$$

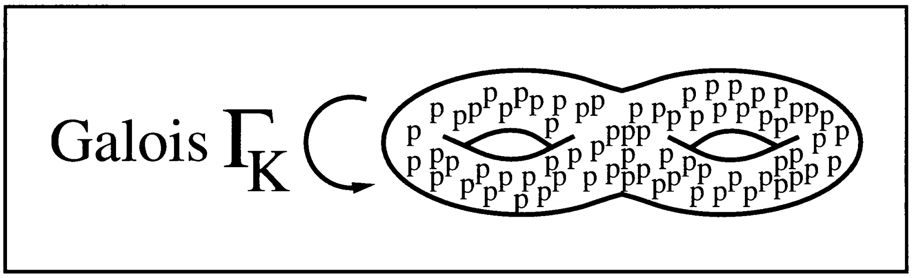
is bijective. Here, the superscripted "dom" denotes dominant ($\Leftrightarrow$ nonconstant) K-morphisms, while $\mathrm{Hom}^{\mathrm{open}}_{\Gamma_K}$ denotes open, continuous homomorphisms compatible with the augmentations to Γ_K, and considered up to composition with an inner automorphism arising from $\pi_1(X_{\overline{K}})$.

Note that this result constitutes an analogue of the "physical aspect" of the Fuchsian uniformization, i.e., it exhibits the *scheme* X_K (in the sense of the functor defined by considering (nonconstant) K-

morphisms from arbitrary smooth S_K to X_K) as equivalent to the "physical/analytic object"

$$\mathrm{Hom}^{\mathrm{open}}_{\Gamma_K}(-, \pi_1(X_K))$$

defined by the topological $\pi_1(X_{\overline{K}})$ together with some additional canonical arithmetic structure (i.e., $\pi_1(X_K)$). This sort of equivalence is depicted in the following illustration, which is meant to remind the reader of the first illustration in §2, (A):



In fact, various slightly stronger versions of Theorem 1 hold. For instance, instead of the whole geometric fundamental group $\pi_1(X_{\overline{K}})$, it suffices to consider its maximal pro-p quotient $\pi_1(X_{\overline{K}})^{(p)}$. Indeed, this is natural relative to the general philosophy discussed in §1, (B) – i.e., one typically expects that the right-hand side of the equivalence of a p-adic Hodge theory should only involve the *pro-p* étale topology. Another strengthening allows one to prove the following result (cf. Theorem B of [Mzk5]), which generalizes a result of Pop ([Pop]):

Theorem 2. *Let p be a prime number. Let K be a subfield of a finitely generated field extension of $\mathbf{Q}_p$. Let L and M be function fields of arbitrary dimension over K. Then the natural map*

$$\mathrm{Hom}_K(\mathrm{Spec}(L), \mathrm{Spec}(M)) \longrightarrow \mathrm{Hom}^{\mathrm{open}}_{\Gamma_K}(\Gamma_L, \Gamma_M)$$

is bijective. Here, $\mathrm{Hom}^{\mathrm{open}}_{\Gamma_K}(\Gamma_L, \Gamma_M)$ *is the set of open, continuous group homomorphisms* $\Gamma_L \to \Gamma_M$ *over* Γ_K*, considered up to composition with an inner homomorphism arising from* $\mathrm{Ker}(\Gamma_M \to \Gamma_K)$.

As discussed in §2, (A), the proof of Theorem 1 consists of implementing the ideas discussed in §2, (A), in the p-adic case by using p-adic Hodge theory.

More precisely, suppose that one is given two *hyperbolic curves* X_K, Y_K over K. For simplicity, let us assume that both X_K and Y_K are both *proper* and *non-hyperelliptic*, and that K is a *finite* extension of $\mathbf{Q}_p$. Suppose, moreover, that we are given an isomorphism

$$\alpha : \pi_1(X_K) \cong \pi_1(Y_K)$$

of the respective arithmetic fundamental groups which is compatible with the projections to Γ_K. Then Theorem 1 states that α necessarily *arises geometrically*, i.e., from a K-isomorphism $X_K \cong Y_K$. In the following, we would like to give a rough sketch of the ideas used to prove this result.

First, observe that α induces an isomorphism

$$\pi_1^{(p)}(X_{\overline{K}})^{\mathrm{ab}} \cong \pi_1^{(p)}(Y_{\overline{K}})^{\mathrm{ab}}$$

between the abelianizations of the maximal pro-p quotients of the respective geometric fundamental groups. Then it follows from p-adic Hodge theory that if one tensors this isomorphism with $\mathbf{C}_p$ (i.e, the p-adic completion of $\overline{K}$), and then takes Γ_K-invariants, one obtains (naturally) on both sides the respective spaces of global differentials, $D_X \stackrel{\mathrm{def}}{=} H^0(X_K, \omega_{X_K})$ and $D_Y \stackrel{\mathrm{def}}{=} H^0(Y_K, \omega_{Y_K})$. Thus, one obtains an isomorphism

$$D_X \cong D_Y$$

induced by α. Let $P_X \stackrel{\mathrm{def}}{=} \mathbf{P}(D_X)$, $P_Y \stackrel{\mathrm{def}}{=} \mathbf{P}(D_Y)$ be the corresponding projective spaces. Thus, one obtains an isomorphism $P_X \cong P_Y$. On the other hand, since we assumed that X_K and Y_K are *non-hyperelliptic*, it follows from elementary algebraic geometry that we have canonical embeddings $X_K \subseteq P_X$, $Y_K \subseteq P_Y$. In other words, we have a diagram:

$$\begin{array}{ccc} P_X & \cong & P_Y \\ \cup & & \cup \\ X_K & \xrightarrow{?} & Y_K \end{array}$$

Thus, the problem of constructing an isomorphism $X_K \cong Y_K$ as desired is reduced to showing that the isomorphism $P_X \cong P_Y$ that we have already constructed maps X_K into Y_K. This is proven precisely by considering certain *p-adic analytic representations* of the differentials of D_X and D_Y as differentials on a certain p-adic space (= the spectrum of a certain large p-adic field) in a fashion reminiscent of the way in which analytic representations (i.e., automorphic forms) of differential forms appeared in the above discussion of the complex case. We refer to [Mzk5], [NTM], for more details.

§3. The Modular Aspect: Canonical Frobenius Actions

(A) The Complex Case

In this section, we discuss the ideas in the complex case that form the philosophical background underlying the theory of [Mzk1], [Mzk2],

[Mzk3], [Mzk4]. Just as in §2, the key phrase was "embedding by automorphic forms", in the present discussion, the key phrase is *"canonical Frobenius actions"*.

First, let us observe that the Fuchsian uniformization of the Riemann surface $\mathcal{X}$ associated to a hyperbolic algebraic curve X gives rise to an action of $\pi_1(\mathcal{X})$ on $\mathfrak{H}$, hence defines a *canonical representation*

$$\rho_{\mathcal{X}} : \pi_1(\mathcal{X}) \longrightarrow PSL_2(\mathbf{R}) \stackrel{\text{def}}{=} SL_2(\mathbf{R})/\{\pm 1\} = \mathrm{Aut}_{\text{Holomorphic}}(\mathfrak{H}).$$

Note that $\rho_{\mathcal{X}}$ may also be regarded as a representation into $PGL_2(\mathbf{C}) = GL_2(\mathbf{C})/\mathbf{C}^\times$, hence as defining an action of $\pi_1(\mathcal{X})$ on $\mathbf{P}^1_{\mathbf{C}}$. Taking the quotient of $\mathfrak{H} \times \mathbf{P}^1_{\mathbf{C}}$ by the action of $\pi_1(\mathcal{X})$ on both factors then gives rise to a projective bundle with connection on $\mathcal{X}$. It is immediate that this projective bundle and connection may be algebraized, so we thus obtain a projective bundle and connection $(P \to X, \nabla_P)$ on X. This pair (P, ∇_P) has certain properties which make it an *indigenous bundle* (terminology due to Gunning).

In general, the notion of an "indigenous bundle on $\mathcal{X}$" may be thought of as the datum of a *projective structure* on $\mathcal{X}$, i.e., a subsheaf of the sheaf of holomorphic functions on $\mathcal{X}$ such that locally any two sections of this subsheaf are related by a linear fractional transformation (with constant coefficients). Thus, the Fuchsian uniformization defines a special *canonical indigenous bundle*, or *canonical projective structure*, on X.

In fact, it is not difficult to see that the notion of an indigenous bundle is entirely *algebraic*. Thus, one has a natural moduli stack

$$\mathcal{S}_{g,r} \longrightarrow \mathcal{M}_{g,r}$$

of hyperbolic curves of type (g, r) equipped with an indigenous bundle, which forms a torsor (under the affine group given by the sheaf of differentials $\Omega_{\mathcal{M}_{g,r}}$ on $\mathcal{M}_{g,r}$) – called the *Schwarz torsor* – over the *moduli stack* $\mathcal{M}_{g,r}$ of hyperbolic curves of type (g, r). Moreover, $\mathcal{S}_{g,r}$ is not only algebraic, it is defined over $\mathbf{Z}[\frac{1}{2}]$. Thus, if, for instance, X is a hyperbolic curve over $\mathbf{C}$, the space of indigenous bundles (or equivalently, of projective structures) on X is a complex affine linear space of dimension $3g - 3 + r$. In particular, (in general) X admits many more indigenous bundles than the canonical one arising from the Fuchsian uniformization.

The canonical indigenous bundle defines a *canonical real analytic section*

$$s : \mathcal{M}_{g,r}(\mathbf{C}) \longrightarrow \mathcal{S}_{g,r}(\mathbf{C})$$

of the Schwarz torsor at the infinite prime. Moreover, not only does s "contain" all the information that one needs to define the Fuchsian uniformization of an individual hyperbolic curve (indeed, this much is obvious from the definition of s!), it also essentially "is" (interpreted properly) the *Bers uniformization* of the universal covering space (i.e., "Teichmüller space") of $\mathcal{M}_{g,r}(\mathbf{C})$. More precisely,

(1) $\overline{\partial} s$ is equal to the *Weil-Petersson metric* (a natural real analytic Kähler metric) on $\mathcal{M}_{g,r}(\mathbf{C})$.

(2) In general, real analytic Kähler metrics may be integrated locally to form *canonical* (*holomorphic*) *coordinates* on the given complex manifold. If one applies this general theory to the Weil-Petersson metric, one obtains the *Bers coordinates* (i.e., the coordinates arising from the Bers uniformization). That is to say, (cf. (1) above), the Bers uniformization may thought of as being precisely the "$\overline{z}$-part" or "*anti-holomorphic part*" of the canonical real analytic section s.

(cf. the discussions in the Introductions of [Mzk1], [Mzk4] for more details). In short, the study of this canonical section s may be regarded as the realization of the Fuchsian uniformization as a *modular IHT*.

Alternatively, from the point of view of classical Teichmüller theory, one may regard the uniformization theory of the moduli of hyperbolic curves as the theory of (so-called) *quasi-fuchsian deformations of the representation* $\rho_{\mathcal{X}}$. Briefly summarized, this point of view runs as follows: Inside $\mathcal{S}_{g,r}(\mathbf{C})$, then is an *open subset* consisting of projective structures defined by certain *quasi-fuchsian groups*. We denote this open subset by

$$\mathrm{Rep}^{\mathrm{QF}}(\pi_1(\mathcal{X}), PGL_2(\mathbf{C})) \subseteq \mathcal{S}_{g,r}(\mathbf{C}).$$

That is to say, this subset parametrizes representations $\pi_1(\mathcal{X}) \to PGL_2(\mathbf{C})$ that arise from Bers' *simultaneous uniformizations* of pairs of hyperbolic Riemann surfaces of type (g, r).

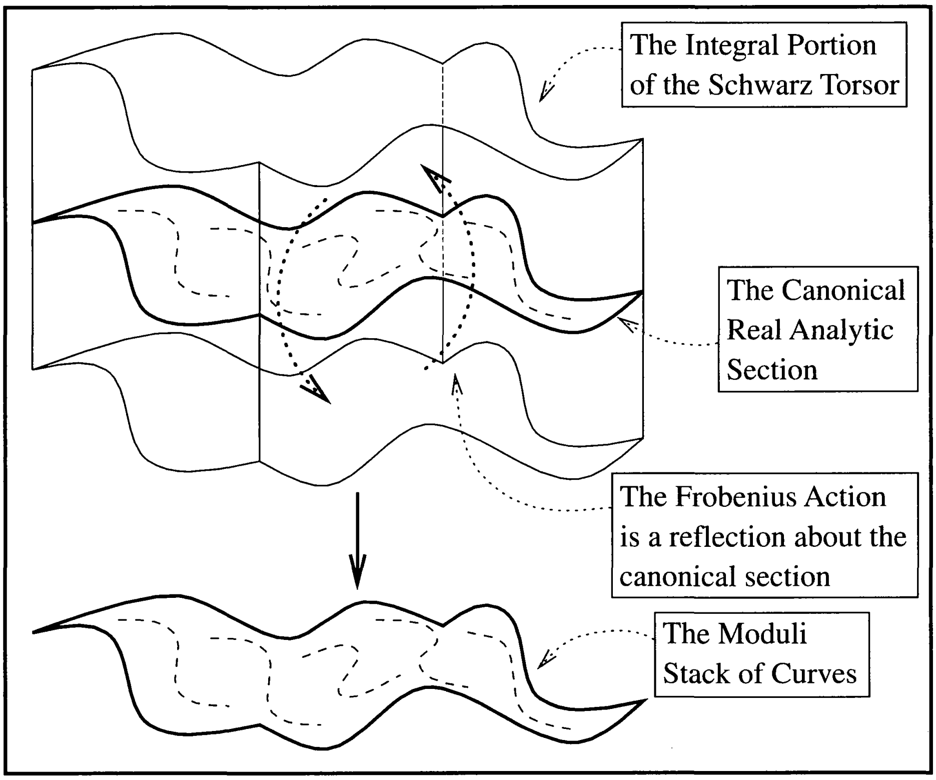
On each of the fibers of $\mathcal{S}_{g,r}(\mathbf{C}) \to \mathcal{M}_{g,r}(\mathbf{C})$ (which are complex affine spaces of dimension $3g - 3 + r$), this open subset is a *bounded contractible subset* of the complex affine space which forms the fiber. From the point of view of Arakelov theory, it is natural to regard such a bounded contractible subset as an *integral stucture* (at the infinite prime) on the complex affine space. Thus, we shall also write

$$\mathcal{S}_{g,r}^{\mathrm{int}_\infty} \stackrel{\mathrm{def}}{=} \mathrm{Rep}^{\mathrm{QF}}(\pi_1(\mathcal{X}), PGL_2(\mathbf{C}))$$

for $\mathrm{Rep}^{\mathrm{QF}}(\pi_1(\mathcal{X}), PGL_2(\mathbf{C}))$.

Next, let us observe that there is a natural action of *complex conjugation* – which we would like to think of as an action of *the Frobenius* Fr_∞ *at the infinite prime* – on $\mathrm{Rep}^{\mathrm{QF}}(\pi_1(\mathcal{X}), PGL_2(\mathbf{C}))$ induced by the action of complex conjugation on the various components of the (equivalence classes of) matrices which form $PGL_2(\mathbf{C})$. Relative to this action of Fr_∞, it is not difficult to see that the image of the canonical real analytic section $s : \mathcal{M}_{g,r}(\mathbf{C}) \to \mathcal{S}_{g,r}(\mathbf{C})$ considered above, i.e., the subset of the set of quasi-fuchsian groups consisting of the *Fuchsian groups*, is precisely equal to the set of Fr_∞-invariants of $\mathcal{S}^{\mathrm{int}_\infty}_{g,r} = \mathrm{Rep}^{\mathrm{QF}}(\pi_1(\mathcal{X}), PGL_2(\mathbf{C}))$. That is to say, we have a natural commutative diagram:

$$\begin{array}{ccc}
\text{`}\mathcal{S}^{\mathrm{int}_\infty}_{g,r}\text{'} \curvearrowleft \mathrm{Fr}_\infty & & \\
\| & & \\
\mathrm{Rep}^{\mathrm{QF}}(\pi_1(\mathcal{X}), PGL_2(\mathbf{C})) & \stackrel{\text{open}}{\hookrightarrow} & \mathcal{S}_{g,r}(\mathbf{C}) \\
\cup & & \downarrow \\
\mathrm{Fr}_\infty\text{-invariants} = \mathrm{Image}(s) & \stackrel{\sim}{\longrightarrow} & \mathcal{M}_{g,r}(\mathbf{C})
\end{array}$$



Moreover, relative to this point of view, the Bers uniformization is the uniformization of $\mathcal{M}_{g,r}(\mathbf{C})$ by a fiber of $\mathcal{S}^{\mathrm{int}_\infty}_{g,r}$ given by the formula:

$$\text{Bers uniformization} = \mathrm{pr}_{\mathcal{M}_{g,r}} \circ (\mathrm{Fr}_\infty|_{\mathrm{Fiber}})$$

i.e., the composite morphism

$$\text{Fiber (of } \mathcal{S}_{g,r}^{\text{int}_\infty} \to \mathcal{M}_{g,r}(\mathbf{C})) \hookrightarrow \mathcal{S}_{g,r}^{\text{int}_\infty} \xrightarrow{\text{Fr}_\infty} \mathcal{S}_{g,r}^{\text{int}_\infty} \longrightarrow (\mathcal{M}_{g,r})_{\mathbf{C}}$$

In other words, from this point of view:

Key Point: *"The Bers uniformization is a Frobenius action!"*

Formulated from this point of view, the ideas of classical Teichmüller theory carry over fairly transparently to the p-adic case. This will be the theme of our discussion of *p-adic Teichmüller theory* in the section's to follow.

(B) Teichmüller Theory in Characteristic p

Let p be an *odd* prime. Then the p-adic theory of Shimura curves (cf., e.g., [Ihara]) suggests that a natural condition to expect of canonical indigenous bundles in characteristic p is that they should have *square nilpotent p-curvature.* The "p-curvature" is a natural invariant of bundles with connection in characteristic p, which, philosophically, may be thought of as a measure of the extent to which the connection ∇ is compatible with (i.e., "commutes with") Frobenius, i.e.:

$$\underline{p\text{-curvature}} = \text{"[Frobenius, } \nabla]\text{"}.$$

In the case, of $\mathbf{P}^1$-bundles, the p-curvature may be thought of as a 2×2-matrix of differentials (conjugated by Frobenius) whose trace is zero. Thus, to say that the p-curvature is "square nilpotent" means simply that the square (in the sense of ordinary matrix multiplication) of this 2×2-matrix is zero.

Let $\mathcal{N}_{g,r} \subseteq (\mathcal{S}_{g,r})_{\mathbf{F}_p}$ denote the closed algebraic substack of indigenous bundles with square nilpotent p-curvature. Then one has the following key result ([Mzk1, Chapter II, Theorem 2.3]):

Theorem 3. *The natural map $\mathcal{N}_{g,r} \to (\mathcal{M}_{g,r})_{\mathbf{F}_p}$ is a finite, flat, local complete intersection morphism of degree p^{3g-3+r}. Thus, up to "isogeny" (i.e., up to the fact that this degree is not equal to one), $\mathcal{N}_{g,r}$ defines a canonical section of the Schwarz torsor $(\mathcal{S}_{g,r})_{\mathbf{F}_p} \to (\mathcal{M}_{g,r})_{\mathbf{F}_p}$ in characteristic p, i.e., it gives rise to a diagram*

$$\begin{array}{ccc} (\mathcal{S}_{g,r})_{\mathbf{F}_p} & = & (\mathcal{S}_{g,r})_{\mathbf{F}_p} \\ \cup & & \downarrow \\ \mathcal{N}_{g,r} & \longrightarrow & (\mathcal{M}_{g,r})_{\mathbf{F}_p} \end{array}$$

reminiscent of the diagram appearing in §3, (A).

It is this stack $\mathcal{N}_{g,r}$ of *nilcurves* – i.e., hyperbolic curves in characteristic p equipped with an indigenous bundle with square nilpotent p-curvature – which is the central object of study in the *p-adic Teichmüller theory* of [Mzk1], [Mzk2], [Mzk3], [Mzk4].

Many facts are now known about the *finer structure of* $\mathcal{N}_{g,r}$. One interesting consequence of this structure theory of $\mathcal{N}_{g,r}$ is that *it gives rise to a new proof of the connectedness of* $(\mathcal{M}_{g,r})_{\mathbf{F}_p}$ (for p large relative to g). This fact is interesting – relative to the claim that this theory is a p-adic version of Teichmüller theory – in that one of the first applications of classical complex Teichmüller theory is to prove the connectedness of $\mathcal{M}_{g,r}$. Also, it is interesting to note that F. Oort has succeeded in giving a proof of the connnectedness of the moduli stack of principally polarized abelian varieties by applying the structure theory of certain natural substacks of this moduli stack in characteristic p.

(C) p-adic Teichmüller Theory

So far, we have been discussing the characteristic p theory. Ultimately, however, we would like to know if the various characteristic p objects discussed in §3, (B), *lift canonically* to objects which are flat over $\mathbf{Z}_p$. Unfortunately, it seems that it is unlikely that $\mathcal{N}_{g,r}$ itself lifts canonically to some sort of natural $\mathbf{Z}_p$-flat object. If, however, we consider the open substack – called the *ordinary locus* – $(\mathcal{N}_{g,r}^{\mathrm{ord}})_{\mathbf{F}_p} \subseteq \mathcal{N}_{g,r}$ which is the étale locus of the morphism $\mathcal{N}_{g,r} \to (\mathcal{M}_{g,r})_{\mathbf{F}_p}$, then (since the étale site is invariant under nilpotent thickenings) we get a canonical lifting, i.e., an étale morphism

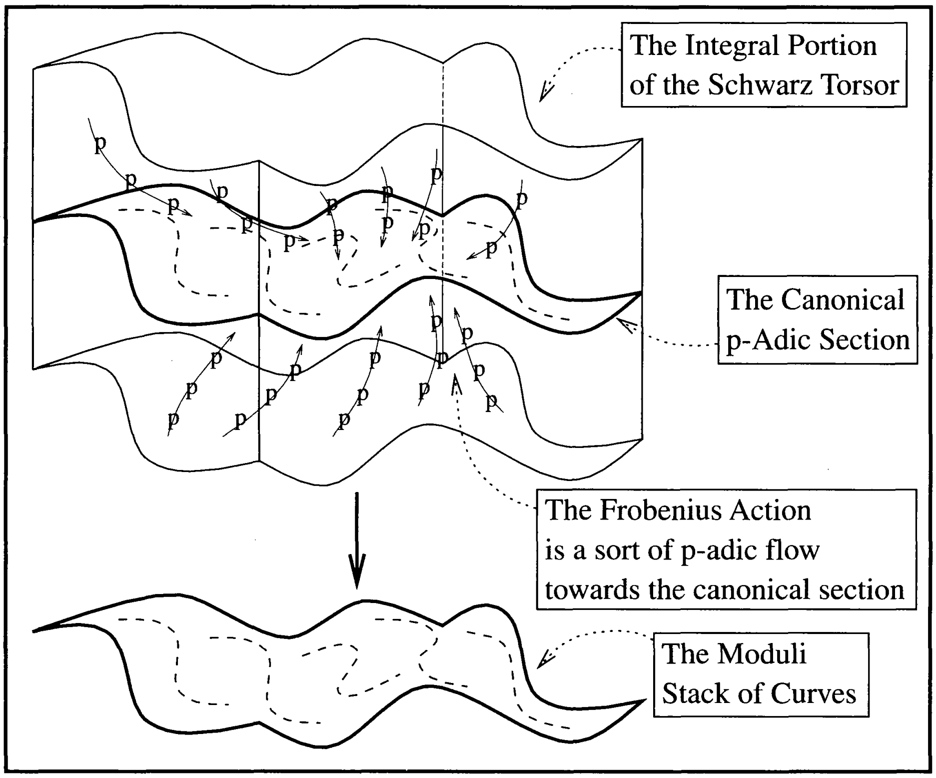
$$\mathcal{N}_{g,r}^{\mathrm{ord}} \longrightarrow (\mathcal{M}_{g,r})_{\mathbf{Z}_p}$$

of p-adic formal stacks. Over $\mathcal{N}_{g,r}^{\mathrm{ord}}$, one has the sought-after *canonical p-adic splitting of the Schwarz torsor*, i.e., the p-adic analogue of the canonical real analytic splitting $s : \mathcal{M}_{g,r}(\mathbf{C}) \to \mathcal{S}_{g,r}(\mathbf{C})$ discussed in §3, (A) (cf. Theorem 0.1 of the Introduction of [Mzk1]; [Mzk4, Chapter X, §3]):

Theorem 4. *There is a canonical section* $\mathcal{N}_{g,r}^{\mathrm{ord}} \to \mathcal{S}_{g,r}$ *of the Schwarz torsor over* $\mathcal{N}_{g,r}^{\mathrm{ord}}$ *which is the unique section having the following property:* $^{\exists}$*a lifting of Frobenius* $\Phi_{\mathcal{N}} : \mathcal{N}_{g,r}^{\mathrm{ord}} \to \mathcal{N}_{g,r}^{\mathrm{ord}}$ *such that the indigenous bundle on the tautological hyperbolic curve over* $\mathcal{N}_{g,r}^{\mathrm{ord}}$ *defined by the section* $\mathcal{N}_{g,r}^{\mathrm{ord}} \to \mathcal{S}_{g,r}$ *is* invariant *with respect to the Frobenius* *action defined by $\Phi_{\mathcal{N}}$. Moreover, this canonical section and canonical Frobenius lifting give rise to a diagram*

$$\begin{array}{ccc}
(\mathcal{S}_{g,r})_{\mathbf{Z}_p} & = & (\mathcal{S}_{g,r})_{\mathbf{Z}_p} \\
\cup & & \downarrow \\
\Phi_{\mathcal{N}} \curvearrowright (\mathcal{N}^{\mathrm{ord}}_{g,r})_{\mathbf{Z}_p} & \stackrel{p\text{-adic et}}{\longrightarrow} & (\mathcal{M}_{g,r})_{\mathbf{Z}_p} \\
\cdot\|\cdot & & \\
\mathrm{Rep}^{\mathrm{Crys}}(\pi_1(X_{\mathbf{Q}_p}), PGL_2(\mathbf{Z}_p)) & &
\end{array}$$

reminiscent of the diagram appearing in §3, (A). *Here, "$\cdot\|\cdot$" means "roughly may be identified with", and "*$\mathrm{Rep}^{\mathrm{Crys}}(\pi_1(X_{\mathbf{Q}_p}), PGL_2(\mathbf{Z}_p))$*" is a certain natural space of crystalline respresentations of the arithmetic fundamental group of the tautological hyperbolic curve into $PGL_2(\mathbf{Z}_p)$. Finally, the resulting action of $\Phi_{\mathcal{N}}$ on "*$\mathrm{Rep}^{\mathrm{Crys}}(\pi_1(X_{\mathbf{Q}_p}), PGL_2(\mathbf{Z}_p))$*" is the natural Frobenius action on this space of crystalline representations* (*cf.* [Mzk4, Chapter X, especially §3], *for more details*).



Next, we observe that the *Frobenius lifting* $\Phi_{\mathcal{N}} : \mathcal{N}^{\mathrm{ord}}_{g,r} \to \mathcal{N}^{\mathrm{ord}}_{g,r}$ (i.e., morphism whose reduction modulo p is the Frobenius morphism) has the special property that $\frac{1}{p} \cdot \mathrm{d}\Phi_{\mathcal{N}}$ induces an isomorphism

$$\Phi^*_{\mathcal{N}} \Omega_{\mathcal{N}^{\mathrm{ord}}_{g,r}} \cong \Omega_{\mathcal{N}^{\mathrm{ord}}_{g,r}}$$

Such a Frobenius lifting is called *ordinary*. It turns out that *any* ordinary Frobenius lifting (i.e., not just $\Phi_{\mathcal{N}}$) defines (by integration) a set of *canonical multiplicative coordinates* in a formal neighborhood of any point α valued in an algebraically closed field k of characteristic p, as well as a *canonical lifting* of α to a point valued in $W(k)$ (Witt vectors with coefficients in k).

Moreover, there is a certain analogy between this general theory of ordinary Frobenius liftings and the theory of *real analytic Kähler metrics* (which also define canonical coordinates by integration). Relative to this analogy, the canonical Frobenius lifting $\Phi_{\mathcal{N}}$ on $\mathcal{N}_{g,r}^{\mathrm{ord}}$ may be regarded as corresponding to the *Weil-Petersson metric* on complex Teichmüller space (a metric whose canonical coordinates are the coordinates arising from the Bers uniformization of Teichmüller space – cf. the discussion of §3, (A)), i.e.,

$$\Phi_{\mathcal{N}} \longleftrightarrow \text{Weil-Petersson metric,}$$
$$\text{`}\int \Phi_{\mathcal{N}}\text{'} \longleftrightarrow \text{Bers coordinates.}$$

Thus, $\Phi_{\mathcal{N}}$ is, in a very real sense, a p-adic analogue of the Bers uniformization in the complex case. Moreover, there is, in fact, a canonical ordinary Frobenius lifting on the "ordinary locus" of the tautological curve over $\mathcal{N}_{g,r}^{\mathrm{ord}}$ whose relative canonical coordinate is analogous to the canonical coordinates arising from the Köbe uniformization of a hyperbolic curve (i.e., from the canonical real analytic Kähler metric obtained by descending the Poincaré metric on $\mathfrak{H}$ via the Köbe uniformization $\mathfrak{H} \to \mathcal{X}$).

Next, we observe that Serre-Tate theory for ordinary (principally polarized) abelian varieties *may also be formulated as arising from a certain canonical ordinary Frobenius lifting.* Thus, the Serre-Tate parameters (respectively, Serre-Tate canonical lifting) may be identified with the canonical multiplicative parameters (respectively, canonical lifting to the Witt vectors) of this ordinary Frobenius lifting. That is to say, *in a very concrete and rigorous sense, Theorem 4 may be regarded as the analogue of Serre-Tate theory for hyperbolic curves.* Nevertheless, we remark that it is *not* the case that the condition that a nilcurve be ordinary (i.e., define a point of $(\mathcal{N}_{g,r}^{\mathrm{ord}})_{\mathbf{F}_p} \subseteq \mathcal{N}_{g,r}$) either implies or is implied by the condition that its Jacobian be ordinary.

Although this fact may disappoint some readers, it is in fact very natural when viewed relative to the general analogy between ordinary Frobenius liftings and real analytic Kähler metrics discussed above. Indeed, relative to this analogy, we see that it corresponds to the fact that, when one equips $\mathcal{M}_g$ with the Weil-Petersson metric and $\mathcal{A}_g$ (the moduli stack of principally polarized abelian varieties) with its natural metric arising from the Siegel upper half-plane uniformization, *the Torelli map* $\mathcal{M}_g \to \mathcal{A}_g$ *is not isometric.*

References

[BK] S. Bloch and K. Kato, L-Functions and Tamagawa Numbers of Motives, in "The Grothendieck Festschrift", Volume I, Birkhäuser, 1990, pp. 333–400.

[Falt1] G. Faltings, Endlichkeitssätze für Abelschen Varietäten über Zahlkörpern, Inv. Math., **73** (1983), 349–366.

[Falt2] G. Faltings, p-adic Hodge Theory, Journal of the Amer. Math. Soc., **1**, No. 1 (1988), 255–299.

[Falt3] G. Faltings, Crystalline Cohomology and p-adic Galois Representations, JAMI Conference, Johns Hopkins Univ. Press (1990), 25–79.

[FC] G. Faltings and C.-L. Chai, "Degenerations of Abelian Varieties", Springer Verlag, 1990.

[Ihara] Y. Ihara, On the Differentials Associated to Congruence Relations and the Schwarzian Equations Defining Uniformizations, Jour. Fac. Sci. Univ. Tokyo, Sect. IA Math., **21** (1974), 309–332.

[LS] P. Lochak and L. Schneps, "Geometric Galois Actions: 1. Around Grothendieck's Esquisse d'un Programme", London Math. Soc. Lect. Note Ser. **242**, Cambridge Univ. Press, 1997.

[Mumf] D. Mumford, An Analytic Construction of Degenerating Curves over Complete Local Rings, Compositio Math., **24** (1972), 129–174.

[Mzk1] S. Mochizuki, A Theory of Ordinary p-adic Curves, RIMS Preprint 1033 (September 1995), Publ. of RIMS, **32**, No. 6 (1996), 957–1151.

[Mzk2] S. Mochizuki, The Generalized Ordinary Moduli of p-adic Hyperbolic Curves, RIMS Preprint 1051 (December 1995), p. 281.

[Mzk3] S. Mochizuki, Combinatorialization of p-adic Teichmüller Theory, RIMS Preprint 1076 (April 1996), p. 32.

[Mzk4] S. Mochizuki, "Foundations of p-adic Teichmüller Theory", to appear (as a book) from International Press.

[Mzk5] S. Mochizuki, The Local Pro-p Anabelian Geometry of Curves, RIMS Preprint 1097 (August 1996), p. 84; to appear in Inv. Math..

[NTM] H. Nakamura, A. Tamagawa, and S. Mochizuki, The Grothendieck Conjecture on the Fundamental Groups of Algebraic Curves, Sugaku, **50**, No. 2 (1998), 1–17. (in Japanese).

[Pop] F. Pop, On Grothendieck's conjecture of birational anabelian geometry II, Preprint (1995).

[SGA1] A. Grothendieck et al., "Revêtements étales et groupe fondamental", Lecture Notes in Mathematics, **224**, Springer, 1971.

[Tama1] A. Tamagawa, The Grothendieck conjecture for affine curves, Compositio Math., **109**, No. 2 (1997), 135–194.

[Tama2] A. Tamagawa, On the fundamental groups of curves over algebraically closed fields of characteristic > 0, RIMS Preprint 1182 (January 1998).

[Tate] J. Tate, p-Divisible Groups, Proceedings of a Conference on Local Fields, Driebergen, Springer Verlag (1967), 158–183.

Research Institute for Mathematical Sciences
Kyoto University
Sakyo-ku
Kyoto, 606-8502
Japan
motizuki@kurims.kyoto-u.ac.jp

Advanced Studies in Pure Mathematics 31, 2001
Taniguchi Conference on Mathematics Nara '98
pp. 235–263

Operator Algebras, Topology and Subgroups of Quantum Symmetry
– Construction of Subgroups of Quantum Groups –

Adrian Ocneanu

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

This article is not available due to permission restrictions.

Advanced Studies in Pure Mathematics 31, 2001
Taniguchi Conference on Mathematics Nara '98
pp. 265–286

Large Deviation and Hydrodynamic Scaling

Srinivasa R. S. Varadhan

§1. What are Large Deviations?

The theory of large deviations is, roughly speaking, a method of describing the rapidity with which probability distributions depending on a parameter approach the degenerate distribution at some point as the parameter becomes large.

Let us suppose that for each n there is a probabilty measure P_n on some space Ω_n defined on some σ-field Σ_n. There is a complete separable metric space X with its Borel sets $\mathcal{B}$, such that for each n there is a measurable map Φ_n of Ω_n into X. We denote the induced measure $P_n \Phi_n^{-1}$ on $(X, \mathcal{B})$ by Q_n. Actually it is the situation $(X, \mathcal{B}, Q_n)$ that will be of interest to us. As $n \to \infty$ the measures Q_n will converge weakly to a limit which will be degenrate at some point x_0 of X. This is usually a 'law of large numbers', statement. In particular for any closed set $A \subset X$, with $x_0 \notin A$,

$$\lim_{n \to \infty} Q_n(A) = 0. \tag{1.1}$$

If the parametrization has been chosen properly, the convergence in the limit (1.1) will often be exponentially fast and

$$\lim_{n \to \infty} \frac{1}{n} \log Q_n(A) = -\Psi(A) \tag{1.2}$$

will exist atleast for a large class nice sets. Since the exponential behavior of a sum is the same as that of the larger of the summands

$$\Psi(A \cup B) = \min\{\Psi(A), \Psi(B)\}$$

and one can expect $\Psi(A)$ to be given by a formula of the type

$$\Psi(A) = \inf_{x \in A} I(x)$$

Received December 2, 1999.

for some function $I(\cdot) : X \to [0, \infty]$. The theory of large deviations is a large collection of interesting examples that fit this model. What makes the class of models interesting is the ability to identify the rate function $I(\cdot)$ in specific cases.

We say that the family Q_n on X satisfies the 'large deviation principle' with rate function $I(\cdot)$ if

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n(A) \le -\inf_{x\in A} I(x) \quad \text{for closed sets } A \in X, \tag{1.3}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n(G) \ge -\inf_{x\in G} I(x) \quad \text{for open sets } G \in X. \tag{1.4}$$

Of course if $E \subset X$ is nice enough that

$$\inf_{x\in E^0} I(x) = \inf_{x\in E} I(x) = \inf_{x\in \overline{E}} I(x)$$

we get

$$\lim_{n\to\infty} \frac{1}{n} \log Q_n(E) = -\inf_{x\in E} I(x). \tag{1.5}$$

It is important that the function $I(\cdot)$ that can take the value $+\infty$ be lower semi-continuous and have compact level sets, i.e., for each $\ell < \infty$, the set

$$K_\ell = \{x : I(x) \le \ell\} \tag{1.6}$$

be a compact (closed and totally bounded) subset of X.

We will begin with some simple examples.

Example. Let α be a probability measure on R. Let P_n on $\Omega = R^n$ be the product measure $\alpha \times \alpha \times \cdots \times \alpha$. Let Φ_n be the map

$$\Phi_n(x_1, \ldots, x_n) = \frac{x_1 + \cdots + x_n}{n}.$$

The law of large numbers asserts that $Q_n \to \delta_a$ with $a = \int x \, d\alpha$. According to a theorem of Cramér [1] Q_n satisfies a large deviation principle with rate function

$$I(y) = \sup_\sigma [\sigma y - \log M(\sigma)] \tag{1.7}$$

where

$$M(\sigma) = \int e^{\sigma x} \alpha(dx). \tag{1.8}$$

Another example is the following.

Example. Let F be a finite alphabet $\mathcal{A}$ consisting of letters $\{a_1, \ldots, a_k\}$. Let Ω_n consist of words $W = \{x_1, \ldots, x_n\}$ of length n in $\mathcal{A}$. The probabilities $P_n(W)$ are all equal and since there are k^n words of length n

$$P_n(W) = \frac{1}{k^n}$$

for every word $W \in \Omega_n$. X is the space of probability distributions on $\{1, \ldots, k\}$, i.e., $\{p_1, \ldots, p_k : p_i \geq 0 \text{ and } \sum_i p_i = 1\}$. The map Φ_n is the empirical distibution

$$\Phi_n(x_1, \ldots, x_n) = \left\{ \frac{\sum_j \delta_{a_i, x_j}}{n} \right\} = \left\{ \frac{n_i}{n} \right\} \tag{1.9}$$

where n_i is the number of times the letter a_i occurs in the word $W = \{x_1, \ldots, x_n\}$. Again, by the law of large numbers, Q_n converges to $\delta_{\{1/k, \ldots, 1/k\}}$.

$$\begin{aligned} Q_n[p_1, \ldots, p_n] &\simeq \frac{n!}{(np_1)! \cdots (np_k)!} \frac{1}{k^n}, \\ I(p) &= -\lim_{n \to \infty} \frac{1}{n} \log Q_n[p_1, \ldots, p_n] \\ &= \log k + \sum p_i \log p_i \\ &= \sum_i p_i \log \frac{p_i}{\frac{1}{k}}. \end{aligned}$$

A slightly more general form of the example is Sanov's theorem.

Example. If we define

$$P_n(W) = \frac{n!}{n_1! \cdots n_k!} \pi_1^{n_1} \cdots \pi_k^{n_k}$$

we get

$$I_\pi(p) = -\sum_i p_i \log \pi_i + \sum_i p_i \log p_i = \sum_i p_i \log \frac{p_i}{\pi_i}$$

or even more generally

Example. We take $\Omega_n = X \times X \times \cdots \times X$, $P_n = \alpha \times \alpha \times \cdots \times \alpha$ and Φ_n the map of Ω into the space $\mathcal{M}$ of all probability measures on X defined by

$$\Phi_n(x_1, \ldots, x_n) = \frac{\delta_{x_1} + \cdots + \delta_{x_n}}{n}.$$

In this example $I_\alpha(\mu) < \infty$ only if $\mu \ll \alpha$ and

$$\frac{d\mu}{d\alpha} \log \frac{d\mu}{d\alpha} \in L_1(\alpha).$$

Then

$$I_\alpha(\mu) = \int_X \log \frac{d\mu}{d\alpha}\, d\mu = \int_X \frac{d\mu}{d\alpha} \log \frac{d\mu}{d\alpha}\, d\alpha. \tag{1.10}$$

Otherwise $I_\alpha(\mu) = +\infty$.

There are some general principles in the theory which are relatively easy to establish. Here is one. A general property known as 'Contarction Principle' is the following:

Theorem 1.1. *Let P_n satisfy the large deviation property with rate function $I(\cdot)$ on X. Let $f : X \to Y$ be a continuous map into Y. Then $Q_n = P_n f^{-1}$ satifies a large deviation principle on Y with rate function $J(y) = \inf[I(x)\ ;\ x : f(x) = y]$.*

We will illustrate the Contarction Principle by showing that Cramér's theorem can be obtained from Sanov's theorem. Consider the map $f : \mathcal{M} \to R$ defined by

$$f(\mu) = \int_R x\, d\mu. \tag{1.11}$$

Then the sample mean $(x_1 + \cdots + x_n)/n$ can be thought of as

$$\frac{x_1 + \cdots + x_n}{n} = f(\Phi_n(x_1, \ldots, x_n))$$

where Φ_n is the empirical distribution. A calculation shows that

$$I(y) = \inf_{\mu : \int x\, d\mu = y} I_\alpha(\mu)$$

which is the contraction principle. Actually Sanov's theorem can be sort of seen as a version of Cramér's theorem as well. We can repalce R by the locally convex topological vector space $\mathcal{M}(R)$ and replace α on R by the distribution β induced on $\mathcal{M}(R)$ by the map $x \to \delta_x$. The empirical distribution is just the sum of n independent $\mathcal{M}(R)$ valued random vectors with the common distribution β. The moment generating function is replaced by

$$M(V) = \int_{\mathcal{M}(R)} e^{\langle V, \mu \rangle}\, d\beta = \int_R e^{V(x)}\, d\alpha \tag{1.12}$$

and

$$H(\mu;\alpha) = I_\alpha(\mu) = \sup_{V(\,\cdot\,)}\left[\int_R V(x)\,d\mu - \log M(V)\right] \tag{1.13}$$
$$= \int_R \frac{d\mu}{d\alpha}\log\frac{d\mu}{d\alpha}\,d\alpha.$$

In particular

$$\int_R V(x)\,d\mu \le \log\int_R e^{V(x)}\,d\alpha + H(\mu;\alpha) \tag{1.14}$$

or for any $\sigma > 0$, replacing V by σV, we get

$$\int_R V(x)\,d\mu \le \frac{1}{\sigma}\log\int_R e^{\sigma V(x)}\,d\alpha + \frac{1}{\sigma}H(\mu;\alpha). \tag{1.15}$$

Another general principle is the following theorem on the exponential growth rate of integrals. It is basically a fancy version of the simple fact that for $a, b > 0$ we have

$$\lim_{n\to\infty}[a^n + b^n]^{1/n} = \max(a,b).$$

Theorem 1.2. *If the large deviation principle holds for some Q_n on X, with a rate function $I(\,\cdot\,)$, then for any real valued bounded continuous function $F(\,\cdot\,)$ on X*

$$\lim_{n\to\infty}\frac{1}{n}\log\int_X \exp[n\,F(x)]\,dQ_n = \sup_{x\in X}[F(x) - I(x)].$$

The book [2] is a good source for a discussion of these topics as well as for additional references.

§2. Hydrodynamic Scaling

The basic example of Hydrodynamic scaling is the derivation of Euler equations from the equations of classical mechanics. Let us start with a collection of $N \simeq \bar\rho\ell^3$ classical particles in a large periodic cube Λ_ℓ of side ℓ in R^3. The motion of the particles are governed by the equations of motion of a classical Hamiltonian dynamical system with energy given by

$$H(p,q) = \frac{1}{2}\sum_{i=1}^{N}\|p_i\|^2 + \frac{1}{2}\sum_{i\neq j}V(q_i - q_j). \tag{2.1}$$

Here, $q_i \in \Lambda_\ell$ is the position of the i-th particle and $p_i \in R^3$ is its velocity. The coordinates $k = 1, 2, 3$ refer to the three components of position or velocity. The repulsive potential $V \geq 0$ is an even function that is not identically zero and has compact support in R^3. The interaction in particular is short range. The classical equations of motion are

$$\frac{dq_i^k}{dt} = \frac{\partial H(p,q)}{\partial p_i^k} = p_i^k, \tag{2.2}$$

$$\frac{dp_i^k}{dt} = -\frac{\partial H(p,q)}{\partial q_i^k} = -\sum_{j=1}^{N} V_k(q_i - q_j), \tag{2.3}$$

where $V_k(q) = \partial V(q)/\partial q^k$ for $k = 1, 2, 3$ are the three components of the gradient of V. The dynamical system has five conserved quantities. The total number N of particles, the total momenta $\sum_{i=1}^N p_i^k$ for $k = 1, 2, 3$ and the total energy $H(p, q)$. The hydrodynamic scaling in this context consists of rescaling space and time by a factor of ℓ. The rescaled space is the unit torus $\mathbf{T}^3$ in 3-dimensions. The macroscopic quantities to be studied correspond to the five conserved quantities. The first one of these is the density, and is measured by a function $\rho(t, x)$ of t and x. For each $\ell < \infty$ it is approximated by $\rho_\ell(t, x)$, defined by

$$\int_{\mathbf{T}^3} J(x)\rho_\ell(t,x)\,dx = \frac{1}{\ell^3}\sum_{i=1}^{N} J\Big(\frac{q_i(\ell t)}{\ell}\Big). \tag{2.4}$$

A straight forward differentiation with respect to t yields

$$\begin{aligned} \frac{d}{dt}\int_{\mathbf{T}^3} J(x)\rho_\ell(t,x)\,dx &= \frac{d}{dt}\frac{1}{\ell^3}\sum_{i=1}^{N} J\Big(\frac{q_i(\ell t)}{\ell}\Big) \\ &= \frac{1}{\ell^3}\sum_{i=1}^{N} (\nabla J)\Big(\frac{q_i(\ell t)}{\ell}\Big)\cdot p_i(\ell t) \\ &\simeq \int_{\mathbf{T}^3} (\nabla J)(x)\cdot \rho_\ell(t,x)u_\ell(t,x)\,dx \end{aligned} \tag{2.5}$$

where $u_\ell(t, x) = u_\ell^k(t, x)$, $k = 1, 2, 3$ are the components of the 'average' velocity of the fluid at the rescaled space time point x, t. This introduces three other macroscopic variables, which represent three coordinates of the momenta that are conserved. We can now write down the first of our five equations

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho u) = 0. \tag{2.6}$$

To derive the next three equations, using a test functions J, we differentiate for $k = 1, 2, 3$

(2.7) $$\frac{d}{dt}\frac{1}{\ell^3}\sum_{i=1}^{N} J\Big(\frac{q_i(\ell t)}{\ell}\Big)p_i^k(\ell t) = \frac{1}{\ell^3}\sum_{i=1}^{N} p_i^k(\ell t)(\nabla J)\Big(\frac{q_i(\ell t)}{\ell}\Big)\cdot p_i(\ell t) - \frac{1}{\ell^2}\sum_{i=1}^{N}\sum_{j=1}^{N} J\Big(\frac{q_i(\ell t)}{\ell}\Big)V_k(q_i(\ell t) - q_j(\ell t)).$$

If we now use the skew-symmetry of $V_k = \partial V/\partial q_k$, we can rewrite the second term of the right hand side of equation (2.7) as

(2.8)
$$\begin{aligned}
&-\frac{1}{2\ell^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(J\Big(\frac{q_i(\ell t)}{\ell}\Big) - J\Big(\frac{q_j(\ell t)}{\ell}\Big)\Big)V_k(q_i(\ell t) - q_j(\ell t))\\
&\simeq -\frac{1}{2\ell^3}\sum_{i=1}^{N}\sum_{j=1}^{N} J_r\Big(\frac{q_i(\ell t)}{\ell}\Big)(q_i^r(\ell t) - q_j^r(\ell t))V_k(q_i(\ell t) - q_j(\ell t))\\
&= \frac{1}{\ell^3}\sum_{i=1}^{N}\sum_{j=1}^{N} J_r\Big(\frac{q_i(\ell t)}{\ell}\Big)\psi_k^r(q_i(\ell t) - q_j(\ell t))
\end{aligned}$$

with
$$\psi_k^r(q) = -\frac{1}{2}q^r V_k(q).$$

The next step is rather mysterious and requires considerable explanation. The quantities
$$\sum_{i=1}^{N} p_i^k p_i^r, \quad \sum_{i,j=1}^{N} \psi_k^r(q_i(t) - q_j(t))$$

are not conserved. They depend on combinations of individual velocities that are not conserved and on spacings between particles both of which change in the microscopic time scale and therefore do so rapidly in the macroscopic time scale. They should therefore be replaced by their space-time averages. By appealing to an 'Ergodic Theorem' they can be replaced by their averages with repect to their equilibrium distributions. The equilibrium 'ensemble' consists of an infinite collection of points $\{p_\alpha, q_\alpha\}$, in the phase space $R^3 \times R^3$. There is a natural five parameter family of measures $\mu_{\rho,u,T}$ that are invariant under spatial translations as well as the Hamiltonian dynamics. The points $\{p_\alpha\}$ are distributed according to a Gibbs Distribution with density ρ and formal interaction energy

$$\frac{1}{2T}\sum_{\alpha,\beta} V(q_\alpha - q_\beta).$$

In other words $\{q_\alpha\}$ is a point process obtained by taking infinite volume limit of $N = \ell^3\rho$ particles distributed in the cube of side ℓ in R^3 according to the joint density

$$\frac{1}{Z}\exp\Big[-\frac{1}{2T}\sum_{1\le i\neq j\le N} V(q_i - q_j)\Big]$$

where Z is the normalization constant. The velocities $\{p_\alpha\}$ are distributed independently of each other as well as of $\{q_\alpha\}$, having a common three dimensional Gaussian distribution with mean u and covariance TI. Assuming that the infinite volume limit exists in a reasonable sense it will be a point process defined as an infinite volume Gibbs measure $\mu_{\rho,T}$. The velocities $\{p_\alpha\}$ will be an independent Gaussian ensemble $\nu_{u,T}$. In the first term the quantities $p_i^k p_i^r$ are replaced by their expectations

$$u^k(t,x)u^r(t,x) + \delta_{k,r}T(t,x)$$

and in the second term $\psi_{k,r}$ are replaced by their expectations that involve the 'pressure' per unit volume in the Gibbs ensemble

$$\mathbf{P}_k^r(\rho,T) = \lim_{\ell\to\infty} E^{\mu_{\rho,T}}\Big\{\frac{1}{\ell^3}\sum_{|q_\alpha|,|q_\beta|\le\ell} \psi_k^r(q_\alpha - q_\beta)\Big\}.$$

This leads to the equation

$$\begin{aligned}(2.9)\qquad &\frac{d}{dt}\int_{\mathbf{T}^3} J(x)u^k(t,x)\,dx \\ &= \int_{\mathbf{T}^3}\sum_{r=1}^{3}\frac{\partial J}{\partial x_r}(x)(u^k(t,x)u^r(t,x) + \delta_{k,r}T(t,x))\,dx \\ &\quad + \int_{\mathbf{T}^3}\sum_{r=1}^{3}\frac{\partial J}{\partial x_r}(x)\mathbf{P}_k^r(\rho(t,x),T(t,x))\,dx.\end{aligned}$$

We now integrate by parts, remove the test function J and obtain from equation (2.9)

$$(2.10)\qquad \frac{d}{dt}(\rho u) + \nabla\cdot(\rho u\otimes u + \rho TI + \mathbf{P}(\rho,T)) = 0.$$

There is an equation of state that expresses the total energy per unit volume e as

$$e(\rho, u, T) = \frac{1}{2}\rho(|u|^2 + 3T) + f(\rho, T) \tag{2.11}$$

where $f(\rho, T)$, the potential energy per unit volume, is given by

$$f(\rho, T) = \lim_{\ell \to \infty} E^{\mu_{\rho,T}} \left\{ \frac{1}{2\ell^3} \sum_{|q_\alpha|, |q_\beta| \le \ell} V(q_\alpha - q_\beta) \right\}.$$

Although we will not derive it, there is a similar equation for $e(t, x)$ that is obtained by differentiating

$$\frac{d}{dt} \frac{1}{2\ell^3} \sum_{i=1}^{N} J\Big(\frac{q_i(\ell t)}{\ell}\Big) \Big[|p_i(\ell t)|^2 + \sum_{j=1}^{N} V(q_i(\ell t) - q_j(\ell t)) \Big]$$

and proceeding in a similar fashion. It looks like

$$\frac{de}{dt} + \nabla \cdot \big[(e + T)u + \mathbf{P}(\rho, T)u\big] = 0. \tag{2.12}$$

The five equations one for density given by equation (2.6), the three for velocities contained in equation (2.10) and finally the enrgy equation (2.12) constitute a first order system of non-linear hyperbolic conservation laws in the six variables $[\rho, u, T, e]$ with one relation between them given by equation (2.11). Given smooth initial data they have local solutions. Rigorous derivation of these equations does not exist.

We have made a basic assumption in the above derivation. If we take a small volume in space around the point (x, t) in macroscopic space-time and blow up the space by a factor of ℓ we will see a bunch of particles with velocities. The positions of these particles will form a point process in a big domain in R^3. The statistics of these points is assumed to be a Gibbs distribution $\mu_{\rho,T}$ corresponding to the density $\rho = \rho(t, x)$ and 'Temperature', $T = T(t, x)$. Given the positions, the velocities are assumed to be mutually independent and have a common Gaussian distribution with mean $u = u(t, x)$ and covariance $TI = T(t, x)I$. The five parameters (ρ, u, T) locally detemine a Gibbs-Gaussian equilibrium. The equations are derived under the assumption that this picture holds asymptotically for large ℓ. There is no known proof of this. While it is possible to prepare the initial state so that this property of loacl equlibrium holds at time $t = 0$, there is no guarantee that this property persists at positive macroscopic times. If ρ, u, T are constants independent of

x at time 0, then we have a global equilibrium and that persists. But hydrodynamically this is the uninteresting case.

While the validity of the principle of local equlibrium is very hard to establish for the Hamiltonian system, it is not nearly so hard for stochastic systems of comparable type. Noise helps to establish local equilibria. This in many cases can be rigorously established and thence the corresponding hydrodynamical equations can be derived with full mathematical rigor.

We will consider a class of stochastic models that are called simple exclusion processes. They make sense on any finite or countable set X and for us it will be either the integer lattice $\mathbf{Z}^d$ in d-dimensions or $\mathbf{Z}^d_N$ obtained from it as a quotient by considering each coordinate modulo N. At any given time a subset of these lattice sites will be occupied by particles, with atmost one particle at each site. In other words some sites are empty while others are occupied with one particle. The particles move randomly. Each particle waits for an exponential random time and then tries to jump from its current site x to a new site y. The new site y is picked randomly according to a probability distribution $\pi(x,y)$. In particular $\sum_y \pi(x,y) = 1$ for every x. Of course a jump to y is not always possible. If the site is empty the jump is possible and is carried out. If the site already has a particle, the jump cannot be carried out and the particle forgets about it and waits for another chance, i.e., waits for a new exponential waiting time. If we normalize so that all waiting times have mean 1, the generator of the process can be written down as

$$(\mathcal{A}f)(\eta) = \sum_{x,y} \eta(x)(1-\eta(y))\pi(x,y)[f(\eta^{x,y}) - f(\eta)] \tag{2.13}$$

where η represents the configuration with $\eta(x) = 1$ if there is a particle at x and $\eta(x) = 0$ otherwise. For each configuration η and a pair of sites x, y the new configuration $\eta^{x,y}$ is defined by

$$\eta^{x,y}(z) = \begin{cases} \eta(y), & \text{if } z = x, \\ \eta(x), & \text{if } z = y, \\ \eta(z), & \text{if } z \neq x, y. \end{cases} \tag{2.14}$$

We will mainly be concerned with the situation where the set X is $\mathbf{Z}^d$ or $\mathbf{Z}^d_N$, viewed naturally as an Abelian group with $\pi(x,y)$ being translation invariant and given by $\pi(x,y) = p(y-x)$ for some probability distribution p. It is convenient to assume that p has finite support. There are various possibilities. We will first consider the case $\sum_z zp(z) = m \neq 0$ that needs hyperbolic scaling and leads to Burgers equation with zero viscosity. In order to convey the idea it is sufficient to restrict ourselves to the case where $d = 1$, $p(1) = 1$ and $p(z) = 0$ for all $z \neq 1$. This is the totally asymmetric nearest neighbor simple exclusion model and of course $m = 1$. In this case the rescaling in time is done by a factor of N and the generator is

$$(\mathcal{A}_N f)(\eta) = N \sum_{x \in \mathbf{Z}_N} \eta(x)(1 - \eta(x+1))[f(\eta^{x,x+1}) - f(\eta)].$$

We can easily calculate

(2.16)

$$\begin{aligned} d\Bigg[\frac{1}{N} \sum_{x \in \mathbf{Z}_N} J\Big(\frac{x}{N}\Big)\eta_t(x)\Bigg] &= \Bigg[\sum_{x \in \mathbf{Z}_N} \Big[J\Big(\frac{x+1}{N}\Big) - J\Big(\frac{x}{N}\Big)\Big]\eta_t(x)(1 - \eta_t(x+1))\Bigg] dt + M_N(t) \\ &\simeq \frac{1}{N}\Bigg[\sum_{x \in \mathbf{Z}_N} J'\Big(\frac{x}{N}\Big)\eta_t(x)(1 - \eta_t(x+1))\Bigg] dt + o(1). \end{aligned}$$

The martingale term is negligible and we need to do simple 'averaging'. The equlibria are the Bernoulli measures μ_ρ indexed by density. Since there is only a single invariant quantity, i.e., the number of particles, we can replace $\eta_t(x)(1-\eta_t(x+1))$ by its expected value $\rho(t,x)(1-\rho(t,x))$.

$$\frac{d}{dt} \int_{\mathbf{T}^1} J(\theta)\rho(t,\theta)\, d\theta = \int_{\mathbf{T}^1} J'(\theta)\rho(t,\theta)(1 - \rho(t,\theta))\, d\theta$$

or equivalently

$$\frac{\partial \rho}{\partial t} + \frac{\partial[\rho(1-\rho)]}{\partial \theta} = 0. \tag{2.17}$$

A different situation occurs when p is symmetric, i.e., $p(x) = p(-x)$. Let us look at the function

$$V_J(\eta) = \sum J(x)\eta(x)$$

and compute

$$(2.18)\qquad \begin{aligned}(\mathcal{A}V_J)(\eta) &= \sum_{x,y} \eta(x)(1-\eta(y))p(y-x)(J(y)-J(x))\\ &= \sum_{x,y} \eta(x)p(y-x)(J(y)-J(x))\\ &= \sum_{x,y} \eta(x)[(\mathbf{P}-I)J](x)\\ &= V_{(\mathbf{P}-I)J}(\eta).\end{aligned}$$

The space of linear functionals is left invariant by the generator. It is not difficult to see that

$$E_\eta\big[V_J(\eta(t))\big] = V_{J(t)}(\eta)$$

where

$$J(t) = \exp[t(\mathbf{P}-I)]J$$

is the solution of

$$\frac{d}{dt}J(t,x) = (\mathbf{P}-I)J(t,x).$$

It is almost as if the interaction had no effect and in fact for the calculation of expectations of 'one particle' functions it clearly does not. Let us start with a configuration on $\mathbf{Z}_N^d$ and scale space by N and time by N^2. The generator becomes $N^2\mathcal{A}$ and the particles can be visualized as moving on a lattice imbedded in the unit torus $\mathbf{T}^d$, with a spacing of $1/N$, that becomes dense as $N \to \infty$.

Let us consider the functional

$$\xi(t) = \frac{1}{N^d}\sum_x J\Big(\frac{x}{N}\Big)\eta_t(x).$$

We can write

$$\xi(t) - \xi(0) = \int_0^t V_N(\eta(s))\,ds + M_N(t)$$

where

$$V_N(\eta) = (N^2\mathcal{A}V_J)(\eta) = V_{J_N}(\eta)$$

with

$$\begin{aligned}(J_N)(x) &= N^2\sum \Big[J\Big(x+\frac{z}{N}\Big) - J(x)\Big]p(z)\\ &\simeq \frac{1}{2}(\Delta_C J)(x).\end{aligned}$$

Here Δ_C refers to the Laplacian

$$\sum_{i,j} C_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}$$

with the covariance matrix C given by

$$C_{i,j} = \sum_x x_i x_j p(x).$$

$M_N(t)$ is a martingale and a very elementary calculation yields

$$E\{[M_N(t)]^2\} \leq CtN^{-d}$$

essentially completing the proof in this case. Technically the empirical distribution $\nu_N(t)$ is viewed as a measure on $\mathbf{T}^d$ and $\nu_N(\,\cdot\,)$ is viewed as a stochastic process with values in the space $\mathcal{M}(\mathbf{T}^d)$ of nonnegative measures on $\mathbf{T}^d$. In the limit it lives on the set of weak solutions of the heat equation

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\Delta_C \rho \tag{2.19}$$

with the initial condition $\rho(0,x) = \rho_0(x)$ determined by

$$\int_{\mathbf{T}^d} J(x)\rho_0(x)\,dx = \lim_{n\to\infty} \frac{1}{N^d} \sum_{x \in \mathbf{Z}_N^d} J\Big(\frac{x}{N}\Big)\eta_0(x) \tag{2.20}$$

and the uniqueness of such weak soultions for given initial density establishes the validity of the scaling limit. We could have have computed two moments as in the noninteracting case. The expectation would have been no different from the noninteracting case since it involves only one particle functions. The variance involves two particle functions and would have involved slighlty more work, because the independence is not there. The martingale argument however is more general.

Let us now turn to the case where p has mean zero but is not symmetric. In this case

$$V_N(\eta) = N^{2-d}\sum_{x,y} \eta(x)(1-\eta(y))p(y-x)\Big[J\Big(\frac{y}{N}\Big) - J\Big(\frac{x}{N}\Big)\Big] \tag{2.21}$$

and we get stuck at this point. If p is symmetric, as we saw, we gain a factor of N^{-2}. Otherwise the gain is only a factor of N^{-1} which is not enough. We seem to end up with

$$\begin{aligned}
N^{-d} & \sum_x \sum_y \eta(x) \\
& \times \left\langle \frac{1}{2}\Big[(\nabla J)\Big(\frac{x}{N}\Big) + (\nabla J)\Big(\frac{y}{N}\Big)\Big], N(1-\eta(y))(y-x)p(y-x) \right\rangle \\
= & \frac{1}{2N^d} \sum_x (\nabla J)\Big(\frac{x}{N}\Big) N \Psi_x
\end{aligned}$$

where

$$\begin{aligned}
\Psi_0 &= \frac{1}{2}\Big[\eta(0)\sum_z (1-\eta(z)) z p(z) + (1-\eta(0))\sum_z \eta(-z) z p(z)\Big] \\
&= \frac{1}{2}\Big[-\eta(0)\sum_z \eta(z) z p(z) + (1-\eta(0))\sum_z \eta(-z) z p(z)\Big] \\
&= \frac{1}{2}\Big[\sum_z \eta(-z) z p(z) - \eta(0)\sum_z (\eta(z)+\eta(-z)) z p(z)\Big].
\end{aligned}$$

The second sum is zero in the symmetric case and Ψ_0 can then be written as a 'gradient' $\Psi_0 = \sum_j \tau_{e_j}\xi_j - \xi_j$ where τ_{e_j} are shifts in the coordinate directions. This allows us to do summation by parts and gain a factor of N^{-1}. When this is not the case, we have a 'nongradient' model and the scaling limit can no longer be established by simple averaging.

Exactly the same situation arises in the symmetric case if we make the probabilities of jumps $p(x) = 1/2d$ for the $2d$ nearest neighbors and 0 otherwise, but change the rates so that the generator reads

$$(\mathcal{A}f)(\eta) = \sum_{|x-y|=1} a_{x,y}(\eta)[f(\eta^{x,y}) - f(\eta)] \tag{2.22}$$

where $a_{x,y}(\eta)$ are translation invariant and satisfy the 'detailed balance', conditions relative to the Bernoulli measures.

There are several good sources for this and related material. In particular the book [4], the monograph [10] and the notes [3] contain all of this material as well as more references.

§3. Large Deviation Methods in Hydrodynamic Scaling

A rigorous proof of the validity of the hydrodynamic scaling limit depends on establishing some sort of a local ergodic theorem. There are several ways of carrying this out depending on the circumstances.

But the methods that are fairly general use ideas from large deiviations in some form. We will consider the example of the totally asymmetric simple exclusion process in one dimension. Suppose we are give a smooth function $\rho(t,x)$ on $[0,T]\times\mathbf{T}$ that satisfies $0<c\le\rho(t,x)\le 1-c<1$ and solves Burgers equation (2.17)

$$\frac{\partial\rho}{\partial t}+\frac{\partial}{\partial x}[\rho(t,x)(1-\rho(t,x)]=0.$$

For any smooth $\rho(x)$ from $\mathbf{T}\to(0,1)$ we can associate a local or slowly varying equlibrium state

$$f^N_{\rho(\,\cdot\,)}(\eta)=\prod_{x=1}^{N}\rho\Big(\frac{x}{N}\Big)^{\eta_x}\Big(1-\rho\Big(\frac{x}{N}\Big)\Big)^{1-\eta_x}. \tag{3.1}$$

We could guess that the state at time t is more or less

$$f_N(t,\eta)=\prod_{x=1}^{N}\rho\Big(t,\frac{x}{N}\Big)^{\eta_x}\Big(1-\rho\Big(t,\frac{x}{N}\Big)\Big)^{1-\eta_x}. \tag{3.2}$$

Even if we start at time 0 with initial distribution $f_N(0,\eta)$ the true state at time t is the solution $g_N(t,\eta)$ of the Kolmogorov forward equation

$$\frac{\partial g_N}{\partial t}=\mathcal{L}^*_N g_N \tag{3.3}$$

with the initial condition

$$g_N(0,\eta)=f_N(0,\eta). \tag{3.4}$$

We wish to compare the true solution g_N to our guess f_N. They match at $t=0$. What about $t>0$?

The comparison is done by

$$H_N(t)=H(g_N(t,\,\cdot\,);f_N(t,\,\cdot\,)). \tag{3.5}$$

It is controlled by establishing a Gronwall type of inequality

$$\frac{dH_N(t)}{dt}\le CH_N(t)+\text{“}\mathit{error}\text{”} \tag{3.6}$$

that leads to

Theorem 3.1. *We have*

$$\lim_{N\to\infty}\sup_{0\le t\le T}\frac{1}{N}H_N(t)=0 \tag{3.7}$$

which in turn implies the validity of the hydrodynamic limit

$$\lim_{N\to\infty}\frac{1}{N}\sum_{x\in\mathbf{Z}_N}J\Big(\frac{x}{N}\Big)\eta_t(x)=\int_{\mathbf{T}}J(x)\rho(t,x)\,dx.$$

One key ingredient is the validity of the local equilibrium principle which needs to be established. This takes the following form. Let ℓ be an intermediate scale, i.e., $1\ll\ell\ll N$. For any local function $g(\eta)$ let us define

$$\hat{g}(\rho)=E^{\mu_\rho}[g(\eta)] \tag{3.8}$$

where μ_ρ is the Bernoulli measure with density ρ. We look at the difference

(3.9)

$$D_{\ell,N,g}(\eta)=\frac{1}{N}\sum_x\left|\frac{1}{2\ell+1}\sum_{y:|y-x|\le\ell}g(\tau_y\eta)-\hat{g}\Big(\frac{1}{2\ell+1}\sum_{y:|y-x|\le\ell}\eta(y)\Big)\right|.$$

Establishing hydrodynamic limit requires

$$\lim_{\epsilon\to0}\limsup_{N\to\infty}E^{P_N}\left[\int_0^T D_{\epsilon N,N,g}(\eta_t)\,dt\right]=0 \tag{3.10}$$

where P_N is the process starting from an arbitrary initial configuration. The relative entropy considerations reduce this to proving estimates of the form

$$\lim_{\ell\to\infty}\limsup_{N\to\infty}E^{P_N}[D_{\ell,N,g}(g)]=0. \tag{3.11}$$

This can be reduced to proving a much stronger estimate for the process Q_N in equilibrium.

$$\lim_{\ell\to\infty}\limsup_{N\to\infty}\frac{1}{N}\log E^{Q_N}\left[\exp N\int_0^T D_{\epsilon N,N,g}(\eta_t)\,dt\right]=0 \tag{3.12}$$

for every g.

This in turn can be estimated in terms of Feynman-Kac representation and variational formulae involving Dirichlet forms.

In the case of diffusive scaling this approach is often powerful enough to yield equation (3.10).

In the last example alluded to in the previous section, i.e., nongradient models, the analysis involves writing the 'current', in the form

$$W_{x,x+1} = c(\eta(x) - \eta(x+1)) + \text{'}\textit{negligible terms}\text{'} \tag{3.13}$$

corresponding to a projection in $L_2(P_\rho)$. The left handside can be thought of as a closed one form while the negligible terms are the exact ones. One has to prove that the codimension of the negligible terms is one and can be represented by the density gradient term. Because the analysis is carried out seperately in each equilibrium this determines $c = c(\rho)$ and one ends up with an equation of the form

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\frac{\partial}{\partial x}\Big[c(\rho)\frac{\partial \rho}{\partial x}\Big]. \tag{3.14}$$

The proofs agian involve establishing superexponetial estimates in equlibrium and use Jensen type inequality (1.15) to go from equilibrium to nonequilibrium.

§4. Large Deviations in Hydrodynamic Scaling

Let us consider $k_N \simeq N^d$ independent random walks on the lattice $\mathbf{Z}_N^d$ of $\mathbf{Z}^d$ modulo N. If we denote their trajectories by $\{x_1(\,\cdot\,), \ldots, x_{k_N}(\,\cdot\,)\}$ and rescale them as $\frac{1}{N}x_i(N^2 t) = y_i(t)$ we have k_N noninteracting rescaled random walks and the empirical process

$$R_{N,\omega} = \frac{1}{N^d}\sum_{i=1}^{k_N} \delta_{y_i(\,\cdot\,)} \tag{4.1}$$

will converge on the Skorohod space of trajectories $D[[0,T];\mathbf{T}^d]$ to a Brownian motion with covariance

$$\langle Ca, b\rangle = \sum_z p(z)\langle z, a\rangle\langle z, b\rangle \tag{4.2}$$

where $p(\,\cdot\,)$ is the probability distribution of a single step and it is assumed to be symmetric. Of course we need to assume that the initial distribution

$$\nu_{N,\omega} = \frac{1}{N^d}\sum_{i=1}^{k_N} \delta_{y_i(0)} \tag{4.3}$$

has a limit μ and if we take Q to be the Brownian motion with Covariance C and initial distribution μ then

$$\lim_{N\to\infty} R_{N,\omega} = Q \tag{4.4}$$

in probability. Because of the way we have normalized, the total mass of Q, which is the same as total mass of μ, is given by

$$\bar{\rho} = \lim_{N\to\infty} \frac{k_N}{N^d}.$$

We can ask about the probabilities of large deviations in this context. It is a minor variation of Sanov's theorem and the rate function with normalization by N^d is given by the relative entropy

$$\mathcal{I}(R) = H(R; Q). \tag{4.5}$$

We would like to see how this changes if we go from the context of independent random walks to an interacting model like simple exclusion. We will keep the jump distribution as the same $p(\,\cdot\,)$.

We saw before that the hydrodynamic limt in this case was still given by equation (2.19). However the behavior of $R_{N,\omega}$ is more complex. For that we have to understand how a tagged particle will behave in equlibrium as well as nonequlibrium. It is known, (see [6]) that a tagged particle in equlibrium will diffuse like a Brownian Motion with some covariane $S(\rho)$ that depends on the density ρ. This is to be expected, because in low density there is very little interaction and one expects $S(\rho) \to C$ as $\rho \to 0$. On the other hand at high density, i.e., when $\rho \to 1$ there is gridlock and one should expect $S(\rho) \to 0$. (The one dimensional nearest neighbor case is different due to blocking and $S(\rho) \equiv 0$ in that one case.)

One expects therefore that if the initial condition is a random configuration chosen from equilibrium with density ρ then, in probability,

$$\lim_{N\to\infty} R_{N,\omega} = Q_\rho \tag{4.6}$$

where Q_ρ is the Browninan motion (with total mass ρ) having covariance $S(\rho)$ and initial density ρ.

In nonequlibrium the situation is a lot more complicated. First of all the density itself is given by the solution $\rho(t,x)$ of the heat equation (2.19) with initial condition $\rho(x)$ which is determined as the limit of the empirical distribution of the initial configuration in the sense of equation (3.4). The tagged particle will only see its immediate neighborhood and will behave as if it is in equilibrium at density $\rho(t,x)$ if it finds itself at time t at the point x. It is reasonable then to expect it to behave like a diffusion with the second order or diffusion coefficients equal to $S(\rho(t,x))$. It could have an additional first order or drift term. It is more convenient to write the expected backward generator in divergence form as

$$\mathcal{L}_t = \frac{1}{2}\nabla S(\rho(t,x))\nabla + c(t,x)\nabla. \tag{4.7}$$

Of course we can tag any are all of the particles and the empirical process is the same whether they or tagged or not. Therefore the solution $\rho(t,x)$ of the heat equation (2.19) must also be a solution of the forward equation corresponding to (4.7), i.e.,

$$\frac{\partial \rho(t,x)}{\partial t} = \frac{1}{2}\nabla S(\rho(t,x))\nabla \rho(t,x) - \nabla \cdot c(t,x)\rho(t,x). \tag{4.8}$$

This means

$$\frac{1}{2}\nabla C \nabla \rho(t,x) = \frac{1}{2}\nabla S(\rho(t,x))\nabla \rho(t,x) - \nabla \cdot c(t,x)\rho(t,x).$$

One can guess (with fingers crossed) that

$$c(t,x) = [S(\rho(t,x) - C]\frac{\nabla \rho(t,x)}{2\rho(t,x)}. \tag{4.9}$$

That this is indeed true is a result in [9] which is based on results of [7].

We now turn to large deviations. To simplify the presentation we assume that the initial configuration is deterministic. Otherwise we have to factor in the large deviation behavior of the initial profile and this adds an extra term to all the rate functions.

Large deviations are invariably obtained by perturbing the dynamics in such a way that the modification produces the needed deviation. The modified process will have, after suitable normalization, some entropy relative to the original process. This can be thought of as 'cost' of the modification. It is conceivable that there are lots of modifications with different costs that produce the same desired deviation. The rate function is always the minimum of such costs. If one can run through a large class of modifications one gets a large deviation lower bound which is the minimum of the costs over that class of modifications. One tries then to match it with an upper bound by some other method.

In our example the possible perturbations are of the jump rates $N^2 p(z)$ of the speeded up dynamics. If the magnitude of the perturbation is $\lambda_N \ll N^2$ then the magnitude of the entropy 'cost' is $\lambda_N^2 N^{-2}$

per particle. This suggests a perurbation of order N to obtain a total entropy 'cost' of order N^d. We therefore consider a perturbed generator of the form

$$(\mathcal{A}_{t,N,q(\,\cdot\,,\,\cdot\,,\,\cdot\,)}f)(\eta) = \sum_{x,y} \eta(x)(1-\eta(y)) \times \left[N^2 p(y-x) + Nq\Big(t, \frac{x}{N}, \cdot\Big)\right][f(\eta^{x,y}) - f(\eta)]$$

where $q(t,x,z)$ is a nice function of t, x and z. If we denote by

$$b(t,x) = \sum zq(t,x,z) \tag{4.11}$$

the effect of the perturbation is to produce a solution of the following modified equation as the hydrodynamic limit.

$$\frac{\partial \rho(t,x)}{\partial t} = \frac{1}{2}\nabla C \nabla \rho(t,x) - \nabla \cdot [b(t,x)\rho(t,x)(1-\rho(t,x))] \tag{4.12}$$

with the same initial condition given by (2.20). Given $\rho(\,\cdot\,,\,\cdot\,)$ we view (4.12) as an equation for $b(\,\cdot\,,\,\cdot\,)$ and denote the set of solutions by $\mathcal{B}_{\rho(\,\cdot\,,\,\cdot\,)}$. For a given $b(\,\cdot\,,\,\cdot\,)$ the set of $q(\,\cdot\,,\,\cdot\,,\,\cdot\,)$ that satisfy (4.11) is denoted by $\mathcal{Q}_{b(\,\cdot\,,\,\cdot\,)}$. The entropy cost when divided by N^d converges to

$$\frac{1}{2}\int_0^T \int_{T^d} \left[\sum_z \frac{q(t,x,z)^2}{p(z)}\right] \rho(t,x)(1-\rho(t,x))\,dxdt. \tag{4.13}$$

Minimizing (4.13) over $\mathcal{Q}_{b(\,\cdot\,,\,\cdot\,)}$ yields

$$\mathcal{E}(b(\,\cdot\,,\,\cdot\,)) = \frac{1}{2}\int_0^T \int_{T^d} \langle b(t,x), C^{-1} b(t,x)\rangle \rho(t,x)(1-\rho(t,x))\,dxdt \tag{4.14}$$

Minimizing $\mathcal{E}(b(\,\cdot\,,\,\cdot\,))$ over $\mathcal{B}_{\rho(\,\cdot\,,\,\cdot\,)}$ gives us the rate function for the large deviation of the empirical density which is the family of one dimensionaal marginals of $R_{N,\omega}$. This was done in [5]. Next, we need to consider the effect of the perturbation on the motion of the tagged particle. This will produce for $R_{N,\omega}$, a weak limit Q_b in probability, with the same initial distribution but with the new backward generator

$$\mathcal{L}_b = \frac{1}{2}\nabla S(\rho(t,x)\nabla + c(t,x)\cdot\nabla + b(t,x)(1-\rho(t,x))\cdot\nabla. \tag{4.15}$$

Finally we can write down the rate function $I(R)$ for the large deviations of $R_{N,\omega}$. From R in addition to its one dimensional marginals

$\rho(\cdot, \cdot)$ we can consider the 'currents'

$$\Lambda_R(f) = E^R\left[\int_0^T \langle f(s, x(s)), dx(s)\rangle\right]. \tag{4.16}$$

From $\mathcal{B}_{\rho(\cdot,\cdot)}$ we look for a $\bar{b}(\cdot, \cdot)$ such that

$$\Lambda_R(f) \equiv \Lambda_{Q_{\bar{b}}}(f)$$

and call it $\overline{R}$. The rate function turns out to be

$$I(R) = \mathcal{E}(\bar{b}(\cdot, \cdot)) + H(R; \overline{R}). \tag{4.17}$$

If the marginal of R does not match the initial density or if we have trouble defining (4.17) at any stage then $I(R)$ is $+\infty$. It turns out that $\bar{b}(\cdot, \cdot)$ if it exists is unique. Details of these results can be found in [8].

References

[1] H. Cramér, Sur un nouveau théeorème-limite de la théorie des probabilités, Actualités Scientifiques et Industrielles, number **736**, Hermann, Paris (1938), 5–23.

[2] J.-D. Dueschel and W. D. Stroock, "Large Deviations", Academic Press, Boston, 1989.

[3] L. Jensen and H.-T. Yau, "Hydrodynamic Scaling Limits of Simple Exclusion Models, Probability Theory and Applications", (P. H. Elton and S. R. S Varadhan, eds.), IAS-Park City Mathematics Series, Vol. **6**, American Mathematical Society, Providence, Rhode Island, 1999, pp. 167–225.

[4] C. Kipnis and C. Landim, "Scaling limits of interacting particle systems", Springer-Varlag, Berlin, 1999.

[5] C. Kipnis, S. Olla and S. R. S. Varadhan, Hydrodynamics and large deviations for simple exclusion processes, Comm. Pure Appl. Math., **42** (1989), 115–137.

[6] C. Kipnis and S. R. S. Varadhan, Central Limit Theorem for additive functionals of reversible Markov processes and applications to simple exclusion, Comm. Math. Phys., **104** (1986), 1–19.

[7] J. Quastel, Diffusion of color in the simple exclusion process, Comm. Pure Appl. Math., **45** (1992), 623–679.

[8] J. Quastel, F. Rezakhanlou and S. R. S. Varadhan, Large deviations for the symmetric simple exclusion process in dimensions $d \geq 3$, Probab. Theory Related Fields, **113,** no. 1 (1999), 1–84.

[9] F. Rezakhanlou, Propagation of chaos for symmetric simple exclusions, Comm. Pure Appl. Math., **47,** no. 7 (1994), 943–957.

[10] H. Spohn, "Large Scale Dynamics of Interacting Particles", Springer-Verlag, New York, 1991.

Courant Institute of Mathematical Sciences
251 Mercer Street
New York, NY 10012
U.S.A.
`varadhan@cims.nyu.edu`