Functional Inequalities, Markov Semigroups and Spectral Theory

Feng-Yu Wang

School of Mathematical Science, Beijing Normal University

SCIENCE PRESS
Beijing/New York

Responsible Editors : Chen Yuzhuo, Lv Hong, Zu Cui'e

Published by Science Press
16 Donghuangchenggen North Street
Beijing 100717, China

Printed in Beijing

ISBN 7-03-014415-5

Preface

Since in a standard situation (e.g. in the symmetric case), any C_0-contraction semigroup (and hence its generator) on a Hilbert space is uniquely determined by the associated quadratic form, it is reasonable to describe the properties of the semigroup and its generator by using functional inequalities of the quadratic form. In particular, if the associated form is a Dirichlet form, then the corresponding semigroup is (sub-) Markovian. The purpose of this book is to present a systematic account of functional inequalities for Dirichlet forms and applications to Markov semigroups (or Markov processes in a regular case).

The functional inequalities considered here only involve in the Dirichlet form and one or two norms of functions, and can be easily illustrated in many cases. On the other hand, these inequalities imply plentiful analytic properties of Markov semigroups and generators, which are related to various behaviors of the corresponding Markov processes. For instance, the Poincaré inequality is equivalent to the exponential convergence of the semigroup and the existence of the spectral gap. Moreover, the Gross log-Sobolev inequality is equivalent to Nelson's hypercontractivity of the semigroup and is strictly stronger than the Poincaré inequality. So, it is natural for us to ask for more spectral information and semigroup properties from more general functional inequalities. This is the starting point of the book.

In this book, we introduce functional inequalities to describe:

(i) the spectrum of the generator: the essential and discrete spectrums, high order eigenvalues, the principal eigenvalue, and the spectral gap;

(ii) the semigroup properties: the uniform integrability, the compactness, the convergence rate, and the existence of density;

(iii) the reference measure and the intrinsic metric: the concentration, the isoperimetric inequality, and the transportation cost inequality.

For reader's convenience and for the completeness of the account, we summarize some necessary preliminaries in Chapter 0. Corresponding to various levels of spectral and semigroup properties, Chapters 1, 3, 4, 5 and 6 focus on several different functional inequalities respectively: Chapter 1 and Chapter 5 introduce the above mentioned Poincaré and log-Sobolev inequalities respectively, Chapter 6 the interpolations of these two inequalities, Chapter 3 the super Poincaré inequality, and Chapter 4 the weak Poincaré inequality. Each of these chapters presents a correspondence between the underlying functional inequality and the properties of the semigroup and its generator, as well as sufficient and necessary conditions for the functional inequality to hold. Moreover, the general results are illustrated by concrete examples, in particular, examples of diffusion processes on manifolds and countable Markov chains. These chapters are relatively (although not absolutely) independent, so that one may read in one's own order without much trouble.

Chapter 2 is devoted to diffusion processes on Riemannian manifolds and applications to geometry analysis. In particular, the estimation of the first eigenvalue is related to the Poincaré inequality, while the results concerning gradient estimates, the Harnack inequality and the isoperimetric inequality will be used in the sequel to illustrate other functional inequalities. The results included in §2.2 concerning the first eigenvalue have been introduced in a recent monograph [47] by Professor Mu-Fa Chen. Chen's monograph emphasizes the main idea of the study which is crucial for understanding the machinery of the work, while the present book provides the technical details which are useful for further study. Finally, in Chapter 7 we establish functional inequalities for three infinite-dimensional models which have been studying intensively in stochastic analysis and mathematical physics.

At the end of each chapter (except Chapter 0), some historical notes and open questions for further studies are addressed. The notes are not intended to summarize the principal results of each paper cited but merely to indicate the connection to the main contents of each chapter in question, while the open problems are listed mainly based on my own interests. Thus, these notes are far from complete in the strict sense. At the end of the book, a list of publications and an index of main notations and key words are presented for reader's reference. These references are presented not for completeness but for a usable guide to the literature. I regret that there might be a lot of related publications which have not been mentioned in the book.

Due to the limitation of knowledge and the experience of writing, I would like to apologize in advance for possible mistakes and incomplete accounts appeared in this book, and to appreciate criticisms and corrections in any sense.

I would like to express my deep gratitude to my advisors Professor Shi-Jian Yan and Professor Mu-Fa Chen for earnest teachings and constant helps. Professor Chen guided me to the cross research field of probability theory and Riemannian manifold, and emphasized probabilistic approaches in research, in particular, the coupling methods which he had worked on intensively. Our fruitful cooperations in this direction considerably stimulated other work included in this book. During the past decade I also greatly benefited from collaborations and communications with Professors M. Röckner, A. Thalmaier, V. I. Bogachev, F.-Z. Gong, K. D. Elworthy, M. Cranston and X.-M. Li. In particular, the work concerning the weak Poincaré inequality and applications is due to effective cooperations with Professor M. Röckner. At different stages I received helpful suggestions and encouragements from many other mathematicians, in particular, Professors S. Aida, S. Albeverio, D. Barkry, D. Chaifi, D.-Y. Chen, T. Couhlon, S. Fang, M. Fukushima, G.-L. Gong, L. Gross, E. Hsu, C.-R. Hwang, W.S. Kendall, R. Leandre, M. Ledoux, Z.-H. Li, Z.-M. Ma, P. Malliavin, Y.-H. Mao, S.-G. Peng, E. Priola, M.-P. Qian, I. Shigekawa, D. Stroock, K.-T. Sturm, Y.-L. Sun, J.-L. Wu, L. Wu, J.-A. Yan, T.-S. Zhang, Y.-H. Zhang and X.-L. Zhao. I would also like to thank Professor Yu-Hui Zhang, Mr Wei Liu and graduate students in our group for reading the draft and checking errors.

Finally, it is a pleasure to acknowledge the generous support of this work by the National Natural Science Foundation of China (1002510, 10121101), the Teaching and Research Awarded Project for Outstanding Young Teachers, and the 973-Project.

Feng-Yu Wang
Beijing, June 2004

Contents

Chapter 0

Preliminaries

In this chapter we briefly recall some necessary preliminaries of the book from Dirichlet forms, Markov processes, spectral theory and Riemannian geometry. Results included in this part are well-known and fundamental in these fields. §0.1 and §0.2 are mainly summarized from [137], most results in §0.3 can be found in [225] and [155], and §0.4 is mainly selected from [33] and [35].

0.1 Dirichlet forms, sub-Markov semigroups and generators

Let us start with some basic facts on semigroups, resolvents and generators. Let $(\mathbb{B}, \|\cdot\|)$ be a real Banach space. A pair $(L, \mathscr{D}(L))$ is called a *linear operator* on $\mathbb{B}$ if $\mathscr{D}(L)$ is a linear subspace of $\mathbb{B}$ and $L : \mathscr{D}(L) \to \mathbb{B}$ is a linear map. We sometimes simply denote the operator by L. The operator L is called *closed* if its graph $\{(f, Lf) : f \in \mathscr{D}(L)\}$ is closed in $\mathbb{B} \times \mathbb{B}$. A linear operator $(L, \mathscr{D}(L))$ is called *closable* if the closure of its graph is the graph of a linear operator $(\bar{L}, \mathscr{D}(\bar{L}))$ which is called the *closure* of $(L, \mathscr{D}(L))$.

Definition 0.1.1 A family $\{P_t\}_{t \geqslant 0}$ of linear operators on $\mathbb{B}$ with $\mathscr{D}(P_t) = \mathbb{B}$ for all $t \geqslant 0$ is called a *strongly continuous* (or C_0-) *contraction semigroup* on $\mathbb{B}$, if

(1) $\lim_{t \to 0} P_t f = P_0 f = f, \qquad f \in \mathbb{B}.$

(2) $\|P_t\| := \sup\{\|P_t f\| : f \in \mathbb{B}, \|f\| \leqslant 1\} \leqslant 1, \qquad t \geqslant 0.$

(3) $P_t P_s = P_{t+s}, \qquad t, s \geqslant 0.$

For a given C_0-contraction semigroup $\{P_t\}_{t \geqslant 0}$ (simply denoted by P_t in the sequel), define

$$\mathscr{D}(L) := \Big\{ f \in \mathbb{B} : \lim_{t \to 0} \frac{1}{t}(P_t f - f) \text{ exists in } \mathbb{B} \Big\},$$

$$Lf := \lim_{t \to 0} \frac{1}{t}(P_t f - f), \qquad f \in \mathscr{D}(L).$$

Then $(L, \mathscr{D}(L))$ is a linear operator on $\mathbb{B}$, which is called the (infinitesimal) *generator* of P_t. The following well-known result provides a complete characterization for generators of C_0-contraction semigroups (see e.g. [225]).

Theorem 0.1.1 (Hille-Yoshida Theorem) *A linear operator $(L, \mathscr{D}(L))$ is the generator of a C_0-contraction semigroup if and only if*

(1) *L is densely defined, i.e. $\mathscr{D}(L)$ is dense in $\mathbb{B}$.*

(2) *For any $\lambda > 0, (\lambda - L)$ is invertible and $\|(\lambda - L)^{-1}\| \leqslant \lambda^{-1}$.*

In this case the corresponding semigroup is uniquely determined by L and is denoted by $P_t = \mathrm{e}^{tL}$, and L is closed.

Let $(L, \mathscr{D}(L))$ be the generator of a C_0-contraction semigroup P_t. We have

$$R_\lambda f := (\lambda - L)^{-1} f = \int_0^\infty \mathrm{e}^{-\lambda s} P_s f \mathrm{d}s, \qquad f \in \mathbb{B}, \quad \lambda > 0.$$

We call $\{R_\lambda : \lambda > 0\}$ the *resolvent* of L or P_t, see §0.3 for this notion of linear operators on complex Banach spaces.

Now, let us consider $\mathbb{B} := \mathbb{H}$, a real Hilbert space with inner product $\langle , \rangle$. Then an operator $(L, \mathscr{D}(L))$ provides a bilinear map $\mathscr{E} : \mathscr{D}(L) \times \mathscr{D}(L) \to \mathbb{R}$ with $\mathscr{E}(f, g) := -\langle Lf, g\rangle$ for $f, g \in \mathscr{D}(L)$. In general, $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ is called a *bilinear form* on $\mathbb{H}$ if $\mathscr{D}(\mathscr{E})$ is a linear subspace of $\mathbb{H}$ and $\mathscr{E} : \mathscr{D}(\mathscr{E}) \times \mathscr{D}(\mathscr{E}) \to \mathbb{H}$ is a bilinear map. If moreover $\mathscr{E}(f, f) \geqslant 0$ for $f \in \mathscr{D}(\mathscr{E})$, then $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ is called a *positive definite form* on $\mathbb{H}$. For a bilinear form $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$, we define its symmetric part by $\tilde{\mathscr{E}}(f, g) := \frac{1}{2}(\mathscr{E}(f, g) + \mathscr{E}(g, f)), f, g \in \mathscr{D}(\mathscr{E})$. Moreover, let $\mathscr{E}_\alpha(f, g) := \alpha\langle f, g\rangle + \mathscr{E}(f, g), f, g \in \mathscr{D}(\mathscr{E}), \alpha \geqslant 0$.

Definition 0.1.2 Let $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ be a densely defined positive definite form on $\mathbb{H}$.

(1) $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ is called *symmetric* if $\mathscr{E}(f, g) = \mathscr{E}(g, f),\ f, g \in \mathscr{D}(\mathscr{E})$.

(2) $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ is called *closed* if $\mathscr{D}(\mathscr{E})$ is complete under the norm $\mathscr{E}_1^{1/2}$.

(3) $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ is called a *coercive closed form* on $\mathbb{H}$ if it is closed and there exists a constant $K > 0$ such that

$$|\mathscr{E}_1(f, g)| \leqslant K \mathscr{E}_1(f, f)^{1/2} \mathscr{E}_1(g, g)^{1/2}, \qquad f, g \in \mathscr{D}(\mathscr{E}). \tag{0.1.1}$$

Condition (0.1.1) is called the *weak sector condition*.

The following result gives a correspondence between the coercive closed forms and the generators of C_0-contraction semigroups.

Theorem 0.1.2 (1) *Let $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ be a coercive closed form. Define*

$$\mathscr{D}(L) := \{f \in \mathscr{D}(\mathscr{E}) : \text{ the map } \mathscr{E}(f, \cdot) : \mathscr{D}(\mathscr{E}) \to \mathbb{R} \text{ is continuous under } \| \cdot \|\},$$

and for $f \in \mathscr{D}(L)$ *define* $Lf \in \mathbb{H}$ *via* $-\langle Lf, g\rangle = \mathscr{E}(f,g)$ *for all* $g \in \mathscr{D}(\mathscr{E})$. *Then* L *is the generator of a* C_0*-contraction semigroup* P_t *with resolvent* $\{R_\lambda\}_{\lambda>0}$ *satisfying*

$$R_\lambda(\mathbb{H}) \subset \mathscr{D}(\mathscr{E}) \ \ \text{and} \ \ \mathscr{E}_\lambda(R_\lambda f, g) = \langle f, g\rangle, \qquad f \in \mathbb{H}, g \in \mathscr{D}(\mathscr{E}), \lambda > 0. \tag{0.1.2}$$

In particular, $(L, \mathscr{D}(L))$ *satisfies the weak sector condition: there exists* $K > 0$ *such that*

$$|\langle (1-L)f, g\rangle| \leqslant K\sqrt{\langle (1-L)f, f\rangle \langle (1-L)g, g\rangle}, \qquad f, g \in \mathscr{D}(L). \tag{0.1.3}$$

(2) *If* $(L, \mathscr{D}(L))$ *satisfies* (0.1.3) *and generates a* C_0*-semigroup, then there exists a unique coercive closed form* $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ *such that* $\mathscr{D}(\mathscr{E})$ *is the completion of* $\mathscr{D}(L)$ *with respect to* $\tilde{\mathscr{E}}_1^{1/2}$ *and*

$$\mathscr{E}(f,g) = -\langle Lf, g\rangle, \qquad f, g \in \mathscr{D}(L),$$

Furthermore, the resolvent $\{R_\lambda\}_{\lambda>0}$ *satisfies* (0.1.2).

Finally, let us consider the Markovian setting. Let $(E, \mathscr{F}, \mu)$ be a measure space and let $\mathbb{H} := L^2(\mu)$, the set of all measurable real functions which are square-integrable with respect to μ, that is, letting $\mathscr{B}(E)$ be the set of all measurable real functions on E, we have

$$L^2(\mu) := \Big\{ f \in \mathscr{B}(E) : \mu(f^2) := \int_E f^2 \mathrm{d}\mu < \infty \Big\}.$$

We write $f \leqslant g$ or $f < g$ if the corresponding inequality holds μ-a.e.

Definition 0.1.3 (1) A bounded linear operator $P : L^2(\mu) \to L^2(\mu)$ is called *sub-Markovian* if $0 \leqslant Pf \leqslant 1$ for all $f \in L^2(\mu)$ with $0 \leqslant f \leqslant 1$. If furthermore $P1 = 1$ then P is called a *Markov operator*. A semigroup $\{P_t\}_{t\geqslant 0}$ is called a sub-Markov (resp. Markov) semigroup if each P_t is sub-Markovian (resp. Markovian).

(2) A closed densely defined linear operator $(L, \mathscr{D}(L))$ on $L^2(\mu)$ is called a *Dirichlet operator* if $\langle Lf, (f-1)^+\rangle \leqslant 0$ for all $f \in \mathscr{D}(L)$. If moreover $1 \in \mathscr{D}(L)$ and $L1 = 0$ then L is called a *conservative Dirichlet generator*.

(3) A coercive closed form $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ on $L^2(\mu)$ is called a *Dirichlet form* if for any $f \in \mathscr{D}(\mathscr{E})$, one has $f^+ \wedge 1 \in \mathscr{D}(\mathscr{E})$ and

$$\mathscr{E}(f + f^+ \wedge 1, f - f^+ \wedge 1) \geqslant 0, \qquad \mathscr{E}(f - f^+ \wedge 1, f + f^+ \wedge 1) \geqslant 0. \tag{0.1.4}$$

A Dirichlet form $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ is called *conservative* if $1 \in \mathscr{D}(\mathscr{E})$ and $\mathscr{E}(f, 1) = \mathscr{E}(1, f) = 0$ for all $f \in \mathscr{D}(\mathscr{E})$.

Proposition 0.1.3 (1) *If* $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ *is a Dirichlet form then so is its symmetric part* $(\tilde{\mathscr{E}},\mathscr{D}(\mathscr{E}))$.

(2) *A symmetric closed form* $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ *is a Dirichlet form if and only if for any* $f\in\mathscr{D}(\mathscr{E})$ *one has* $f^+\wedge 1\in\mathscr{D}(\mathscr{E})$ *and*

$$\mathscr{E}(f^+\wedge 1, f^+\wedge 1)\leqslant \mathscr{E}(f,f). \tag{0.1.5}$$

(3) *A symmetric closed form* $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ *is a Dirichlet form if and only if for any* $T:\mathbb{R}\to\mathbb{R}$ *with* $T(0)=0$ *and* $|T(x)-T(y)|\leqslant|x-y|$ *for all* $x,y\in\mathbb{R}$, *one has* $T\circ f\in\mathscr{D}(\mathscr{E})$ *and* $\mathscr{E}(T\circ f,T\circ f)\leqslant\mathscr{E}(f,f)$ *for all* $f\in\mathscr{D}(\mathscr{E})$.

(4) *Let* $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ *be a Dirichlet form. If* $f\in\mathscr{D}(\mathscr{E})$ *and* $g\in L^2(\mu)$ *satisfies* $|g|\leqslant|f|, |g(x)-g(y)|\leqslant|f(x)-f(y)|$, *then* $g\in\mathscr{D}(\mathscr{E})$ *and* $\mathscr{E}(g,g)\leqslant\mathscr{E}(f,f)$.

Theorem 0.1.4 *Let* $(L,\mathscr{D}(L))$ *generate a* C_0*-contraction semigroup* $\{P_t\}_{t\geqslant 0}$, *and let* $\{R_\lambda\}_{\lambda>0}$ *be the corresponding resolvent. Then the following are equivalent.*

(1) *L is a Dirichlet operator* (*resp. conservative Dirichlet operator*).

(2) $\{P_t\}_{t\geqslant 0}$ *is sub-Markovian* (*resp. Markovian*).

(3) *For each* $\lambda>0, \lambda R_\lambda$ *is sub-Markovian* (*resp. Markovian*).

If $(L,\mathscr{D}(L))$ *satisfies the weak sector condition* (0.1.3) *and* $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ *is the associated coercive closed form, then they are also equivalent.*

(4) *For any* $f\in\mathscr{D}(\mathscr{E})$ *one has* $f^+\wedge 1\in\mathscr{D}(\mathscr{E})$ *and* $\mathscr{E}(f+f^+\wedge 1, f-f^+\wedge 1)\geqslant 0$ (*resp. moreover* $1\in\mathscr{D}(\mathscr{E})$ *with* $\mathscr{E}(1,f)=1$ *for all* $f\in\mathscr{D}(\mathscr{E})$).

Corollary 0.1.5 *Let* $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ *be a coercive closed form associated to the generator* $(L,\mathscr{D}(L))$, *the semigroup* $\{P_t\}_{t\geqslant 0}$ *and the resolvent* $\{R_\lambda\}_{\lambda>0}$. *Let* P_t^* (*resp.* R_λ^*) *be the adjoint operator* (*see Definition 0.3.2 below*) *of* P_t (*resp.* R_λ) *on* $L^2(\mu)$ *for* $t\geqslant 0$ (*resp.* $\lambda>0$), *and let* $(L^*,\mathscr{D}(L^*))$ *be the corresponding generator. Then the following are equivalent.*

(1) $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ *is a Dirichlet form* (*resp. conservative Dirichlet form*).

(2) *L and L^* are Dirichlet operators* (*resp. conservative Dirichlet operators*).

(3) $\{P_t\}_{t\geqslant 0}$ *and* $\{P_t^*\}_{t\geqslant 0}$ *are sub-Markovian* (*resp. Markovian*).

(4) λR_λ *and* λR_λ^* *are sub-Markovian* (*resp. Markovian*) *for each* $\lambda>0$.

In applications, $\mathscr{E}$ is often explicitly defined on a smaller domain $\mathscr{D}(\mathscr{E})$ so that $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ is not closed. To determine a closed form, one needs to find a closed extension of $(\mathscr{E},\mathscr{D}(\mathscr{E}))$. To this end, we introduce the notion of closability of the form.

Definition 0.1.4 A positive definite bilinear form $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ is called *closable* if it has a closed extension $(\mathscr{E}',\mathscr{D}(\mathscr{E}'))$, i.e. $(\mathscr{E}',\mathscr{D}(\mathscr{E}'))$ is a closed form with $\mathscr{D}(\mathscr{E}')\supset\mathscr{D}(\mathscr{E})$ and $\mathscr{E}'|_{\mathscr{D}(\mathscr{E})}=\mathscr{E}$.

Proposition 0.1.6 *Let $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ be a positive definite bilinear form satisfying the weak sector condition* (0.1.1)*.*

(1) *$(\mathscr{E},\mathscr{D}(\mathscr{E}))$ is closable if and only if for any $\mathscr{E}$-Cauchy sequence $\{f_n\}\subset\mathscr{D}(\mathscr{E})$ (f i.e. $\mathscr{E}(f_n-f_m,f_n-f_m)\to 0$ as $n,m\to\infty$) with $f_n\to 0 (n\to\infty)$ in $L^2(\mu)$, one has $\mathscr{E}(f_n,f_n)\to 0(n\to\infty)$.*

(2) *$(\mathscr{E},\mathscr{D}(\mathscr{E}))$ is closable if and only if so is its symmetric part $(\tilde{\mathscr{E}},\mathscr{D}(\mathscr{E}))$.*

(3) *If $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ is closable, then it extends uniquely to the completion of $\mathscr{D}(\mathscr{E})$ with respect to the norm $\mathscr{E}_1^{1/2}$, denoted by $(\bar{\mathscr{E}},\mathscr{D}(\bar{\mathscr{E}}))$. If moreover $\mathscr{D}(\mathscr{E})$ is dense in $L^2(\mu)$ then $(\bar{\mathscr{E}},\mathscr{D}(\bar{\mathscr{E}}))$ is the smallest coercive closed form extending $(\mathscr{E},\mathscr{D}(\mathscr{E}))$, and is called the closure of $(\mathscr{E},\mathscr{D}(\mathscr{E}))$.*

Proposition 0.1.7 (1) *Let $(L,\mathscr{D}(L))$ be a negative definite operator on $L^2(\mu)$, satisfying the weak sector condition* (0.1.3)*. Define*

$$\mathscr{E}(f,g):=-\langle Lf,g\rangle,\qquad f,g\in\mathscr{D}(L).$$

Then $(\mathscr{E},\mathscr{D}(L))$ is closable on $L^2(\mu)$.

(2) *Let $(\mathscr{E}^{(k)},\mathscr{D}(\mathscr{E}^{(k)})), k\in\mathbb{N}$, be closable (resp. closed) positive definite symmetric forms on $L^2(\mu)$. Let*

$$\mathscr{D}(\mathscr{E}):=\left\{f\in\bigcap_{k\geqslant 1}\mathscr{D}(\mathscr{E}^{(k)}):\sum_{k=1}^{\infty}\mathscr{E}^{(k)}(f,f)<\infty\right\},$$

$$\mathscr{E}(f,g):=\sum_{k=1}^{\infty}\mathscr{E}^{(k)}(f,g),\qquad f,g\in\mathscr{D}(\mathscr{E}).$$

Then $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ is closable (resp. closed) on $L^2(\mu)$.

(3) *Let $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ be a coerceive closed form and $\{f_n\}\subset\mathscr{D}(\mathscr{E})$ such that $\{\mathscr{E}(f_n,f_n)\}$ is bounded and $f_n\to f\in L^2(\mu)$ as $n\to\infty$, then $f\in\mathscr{D}(\mathscr{E})$ and*

$$\lim_{n\to\infty}\tilde{\mathscr{E}}(f_n,f)=\mathscr{E}(f,f)\leqslant\varliminf_{n\to\infty}\mathscr{E}(f_n,f_n).$$

Finally, the following result (see [91, Theorem 1.5.2]) enables us to extend the domain of a Dirichlet form.

Theorem 0.1.8 *Let $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ be a symmetric Dirichlet form on $L^2(\mu)$. For any measurable function f, if there exists an $\mathscr{E}$-Cauchy sequence $\{f_n\}\subset\mathscr{D}(\mathscr{E})$ such that $f_n\to f$ μ-a.e., then the limit $\mathscr{E}(f,f):=\lim_{n\to\infty}\mathscr{E}(f_n,f_n)$ exists and does not depend on the choice of $\{f_n\}$. If moreover $f\in L^2(\mu)$ then $f\in\mathscr{D}(\mathscr{E})$.*

According to Theorem 0.1.8 we may extend the Dirichlet form to the extended domain

$$\mathscr{D}_e(\mathscr{E}):=\{f\in\mathscr{B}(E):\ f_n\to f\ \mu\text{-a.e. for some}$$

$$\mathscr{E}\text{-Cauchy sequence } \{f_n\} \subset \mathscr{D}(\mathscr{E})\},$$

where $\mathscr{B}(E)$ is the set of all measurable real functions on E. Throughout the book, all real or complex functions are assumed to be finite.

0.2 Dirichlet forms and Markov processes

In this section we introduce the correspondence between Dirichlet forms and Markov processes, i.e. to show how these two objects determine each other. We first recall the notion of Markov processes.

Let E be a Hausdorff topological space with the Borel σ-field $\mathscr{F}$, that is, the σ-field induced by open sets. A stochastic process with state space E describes the behavior of a particle randomly moving on E. If the particle is allowed to move out from E, then we add a new point Δ to stand for the "died state" of the particle. Thus, the whole state space becomes $E_\Delta := E \bigcup \{\Delta\}$ equipped with the natural one-point compaction topology, i.e. a subset G of E_Δ is open if it is either an open set in E or a set containing Δ with compact complement. If in particular E itself is compact, then Δ is isolated. Let $\mathscr{F}_\Delta$ be the corresponding Borel σ-field. For any function f on E, we extend it to E_Δ by letting $f(\Delta) = 0$.

For simplicity, we only consider the standard Markov process defined on the *canonical path space* over E. A map $\omega_\cdot : [0, \infty) \to E_\Delta$ is called a canonical path if it is right continuous and has left limit at each point $t > 0$ with $\omega_t \neq \Delta$. Let Ω denote the set of all canonical pathes over E such that $\omega_t = \Delta$ for all $t \geqslant \xi(\omega) := \inf\{t \geqslant 0 : \omega_t = \Delta\}$, where ξ is called the *lifetime*. For each $t \geqslant 0$, let

$$x_t : \Omega \to E; \quad x_t(\omega) := \omega_t, \qquad \omega \in \Omega,$$

and let $\mathscr{F}_t := \sigma(x_s : s \leqslant t)$ be the smallest σ-field on Ω such that x_s is measurable for all $s \leqslant t$. Let $\mathscr{F}_\infty := \sigma(x_t : t \geqslant 0)$. The family $\{\mathscr{F}_t\}_{t\geqslant 0}$ is called the *natural filtration* of the path process over E. To make the filtration *right continuous*, let $\mathscr{F}_t^+ := \bigcap_{s>t} \mathscr{F}_s, t \geqslant 0$. In the sequel, whenever $(\Omega, \mathscr{F}_\infty)$ is equipped with a probability measure $\mathbb{P}$, the filtration under consider is automatically extended to its completion with respect to $\mathbb{P}$.

Definition 0.2.1 A family of probability measures $\{\mathbb{P}^x : x \in E_\Delta\}$ on $(\Omega, \mathscr{F}_\infty)$ is called a *Markov process* on E, if

(1) $\mathbb{P}^x(x_0 \in A) = \delta_x(A) := \mathbf{1}_A(x), x \in E_\Delta, A \in \mathscr{F}_\Delta$, where $\mathbf{1}_A$ is the indicator function of A. In particular, $\mathbb{P}^\Delta(x_t = \Delta, t \geqslant 0) = 1$.

(2) For any $\Gamma \in \mathscr{F}_\infty, \mathbb{P}^\cdot(\Gamma)$ is $\mathscr{F}_\Delta$-measurable.

(3) For any $s, t \geqslant 0$ and any $x \in E, A \in \mathscr{F}_\Delta$,

$$\mathbb{P}^x(x_{t+s} \in A|\mathscr{F}_s) = \mathbb{P}^x(x_{t+s} \in A|x_s), \qquad \mathbb{P}^x\text{-a.s.}, \tag{0.2.1}$$

where $\mathbb{P}^x(\cdot|x_s)$ is the conditional probability of $\mathbb{P}^x$ under the σ-field induced by x_s. The equation (0.2.1) is called the *Markov property* (with respect to the filtration $\{\mathscr{F}_t\}$). If for any $x \in E$ one has $\mathbb{P}^x(\xi=\infty)=1$, then we may drop Δ and call $\{\mathbb{P}^x : x \in E\}$ a nonexplosive (or conservative) Markov process on E.

Given a Markov process $\{\mathbb{P}^x : x \in E_\Delta\}$, and given $\nu \in \mathscr{P}(E_\Delta)$, the set of all probability measures on E_Δ, let $\mathbb{P}^\nu := \int_{E_\Delta} \mathbb{P}^x \nu(\mathrm{d}x)$, which is called the distribution of the Markov process starting from ν.

In this book, we only consider the *time-homogeneous* Markov process for which the Markov property (0.2.1) can be written as

$$\mathbb{P}^x(x_{t+s} \in A|\mathscr{F}_s) = \mathbb{P}^{x_s}(x_t \in A), \qquad \mathbb{P}^x\text{-a.s.}, x \in E_\Delta, A \in \mathscr{F}_\Delta, s, t \geqslant 0. \tag{0.2.2}$$

For a time-homogenous Markov process $\{\mathbb{P}^x : x \in E_\Delta\}$, define

$$P_t f(x) := \mathbb{E}^x f(x_t) := \int_\Omega f(x_t)\mathrm{d}\mathbb{P}^x, \qquad f \in \mathscr{B}_+(E), x \in E,$$

where $\mathscr{B}_+(E)$ is the set of nonnegative measurable functions on E. Since $f(\Delta) = 0$ by convention, we have

$$P_t f(x) = \mathbb{E}^x f(x_t) 1_{\{t<\xi\}}, \qquad x \in E, t \geqslant 0.$$

It is easy to see from (0.2.2) that $\{P_t\}_{t\geqslant 0}$ is a sub-Markov semigroup on $\mathscr{B}_b(E) := \{f \in \mathscr{B}(E) : \|f\| := \sup|f| < \infty\}$, which is a Banach space and the norm is called the *uniform norm*.

To introduce the definition of strong Markov processes, let us recall the notion of a *stopping time*. A mapping $\tau : \Omega \to [0, \infty]$ is called a $\{\mathscr{F}_t\}$-stopping time if $\{\tau \leqslant t\} \in \mathscr{F}_t$ for all $t \geqslant 0$. Given a stopping time τ we define the σ-field

$$\mathscr{F}_\tau := \{\Gamma \in \mathscr{F}_\infty : \Gamma \bigcap \{\tau \leqslant t\} \in \mathscr{F}_t,\ t \geqslant 0\}.$$

Definition 0.2.2 A time-homogenous Markov process $\{\mathbb{P}^x : x \in E_\Delta\}$ is called a *strong Markov process* if for any $\{\mathscr{F}_t\}$-stopping time τ, any $\nu \in \mathscr{P}(E_\Delta)$ and any $A \in \mathscr{F}_\Delta$,

$$\mathbb{P}^\nu(x_{t+\tau} \in A|\mathscr{F}_\tau) = \mathbb{P}^{x_\tau}(x_t \in A), \qquad \mathbb{P}^\nu\text{-a.s.} \quad \text{on} \quad \{\tau < \infty\}. \tag{0.2.3}$$

Now, let us connect Markov processes with Dirichlet forms.

Definition 0.2.3 Let μ be a measure on $(E,\mathscr{F})$. Let $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ be a Dirichlet form on $L^2(\mu)$ with $\{T_t\}_{t\geqslant 0}$ the associated sub-Markov C_0-contraction semigroup. A Markov process $\{\mathbb{P}^x : x\in E_\Delta\}$ is called associated with $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ if its semigroup P_t satisfies $P_tf = T_tf$ μ-a.e. for all $t>0$ and all $f\in\mathscr{B}_b(E)\bigcap L^2(\mu)$.

Proposition 0.2.1 *Let $\{\mathbb{P}^x : x\in E_\Delta\}$ be a Markov process with semigroup $\{P_t\}_{t\geqslant 0}$. If $\{P_t\}_{t\geqslant 0}$ is strongly continuous on $L^2(\mu)$ with generator satisfying the weak sector condition, then P_t is associated with a unique coercive closed form $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ such that for any $f\in\mathscr{D}(\mathscr{E})$ one has $f^+\wedge 1\in\mathscr{D}(\mathscr{E})$ and $\mathscr{E}(f+f^+\wedge 1, f-f^+\wedge 1)\geqslant 0$. If moreover P_t^* is also sub-Markovian, it is the case when μ is a P_t-supermedian measure, i.e. $\int_E P_tf\mathrm{d}\mu\leqslant\int_E f\mathrm{d}\mu$ for all $f\in\mathscr{B}_+(E)$, then $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ is a Dirichlet form.*

Remark 0.2.1 Assume that μ is finite and P_t-supermedian. Let $C(E)$ (resp. $C_b(E)$) be the set of all continuous (resp. bounded continuous) real functions on E. Then for any $f\in C_b(E)$, the right-continuity of the process implies $P_tf\to f$ and hence $\|P_tf-f\|_{L^2(\mu)}\to 0$ as $t\to 0$. If moreover $\sigma(C(E))=\mathscr{F}$, i.e. the Borel σ-field is induced by the class of continuous functions, then $C_b(E)$ is dense in $L^2(\mu)$ and hence $\{P_t\}_{t\geqslant 0}$ is a sub-Markov C_0-contraction semigroup on $L^2(\mu)$. For general μ , $\{P_t\}_{t\geqslant 0}$ is strongly continuous if

$$\mathscr{C} := \left\{f\in C_b(E)\bigcap L^2(\mu) : \{P_tf\}_{t\in[0,1)} \text{ is uniformly integrable in } L^2(\mu)\right\}$$

is dense in $L^2(\mu)$.

Definition 0.2.4 Let $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ be a Dirichlet form on $L^2(\mu)$.

(1) For any $K\subset E$, let $\mathscr{D}(\mathscr{E})_K := \{f\in\mathscr{D}(\mathscr{E}) : f|_{K^c}=0\ \mu\text{-a.e.}\}$.

(2) An increasing sequence $\{K_n\}_{n\geqslant 1}$ of closed subsets of E is called an $\mathscr{E}$*-nest* if $\bigcup_{n\geqslant 1}\mathscr{D}(\mathscr{E})_{K_n}$ is dense in $\mathscr{D}(\mathscr{E})$ with respect to $\tilde{\mathscr{E}}_1^{1/2}$.

(3) A subset $N\subset E$ is called $\mathscr{E}$*-exceptional* if $N\subset\bigcap_{n\geqslant 1}K_n^c$ for some $\mathscr{E}$-nest $\{K_n\}_{n\geqslant 1}$.

(4) A function f is called $\mathscr{E}$*-quasi-continuous* if there exists an $\mathscr{E}$-nest $\{K_n\}_{n\geqslant 1}$ such that $f|_{K_n}$ is continuous for each $n\geqslant 1$.

We now introduce the notion of quasi-regular Dirichlet form which is associated with the *special standard process.*

Definition 0.2.5 A Dirichlet form $(\mathscr{E},\mathscr{D}(\mathscr{E}))$ on $L^2(\mu)$ is called *quasi-regular* if

(1) There exists an $\mathscr{E}$-nest $\{K_n\}_{n\geqslant 1}$ of compact sets.

(2) The set $\{f \in \mathscr{D}(\mathscr{E}) : f$ has a $\mathscr{E}$-quasi-continuous μ-version$\}$ is dense in $\mathscr{D}(\mathscr{E})$ with respect to $\tilde{\mathscr{E}}_1^{1/2}$.

(3) There exist an $\mathscr{E}$-exceptional set $N \subset E$ and $\{f_n\}_{n\geqslant 1} \subset \mathscr{D}(\mathscr{E})$ having $\mathscr{E}$-quasi-continuous μ-versions $\{\tilde{f}_n\}_{n\geqslant 1}$ separating the points of N^c.

Theorem 0.2.2 *Assume that $\sigma(C(E)) = \mathscr{F}$. A Dirichlet form $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ on $L^2(\mu)$ is quasi-regular if and only if it is associated with a μ-tight special standard process $\{\mathbb{P}^x : x \in E_\Delta\}$, i.e. the process is strongly Markovian with respect to $\{\mathscr{F}_t^+\}$ and satisfies*

(1) *For some (and hence all) $\nu \in \mathscr{P}(E_\Delta)$ equivalent to μ, if $\tau, \tau_n (n \geqslant 1)$ are $\mathscr{F}_t^+$-stopping times such that $\tau_n \to \tau$, then $x_{\tau_n} \to x_\tau (n \to \infty)$ P^ν-a.s. on $\{\tau < \xi\}$, and x_τ is $\bigvee_{n\geqslant 1} \mathscr{F}_{\tau_n}^+$-measurable. As mentioned above the filtration is understood as its completion with respect to $\mathbb{P}^\nu$.*

(2) *There exists an increasing sequence $\{K_n\}_{n\geqslant 1}$ of compact metrizable sets in E such that $\mathbb{P}^\mu(\lim_{n\to\infty} \sigma_{K_n^c} < \xi) = 0$, where $\sigma_{K_n^c} := \inf\{t \geqslant 0 : x_t \notin K_n\}$. In this case $P_t f$ is $\mathscr{E}$-quasi-continuous for all $t > 0$ and all $f \in \mathscr{B}_b(E) \bigcap L^2(\mu)$, i.e. the process is properly associated with $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$.*

Let E be a locally compact separable metric space with μ being a positive Radon measure of full support, i.e. μ are finite on compact sets and strictly positive on non-empty open sets. Let $C_0(E)$ be the set of all continuous real functions with compact supports. Then we call a symmetric Dirichlet form $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ on $L^2(\mu)$ *regular* if $\mathscr{D}(\mathscr{E}) \bigcap C_0(E)$ is a core, i.e. it is dense in $\mathscr{D}(\mathscr{E})$ under the norm $\mathscr{E}_1^{1/2}$ and dense in $C_0(E)$ under the uniform norm. Moreover, a strong Markov process is called a *Hunt process* if it is strongly Markovian with respect to $\{\mathscr{F}_t^+\}$ and *quasi-left continuous,* i.e. for any $\nu \in \mathscr{P}(E_\Delta)$ and any $\mathscr{F}_t^+$-stopping times $\tau, \tau_n (n \geqslant 1)$ with $\tau_n \to \tau$, one has $x_{\tau_n} \to x_\tau (n \to \infty)$ P^ν-a.s. on $\{\tau < \infty\}$. The following result is taken from [91, Theorem 7.2.1].

Theorem 0.2.3 *If $(\mathscr{E}, \mathscr{D}(\mathscr{E}))$ is a symmetric regular Dirichlet form on $L^2(\mu)$, then it is associated with a μ-symmetric Hunt process.*

0.3 Spectral theory

Let $(L, \mathscr{D}(L))$ be a linear operator on the complex Banach space $(\mathbb{B}, \|.\|)$.

Definition 0.3.1 Let $\rho(L)$ be the set of all $\lambda \in \mathbb{C}$ such that the range $\mathscr{R}(\lambda - L)$ of $\lambda - L$ is dense in $\mathbb{B}$ and $\lambda - L$ has a bounded inverse $R_\lambda := (\lambda - L)^{-1}$. We call $\rho(L)$ the *resolvent set* of L and R_λ the resolvent of L at λ for each $\lambda \in \rho(L)$. Moreover, $\sigma(L) := \mathbb{C} \setminus \rho(L)$ is called the *spectrum* of L and is decomposed into the following three disjoint parts:

(a) The *point spectrum* $\sigma_p(L)$ is the set of all $\lambda \in \mathbb{C}$ such that there exists $f \neq 0$ with $Lf = \lambda f$, i.e. $\lambda - L$ does not have any inverse. The element f is called an eigenvector of L corresponding to the eigenvalue λ.

(b) The *continuous spectrum* $\sigma_c(L)$ is the set of all $\lambda \in \mathbb{C}$ such that $\lambda - L$ has an unbounded densely defined inverse.

(c) The *residual spectrum* $\sigma_r(L)$ is the set of all $\lambda \in C$ such that $\lambda - L$ has an inverse whose domain is not dense.

Finally, the above notions of a linear operator on a real Banach space are defined by those of the complexification of the operator.

Proposition 0.3.1 *Let $(L, \mathscr{D}(L))$ be closed. Then for any $\lambda \in \rho(L)$, $R_\lambda := (\lambda - L)^{-1}$ is defined on the whole space, i.e. $\mathscr{R}(\lambda - L) = \mathbb{B}$.*

Definition 0.3.2 Let $(\mathbb{H}, \langle , \rangle)$ be a real or complex Hilbert space, and let $(L, \mathscr{D}(L))$ be a densely defined linear operator on $\mathbb{H}$.

(1) Let $\mathscr{D}(L^*)$ be the set of all $f \in \mathbb{H}$ such that there exists (and hence unique) $g \in \mathbb{H}$ satisfying $\langle Lh, f\rangle = \langle h, g\rangle,\ h \in \mathscr{D}(L)$. Let $L^*f := g$ for $f \in \mathscr{D}(L^*)$. We call $(L^*, \mathscr{D}(L^*))$ the *adjoint operator* of $(L, \mathscr{D}(L))$.

(2) $(L, \mathscr{D}(L))$ is called *symmetric* if $L^* \supset L$, i.e. $(L^*, \mathscr{D}(L^*))$ is an extension of $(L, \mathscr{D}(L))$.

(3) If $L^* = L$, then $(L, \mathscr{D}(L))$ is called *self-adjoint.* If L^* is self-adjoint, i.e., $L^{**} = L^*$, then $(L, \mathscr{D}(L))$ is called *essentially self-adjoint.*

Proposition 0.3.2 *Let $(L, \mathscr{D}(L))$ be a densely defined operator on $\mathbb{H}$. Then:*

(1) $(L^*, \mathscr{D}(L^*))$ is closed, hence any self-adjoint operator is closed.

(2) L is closable if and only if $\mathscr{D}(L^*)$ is dense, in this case $\bar{L} = L^{**}$ and $\bar{L}^* = L^*$.

Theorem 0.3.3 *Let $(L, \mathscr{D}(L))$ be a densely defined negative definite symmetric operator on $\mathbb{H}$. Then the quadric from $\mathscr{E}(f, g) := -\langle Lf, g\rangle$ defined on $\mathscr{D}(L)$ is closable and its closure is associated with a unique self-adjoint operator which is called the* Friedrichs extension *of $(L, \mathscr{D}(L))$. $(L, \mathscr{D}(L))$ is essentially self-adjoint if and only if its closure is self-adjoint, in this case the closure is the unique self-adjoint extension.*

We now consider the spectrum of a self-adjoint operator, which is determined by the resolution of identity (or the spectral family).

Definition 0.3.3 A family of projection operators $\{E_\lambda : \lambda \in \mathbb{R}\}$ on $\mathbb{H}$ is called a *resolution of the identity* if

(1) $E_{\lambda_1} E_{\lambda_2} = E_{\lambda_1 \wedge \lambda_2}, \qquad \lambda_1,\ \lambda_2 \in \mathbb{R}.$

(2) $E_{-\infty} f := \lim_{\lambda \to -\infty} E_\lambda f = 0$, $E_\infty f := \lim_{\lambda \to \infty} E_\lambda f = f, \qquad f \in \mathbb{H}.$

(3) $E_{\lambda+} = E_\lambda$, i.e. $\lim\limits_{\lambda'\to\lambda} E_{\lambda'}f = E_\lambda f, \qquad f \in \mathbb{H}$.

Proposition 0.3.4 *Let $\{E_\lambda : \lambda \in \mathbb{R}\}$ be a resolution of the identity. For any $f, g \in \mathbb{H}$, $\langle E.f, g\rangle$ is a boundedly variational function and hence $\mathrm{d}\langle E_\lambda f, g\rangle$ is a signed measure on $\mathbb{R}$. In particular, if $\|f\| = 1$ then $\mathrm{d}\|E_\lambda(f)\|^2 := \mathrm{d}\langle E_\lambda f, f\rangle$ is a probability measure.*

Theorem 0.3.5 *$(L, \mathscr{D}(L))$ is a self-adjoint operator on $\mathbb{H}$ if and only if there exists a unique resolution of the identity $\{E_\lambda : \lambda \in \mathbb{R}\}$ such that*

$$Lf = \int_{-\infty}^{\infty} \lambda \mathrm{d}E_\lambda f, \qquad \mathscr{D}(L) = \left\{f \in \mathbb{H} : \int_{-\infty}^{\infty} \lambda \mathrm{d}E_\lambda f \text{ exists in } \mathbb{H}\right\}, \quad (0.3.1)$$

where $\int_{-\infty}^{\infty} \lambda \mathrm{d}E_\lambda f$ is defined as the Riemann integral on $\mathbb{H}$.

Remark 0.3.1 Let F be a complex-valued Borel function on $\mathbb{R}$. Then $\int_{-\infty}^{\infty} F(\lambda)\mathrm{d}E_\lambda f$ exists if and only if $\int_{-\infty}^{\infty} |F(\lambda)|^2 \mathrm{d}\|E_\lambda f\|^2 < \infty$. In particular, if F is a real-valued continuous function, then $(F(L), \mathscr{D}(F(L)))$ defined below is a self-adjoint operator:

$$F(L)f := \int_{-\infty}^{\infty} F(\lambda)\mathrm{d}E_\lambda f, \qquad \mathscr{D}(F(\lambda)) := \left\{f : \int_{-\infty}^{\infty} F(\lambda)^2 \mathrm{d}\|E_\lambda f\|^2 < \infty\right\}.$$

Definition 0.3.4 The resolution of the identity $\{E_\lambda : \lambda \in \mathbb{R}\}$ satisfying (0.3.1) is called the *spectral family* of $(L, \mathscr{D}(L))$.

Theorem 0.3.6 *Let $(L, \mathscr{D}(L))$ be a self-adjoint operator on $\mathbb{H}$. Then:*

(1) *$\sigma(L) \subset \mathbb{R}$ and $\sigma_r(L) = \varnothing$.*

(2) *$\lambda \in \sigma_p(L)$ if and only if $E_\lambda \neq E_{\lambda-}$, and the eigenspace of λ, i.e. the space spanned by eigenvectors with eigenvalue λ, is $\mathscr{R}(E_\lambda - E_{\lambda-})$.*

(3) *$\lambda \in \sigma_c(L)$ if and only if $E_\lambda = E_{\lambda-}$ and $E_{\lambda_1} \neq E_{\lambda_2}$ for any $\lambda_1 < \lambda < \lambda_2$. In other words, $\sigma_c(L) = \sigma(L) \setminus \sigma_p(L)$.*

(4) *For any real continuous function F, one has $\sigma(F(L)) = \overline{F(\sigma(L))}$. If moreover F is strictly monotonic then $\sigma_p(F(L)) = F(\sigma_p(L))$.*

Theorem 0.3.6 (4) is called the spectral mapping theorem, which follows immediately from the definition of $F(L)$. Indeed, for any $\lambda_0 \in \sigma(L)$, there exists $\{f_n\} \subset \mathscr{D}(L)$ with $\|f_n\| = 1$ such that the support of $\mathrm{d}E_\lambda f_n$ is contained in $\left[\lambda_0 - \frac{1}{n}, \lambda_0 + \frac{1}{n}\right]$. Then it is trivial to see that $\|F(L)f_n - F(\lambda_0)f_n\| \leqslant \sup_{|\lambda-\lambda_0|\leqslant n^{-1}} |F(\lambda) - F(\lambda_0)| \to 0$ as $n \to \infty$. Thus, $F(\lambda_0) \in \sigma(F(L))$. Since the spectrum is closed, we obtain $\sigma(F(L)) \supset \overline{F(\sigma(L))}$. On the other hand, if $\lambda_0 \notin \overline{F(\sigma(L))}$, then there exists $\varepsilon > 0$ such that $|F(\lambda) - \lambda_0| > \varepsilon$ for all $\lambda \in \sigma(L)$. Therefore, for any $f \in \mathbb{H}$ with $\|f\| = 1$,

$$\|F(L)f - \lambda_0 f\|^2 = \int_{\sigma(L)} (F(\lambda) - \lambda_0)^2 \mathrm{d}\|E_\lambda(f)\|^2 \geqslant \varepsilon^2.$$

Thus, $\lambda_0 \notin \sigma(F(L))$. The second assertion of (4) holds since when F is strictly monotonic, λ is an eigenvalue of L if and only if so is $F(\lambda)$ of $F(L)$.

Definition 0.3.5 Let $(L, \mathscr{D}(L))$ be a self-adjoint operator on $\mathbb{H}$. We write $\lambda \in \sigma_{\text{ess}}(L)$, the *essential spectrum* of L, if for any $\varepsilon > 0$ the closure of the range of $\mathbf{1}_{(\lambda-\varepsilon,\lambda+\varepsilon)}(L)$ is infinite dimensional. We call $\sigma_d(L) := \sigma(L) \setminus \sigma_{\text{ess}}(L)$ the *discrete spectrum* of L.

Proposition 0.3.7 *Let $(L, \mathscr{D}(L))$ be a self-adjoint operator on $\mathbb{H}$.*

(1) *$\lambda \in \sigma_d(L)$ if and only if λ is isolated in $\sigma(L)$ and is an eigenvalue with finite multiplicity, i.e. its eigenspace is finite dimensional.*

(2) *$\lambda \in \sigma_{\text{ess}}(L)$ if and only if λ is either a limit point in $\sigma(L)$ or an eigenvalue with infinite multiplicity.*

The following *Weyl's criterion* follows immediately from the definitions of $\sigma(L)$ and $\sigma_{\text{ess}}(L)$ but is very useful in applications (see [155, VII.12]).

Theorem 0.3.8 (Weyl's Criterion) *Let $(L, \mathscr{D}(L))$ be a self-adjoint operator on $\mathbb{H}$.*

(1) *$\lambda \in \sigma(L)$ if and only if for any $\varepsilon > 0$ there exists a unit $f \in \mathscr{D}(L)$ such that $\|Lf - \lambda f\| < \varepsilon$.*

(2) *$\lambda \in \sigma_{\text{ess}}(L)$ if and only if for any $\varepsilon > 0$ there exists an orthonormal sequence $\{f_n\}_{n\geqslant 1} \subset \mathscr{D}(L)$ such that $\|Lf_n - \lambda f_n\| < \varepsilon$ for any $n \geqslant 1$.*

To define the discrete spectrum for general closed operators on $\mathbb{B}$, let $\lambda \in \sigma(L)$ be isolated, i.e. there exists $\varepsilon > 0$ such that $\{z \in \mathscr{C} : |z - \lambda| < \varepsilon\} \bigcap \sigma(L) = \{\lambda\}$. Then (see [155, Theorem XII.5]) for any $r \in (0, \varepsilon)$,

$$P_\lambda := -\frac{1}{2\pi \mathrm{i}} \oint_{|z-\lambda|=r} (L - z)^{-1} \mathrm{d}z$$

exists and is independent of r. Moreover, P_λ is a projection, i.e. $P_\lambda^2 = P_\lambda$.

Definition 0.3.6 We write $\lambda \in \sigma_d(L)$, the *discrete spectrum* of a closed operator $(L, \mathscr{D}(L))$, if $\lambda \in \sigma(L)$ is isolated and the range of P_λ is finite-dimensional. We call $\sigma_{\text{ess}}(L) := \sigma(L) \setminus \sigma_d(L)$ the essential spectrum of L.

The following result provides some useful descriptions of the essential spectrum (see [114, Theorem 3.6.29]), where the first assertion is called the Weyl theorem (see [155, XIII.4]). A linear operator P on $\mathbb{B}$ is called *compact* if it sends bounded sets onto relatively compact sets.

Theorem 0.3.9 (1) *If $\mathbb{B} = \mathbb{H}$ and $(L, \mathscr{D}(L))$ is self-adjoint, then for any compact operator $\tilde{L}$, $L + \tilde{L}$ is well-defined on $\mathscr{D}(L)$ and $\sigma_{\text{ess}}(L + \tilde{L}) = \sigma_{\text{ess}}(L)$.*

(2) Let $(L, \mathscr{D}(L))$ be a closed operator generating a C_0-contraction semigroup $\{P_t\}_{t \geqslant 0}$ on $\mathbb{B}$. If $\mathbb{B} = \mathbb{H}$ and L is self-adjoint, then the following are equivalent.

(i) *$\sigma_{\text{ess}}(L) = \varnothing$, i.e. the spectrum of L is discrete.*

(ii) *$R_\lambda := (\lambda - L)^{-1}$ is compact for some $\lambda \in \rho(L)$.*

(iii) *R_λ is compact for all $\lambda \in \rho(L)$.*

(iv) *P_t is compact for all $t > 0$.*

(v) *P_t is compact for some $t > 0$.*

In general, (iv) *implies* (iii), (iii) *is equivalent to* (ii) *and implies* (i)*.*

In general, a C_0-contraction semigroup P_t on a Banach space satisfies (iv) if and only if (iii) holds and P_t is *equicontinuous*, that is, $\|P_{t+s} - P_t\| \to 0$ as $s \to 0$ for all $t > 0$, see e.g. [181, Theorem 6.2.1].

Since the compactness of P_t (or R_λ) is crucial for the generator to have discrete spectrum, we focus on the analysis of compact operators. First, let us introduce the notion of the spectral radius and the essential spectral radius.

Definition 0.3.7 Let P be a bounded linear operator on $\mathbb{B}$. We call $r(P) := \sup\{|\lambda| : \lambda \in \sigma(P)\}$ the *spectral radius* and $r_{\text{ess}}(P) := \sup\{|\lambda| : \lambda \in \sigma_{\text{ess}}(P)\}$ the *essential spectral radius* of P.

It is clear that

$$r(P) = \lim_{n\to\infty} \|P^n\|^{1/n} = \inf_{n \geqslant 1} \|P^n\|^{1/n}. \tag{0.3.2}$$

Next, by Theorem 0.3.9(1), the essential spectral radius of a compact operator is zero. This leads us to search for a formula of $r_{\text{ess}}(P)$ analogous to (0.3.2) by using the measure of noncompactness instead of the operator norm.

Definition 0.3.8 For a set $D \subset \mathbb{B}$, the quantity

$$\beta(D) := \inf\left\{ r > 0 : \text{there exist } f_1, \cdots, f_n \in \mathbb{B} \text{ such that } D \subset \bigcup_{i=1}^n B(f_i, r) \right\}$$

is called the *measure of noncompactness* of D, where $B(f_i, r) := \{f : \|f - f_i\| < r\}$. For a linear operator P we call $\beta(P) := \beta(PB(0,1))$ its measure of noncompactness.

The following theorem due to [149] (see also [145]) provided a formula of $r_{\text{ess}}(P)$.

Theorem 0.3.10 *For any bounded linear operator P on $\mathbb{B}$ (i.e. P sends bounded sets onto bounded sets), we have*

$$r_{\text{ess}}(P) = \lim_{n\to\infty} \beta(P^n)^{1/n} = \inf_{n\geqslant 1} \beta(P^n)^{1/n}. \tag{0.3.3}$$

We now consider $\mathbb{B} := L^p_{\mathscr{C}}(\mu)$ ($p \in [1,\infty)$), the complexification of $L^p(\mu) := \{f \in \mathbb{B}(E) : \mu(|f|^p) < \infty\}$ for a σ-finite measure space $(E, \mathscr{F}, \mu)$. In this case it is trivial to see that if P is compact then for any $g \in L^p_{\mathscr{C}}(\mu)$ with $|g| > 0$, one has

$$\lim_{r\to\infty} \sup_{\|f\|_p\leqslant 1} \mu(|Pf|^p \mathbf{1}_{\{|Pf|>r|g|\}}) = 0, \tag{0.3.4}$$

where $\|\cdot\|_p$ is the norm in $L^p_{\mathscr{C}}(\mu)$. This leads to the following notion of semicompactness.

Definition 0.3.9 (1) The following quantity is called the *measure of non-semicompactness* of a set $D \subset L^p_{\mathscr{C}}(\mu)$:

$$\rho(D) := \inf_{g\in L^p_{\mathscr{C}}(\mu)} \sup_{f\in D} \|f\mathbf{1}_{\{|f|>|g|\}}\|_p,$$

and $\rho(P) := \rho(PB(0,1))$ is called the *measure of non-semicompactness* of P. Moreover, D (resp. P) is called semicompact if $\rho(D) = 0$ (resp. $\rho(P) = 0$).

(2) D is called order bounded in $L^p_{\mathscr{C}}(\mu)$ if there exists nonnegative $u \in L^p(\mu)$ such that $D \subset [-u, u] := \{f \in L^p_{\mathscr{C}}(\mu) : |f| \leqslant u\}$. An operator is called order bounded if it sends order bounded sets onto order bounded sets.

(3) An operator P is called AM-compact if it sends semicompact sets onto relatively compact sets, or equivalently sends order bounded sets onto relatively compact sets.

The following result is due to [183] (see also [71] for the context of Banach lattice).

Theorem 0.3.11 *If P is order bounded and AM-compact on $L^p_{\mathscr{C}}(\mu)$ then $\beta(P) = \rho(P)$. Consequently, an order bounded operator is compact if and only if it is semicompact and AM-compact.*

In applications, it is convenient to rewrite the quantity ρ in the following simple way. We leave the proof as exercise.

Proposition 0.3.12 *For any strictly positive $\phi \in L^p(\mu)$ (which exists since μ is σ-finite), one has $\rho(D) = \lim_{r\to\infty} \sup_{f\in D} \|f\mathbf{1}_{\{|f|>r\phi\}}\|_p$. Consequently, $\rho(P)$ is equal to the following L^p-tail norm of P: $\|P\|_{p,T} = \lim_{r\to\infty} \sup_{\|f\|_p\leqslant 1} \|(Pf)\mathbf{1}_{\{|Pf|>r\phi\}}\|_p$.*

We recall the following two interpolation theorems which are crucial in the analysis of semigroups.

Theorem 0.3.13 (1) Riesz-Thorin's interpolation theorem *Let* $1 \leqslant p_1, p_2, q_1, q_2 \leqslant \infty$ *and let* P *be a linear operator from* $L^{p_i}(\mu)$ *to* $L^{q_i}(\mu)$ *with* $\|P\|_{p_i \to q_i} < \infty$, $i = 1, 2$. *Then for any* $r \in (0,1)$ *and* p_0, q_0 *satisfying*

$$\frac{1}{p_0} = \frac{r}{p_1} + \frac{1-r}{p_2}, \qquad \frac{1}{q_0} = \frac{r}{q_1} + \frac{1-r}{q_2},$$

we have

$$\|P\|_{p_0 \to q_0} \leqslant \|P\|_{p_1 \to q_1}^r \|P\|_{p_2 \to q_2}^{1-r}.$$

(2) Stein's interpolation theorem *Let* p_i, q_i *and* r *be as above. Let* $S := \{z \in \mathbb{C} : \operatorname{Re} z \in [0,1]\}$ *and* A_z *a linear operator from* $L^{p_1}(\mu) \bigcap L^{p_2}(\mu)$ *to* $L^{q_1}(\mu) \bigcup L^{q_2}(\mu)$ *for each* $z \in S$. *Assume that for any* $f \in L^{p_1}(\mu) \bigcap L^{p_2}(\mu)$ *and* $g \in L^{p_1'}(\mu) \bigcap L^{p_2'}(\mu)$, *where* p_i' *is the conjugate number to* $p_i, i = 1, 2$, *the function* $\langle A.f, g \rangle$ *is uniformly bounded and continuous on* S *and analytic in* $\mathring{S}$. *If for any* $f \in L^{p_1}(\mu) \bigcap L^{p_2}(\mu)$ *and* $y \in \mathbb{R}$ *one has* $\|A_{iz} f\|_{q_1} \leqslant M_1 \|f\|_{p_1}$ *and* $\|A_{1+iy} f\|_{q_2} \leqslant M_2 \|f\|_{p_2}$, *then*

$$\|A_r f\|_{q_0} \leqslant M_1^r M_2^{1-r} \|f\|_{p_0}, \qquad f \in L^{p_1}(\mu) \bigcap L^{p_2}(\mu).$$

Assume that P_t is the semigroup of a self-adjoint operator L on $L^2(\mu)$. Then for any $t > 0$, the operator $A_z := \mathrm{e}^{ztL}$ is analytic in $L^2(\mu)$ and hence the above Stein's interpolation theorem works as follows: if $\|P_t\|_{2 \to p} < \infty$ for some $p > 2$ and $t > 0$, then $\|P_{rt}\|_{2 \to p(r)} \leqslant \|P_t\|_{2 \to p}^r$, $p(r) := 2p/(2r + p(1-r))$.

Finally, we introduce the following result concerning the weak* compactness for compact sets on a Banach space.

Theorem 0.3.14(Bourbaki) *The unit ball of the dual space* $\mathbb{B}'$ *of a separable Banach space* $\mathbb{B}$ *is compact and metrizable with respect to the topology* $\sigma(\mathbb{B}', \mathbb{B})$. *Consequently, for any* $p \in (1, \infty]$ *and any bounded sequence* $\{f_n\} \subset L^p(\mu)$, *there exists* $f \in L^p(\mu)$ *and a subsequence* $\{f_{n'}\}$ *such that* $\mu(f_{n'} g) \to \mu(fg)$ *for any* $g \in L^q(\mu)$, *where* $q \geqslant 1$ *satisfies* $q^{-1} + p^{-1} = 1$.

Proof. The first assertion is well-known so that the second follows immediately as soon as $L^q(\mu)$ is separable. If $L^q(\mu)$ is not separable, let us consider the sub-σ-field $\mathscr{C} := \sigma(f_n : n \geqslant 1)$. Then $L^q(\mu|_{\mathscr{C}})$ is separable for any $q \geqslant 1$. Thus, there exists $f \in L^q(\mu|_{\mathscr{C}}) \subset L^q(\mu)$ and a subsequence $\{f_{n'}\}$ such that $\mu(f_{n'} g) \to \mu(fg)$ for any $g \in L^q(\mu|_{\mathscr{C}})$. Therefore, for any $g \in L^1(\mu)$, letting $g' := \mu(g|\mathscr{C})$ be the conditional expectation given $\mathscr{C}$, we obtain $\mu(f_{n'} g) = \mu(f_{n'} g') \to \mu(fg') = \mu(fg)$. □

0.4 Riemannian geometry

Let M be a Hausdorff topological space with a countable basis of open sets. For each open set $U \subset M$, if $\varphi : U \to \mathbb{R}^d$ is one-to-one and $\varphi(U)$ is open, then (U, φ) is called a *coordinate neighborhood* on M. A *d-dimensional differential structure* on M is a family $\mathscr{U} := \{(U_\alpha, \varphi_\alpha)\}$ of coordinate neighborhoods such that

(i) $\bigcup_\alpha U_\alpha \supset M$,

(ii) For any $\alpha,\ \beta,\ \varphi_\alpha \circ \varphi_\beta^{-1} :\ \varphi_\beta(U_\beta \bigcap U_\alpha) \to \varphi_\alpha(U_\alpha \bigcap U_\beta)$ is C^∞-smooth, i.e. $(U_\alpha, \varphi_\alpha)$ and (U_β, φ_β) are C^∞*-compatible*,

(iii) If a coordinate neighborhood (U, φ) is C^∞-compatible with each $(U_\alpha, \varphi_\alpha)$ in $\mathscr{U}$, then $(U, \varphi) \in \mathscr{U}$.

If M is equipped with a differential structure, then it is called a d-dimensional *differential manifold*, and each $(U, \varphi) \in \mathscr{U}$ is called a *local* (*coordinate*) *chart*.

A function $f : M \to \mathbb{R}$ is called C^p-smooth, if for any $(U, \varphi) \in \mathscr{U}$ so is $f \circ \varphi^{-1} : \varphi(U) \to \mathbb{R}$. Let $C^p(M)$ denote the set of all C^p-smooth functions on M, and $C_0^p(M)$ the set of such functions with compact supports. Given $x \in M$, let $C^\infty(x)$ be the set of C^∞-functions defined in a neighborhood of x.

Definition 0.4.1 Let M be a differential manifold. The *tangent space* T_xM at a point $x \in M$ is the set of all mappings $X :\ C^\infty(x) \to \mathbb{R}$ satisfying

(i) $X(c_1 f + c_2 g) = c_1 Xf + c_2 Xg, \qquad f,\ g \in C^\infty(x),\ c_1,\ c_2 \in \mathbb{R}^d,$

(ii) $X(fg) = (Xf)g(x) + f(x)Xg, \qquad f,\ g \in C^\infty(x).$

Obviously, T_xM is a vector space by the convention:

$$(X+Y)f := Xf + Yf, \quad (cX)f := c(Xf), \qquad c \in \mathbb{R},\ f \in C^\infty(x).$$

Let $x \in U$ with $(U, \varphi) \in \mathscr{U}$, then for any vector Z at $\varphi(x)$ on $\mathbb{R}^d$, one may define $\varphi^* Z \in T_xM$ by

$$(\varphi^* Z)f := Z(f \circ \varphi^{-1}), \qquad f \in C^\infty(x).$$

Let $(u_1, \cdots, u_d)$ be the Euclidean coordinate on $\varphi(U)$, and let $\dfrac{\partial}{\partial x_i} := \varphi^* \dfrac{\partial}{\partial u_i}$, $1 \leqslant i \leqslant d$. For any $X \in T_xM$ one has $X = \varphi^* \varphi_* X$, where $\varphi_* X$ is a vector at $\varphi(x)$ satisfying

$$(\varphi_* X)g := X(g \circ \varphi), \qquad g \in C^\infty(\varphi(x)).$$

Then

$$X = \varphi^* \left(\sum_{i=1}^d \langle \varphi_* X, \frac{\partial}{\partial u_i} \rangle_{\mathbb{R}^d} \frac{\partial}{\partial u_i} \right) = \sum_{i=1}^d \langle \varphi_* X, \frac{\partial}{\partial u_i} \rangle_{\mathbb{R}^d} \frac{\partial}{\partial x_i}.$$

Therefore, $\dfrac{\partial}{\partial x_i}$ is a basis of T_xM.

Now, let $TM := \bigcup_{x\in M} T_xM$, which is called the *vector bundle* on M. A *vector field* on M is a mapping

$$X : M \to TM, \qquad X_x \in T_xM,\ x \in M.$$

Let $\Gamma(TM)$ be the set of all vector fields on M. A vector field X is called C^p-smooth if in any local chart there exist C^p-smooth functions $f_1, \cdots, f_d$ such that

$$X = \sum_{i=1}^{d} f_i \frac{\partial}{\partial x_i}.$$

Let $\Gamma^p(TM)$ denote the set of all C^p-vector fields.

Definition 0.4.2 Let M be a differential manifold. A mapping $\nabla : TM \times \Gamma^1(TM) \to TM$ is called a *connection* on M, if it is bilinear and $\nabla_X Y := \nabla(X, Y)$ has the following properties:

(i) If $X \in T_xM$ and $Y \in \Gamma^1(TM)$, then $\nabla_X Y \in T_xM$;

(ii) For any $f \in C^1(M)$, $\nabla_X(fY) = (Xf)Y_x + f(x)\nabla_X Y,\ X \in T_xM,\ x \in M,\ Y \in \Gamma^1(TM)$.

Definition 0.4.3 Let M be a differential manifold. For each $x \in M$, let g_x be an inner product on the vector space T_xM. If for any local chart (U, φ) and any $X, Y \in \Gamma^\infty(TM)$, $g_x(X_x, Y_x)$ is C^∞-smooth in x, then g is called a *Riemannian metric* on M. A differential manifold equipped with a Riemannian metric is called a *Riemannian manifold*.

It is clear that under a local chart (U, φ) a Riemannian metric has the representation

$$g\Big(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\Big) = g_{ij}(x), \qquad 1 \leqslant i,\ j \leqslant d,$$

so that $g_{ij} \in C^\infty$ and $(g_{ij}(x))$ is strictly positive definite at each $x \in U$. Moreover, the Riemannian metric determines a unique measure such that for any local chart (U, φ),

$$\mathrm{vol}(A) = \int_{\varphi(A)} \sqrt{\det g \circ \varphi^{-1}(u)}\, \mathrm{d}u, \qquad A \subset U.$$

We call this measure the *volume measure* of M and simply denote it by $\mathrm{d}x$.

Theorem 0.4.1 (Levi-Civita) *If M is a Riemannian manifold, then there exists a unique connection ∇ (called Levi-Civita connection) satisfying*

$$X\langle Y,Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle, \qquad \nabla_X Y = \nabla_Y X + [X,Y]$$

for all $X, Y, Z \in \Gamma^1(TM)$, where $\langle , \rangle$ denotes the inner product under the Riemannian metric and $[X,Y] := XY - YX$.

Throughout the book, we only use the Levi-Civita connection. It is useful to note that $[X,Y]$ is a vector field for any $X, Y \in \Gamma^1(TM)$. A mapping $\gamma : [\alpha,\beta] \to M$ is called a C^p-curve on M if it is continuous and for any local chart (U,φ), $\varphi \circ \gamma : [\alpha,\beta] \bigcap \gamma^{-1}(U) \to \mathbb{R}^d$ is C^p-smooth. For a C^1-curve γ, we may define the *tangent vector* along γ by

$$\dot\gamma_t f := \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma_t), \qquad f \in C^\infty(\gamma_t).$$

Definition 0.4.4 (1) Let $\gamma : [\alpha,\beta] \to M$ be a C^1-curve on M. A vector field X is said to be constant (or parallel) along γ if $\nabla_{\dot\gamma_t} X = 0$ for $t \in [\alpha,\beta]$. Given $V \in T_{\gamma_\alpha} M$, there exists a unique constant vector field X along γ satisfying $X_{\gamma_\alpha} = V$. We call this vector field the *parallel transportation* of V along γ. A C^2-curve γ is called *geodesic* if $\nabla_{\dot\gamma}\dot\gamma = 0$.

(2) For any $x \in M$ and any $X \in T_x M$, $X \neq 0$, there exists a unique geodesic $\gamma : [0,\infty) \to M$ such that $\gamma_0 = x$ and $\dot\gamma_0 = X$. We denote $\gamma_t := \exp_x(tX)$ and call $\exp_x : T_x M \to M$ the *exponential map* at x. By convention we set $\exp_x(0) = x$.

For any $x \neq y$, one may define the Riemannian distance between x and y by

$$\rho(x,y) := \inf\left\{ \int_0^1 |\dot\gamma_s| \mathrm{d}s : \gamma : [0,1] \to M \text{ is a } C^1\text{-curve such that } \gamma_0 = x \text{ and } \gamma_1 = y \right\},$$

where $|X| = \langle X, X\rangle^{1/2} := g(X,X)^{1/2}$. It is classical that $\rho(x,y)$ can be reached by a geodesic. On the other hand, however, geodesics linking two points may be not unique. Thus, the one with length $\rho(x,y)$ is called the *minimal geodesic.*

In many cases, the minimal geodesic is still not unique. For instance, for the unit sphere $\mathbb{S}^d$, each half circle linking the highest and the lowest points is a minimal geodesic. This fact leads to the following notion of cut-locus.

Definition 0.4.5 Let $x \in M$. For any $X \in \mathbb{S}_x := \{X \in T_x M : |X| = 1\}$, let

$$r(X) := \sup\{t > 0 : \rho(x, \exp_x(tX)) = t\}.$$

If $r(X) < \infty$ then we call $\exp_x(r(X)X)$ a cut-point of x. The set

$$\mathrm{cut}(x) := \{\exp_x(r(X)X) : X \in \mathbb{S}_x,\ r(X) < \infty\}$$

is called the *cut-locus* of the point x. Moreover, the quantity

$$i_x := \inf\{r(X) : X \in \mathbb{S}_x\}$$

is called the *injectivity radius* of x. Finally, we call $i_M := \inf_{x\in M} i_x$ the injectivity radius of M.

The following result summarizes some property of the cut-locus.

Theorem 0.4.2 (1) $\mathrm{cut}(x)$ *is closed and has volume zero.*
(2) $\rho(x,\cdot)$ *is* C^∞*-smooth on* $M \setminus (x \bigcup \mathrm{cut}(x))$.
(3) $i_x > 0$ *for any* $x \in M$ *and the function* $i : M \to (0,\infty]$ *is continuous.*
(4) *The set* $D_x := \exp_x^{-1}(M \setminus \mathrm{cut}(x))$ *is starlike in* T_xM *and*

$$\exp_x : D_x \to \exp_x(D_x)$$

is a diffeomorphism. Consequently, if $y \notin \mathrm{cut}(x)$ *then the minimal geodesic linking* x *and* y *is unique.*

We now introduce the curvature on M. For any $X, Y, Z \in \Gamma^2(TM)$, let

$$\mathscr{R}(X,Y)Z := \nabla_Y\nabla_X Z - \nabla_X\nabla_Y Z - \nabla_{[X,Y]}Z.$$

For any $X_x, Y_x, Z_x \in T_xM$, let X, Y, Z be their smooth extensions respectively. Then the value of $\mathscr{R}(X,Y)Z$ at point x is independent of the choice of extensions and hence, $\mathscr{R}$ is a well-defined tensor which is called the *curvature* tensor of the connection ∇.

The curvature tensor satisfies the following identities:

$$\mathscr{R}(X,Y)Z + \mathscr{R}(Y,X)Z = 0, \tag{0.4.1}$$

$$\mathscr{R}(X,Y)Z + \mathscr{R}(Z,X)Y + \mathscr{R}(Y,Z)X = 0, \tag{0.4.2}$$

$$\langle \mathscr{R}(X,Y)Z, V\rangle = \langle \mathscr{R}(Z,V)X, Y\rangle = -\langle \mathscr{R}(X,Y)V, Z\rangle. \tag{0.4.3}$$

Definition 0.4.6 (1) For $X, Y \in T_xM$, the quantity

$$\mathrm{Sect}(X,Y) := \frac{\langle \mathscr{R}(X,Y)X, Y\rangle}{|X|^2|Y|^2 - \langle X, Y\rangle^2}$$

is called the *sectional curvature* of the plane spanned by X and Y. If X is parallel to Y then we set $\mathrm{Sect}(X,Y) = 0$.

(2) Let $\{W_i\}_{i=1}^d$ be an orthonormal basis on T_xM. The quantity

$$\mathrm{Ric}(X,Y) := \sum_{i=1}^d \langle \mathscr{R}(X,W_i)Y, W_i\rangle$$

is independent of the choice of $\{W_i\}$ and Ric is called the *Ricci curvature tensor*.

(3) Let γ be a geodesic. A smooth vector field J is called a *Jacobi field* along γ if

$$\nabla_{\dot\gamma}\nabla_{\dot\gamma} J = -\mathscr{R}(\dot\gamma, J)\dot\gamma.$$

This equation is called the *Jacobi equation.*

Since the Jacobi equation is a second order ordinary equation, given $X, Y \in T_{\gamma_0}M$, there exists a Jacobi field along γ such that $J_0 = X$, $\nabla_{\dot\gamma_t} J_t|_{t=0} = Y$. Moreover, let $\gamma : [0,t] \to M$ be a geodesic, for any $X \in T_{\gamma_0}M$ and $Y \in T_{\gamma_t}M$, there exists a Jacobi field J along γ satisfying $J_0 = X$ and $J_t = Y$. Concerning the uniqueness of Jacobi fields, we introduce the notion of conjugate points.

Definition 0.4.7 Let $x \in M$, a point $y \in M$ is called a conjugate point of x, if there exists a nontrivial Jacobi field J along a minimal geodesic linking x and y such that J vanishes at x and y.

Proposition 0.4.3 $\mathrm{cut}(x)$ *consists of conjugate points of* x *and points having more than one minimal geodesics to* x.

To make analysis on Riemannian manifolds, let us introduce some fundamental operators including the divergence, the gradient and the Laplace operators.

Definition 0.4.8 Let $X \in \Gamma^1(TM)$, we define its *divergence* by

$$(\mathrm{div}X)(x) := (\mathrm{tr}\nabla X)(x) = \sum_{i=1}^d \langle \nabla_{W_i} X, W_i\rangle,$$

where $\{W_i\}$ is an orthonormal basis of T_xM. It is easy to check that $\mathrm{div}X$ is independent of the choice of $\{W_i\}$. For $f \in C^1(M)$, define its *gradient* $\nabla f \in \Gamma(TM)$ by

$$\langle \nabla f, X\rangle := Xf, \qquad X \in \Gamma(TM).$$

Finally, the *Laplace operator* is defined by $\Delta := \mathrm{div}\nabla$ which well acts on C^2-functions.

For any $f \in C^2(M)$, define $\mathrm{Hess}_f(X,Y) := \langle \nabla_X \nabla f, Y\rangle$ for $X, Y \in T_xM, x \in M$. We call Hess_f the *Hessian tensor* of f. It is clear that the Hessian tensor is symmetric. The following Bochner-Weitzenböck formula connects the Ricci curvature with the Laplace and gardient operators and the Hessian tensor.

Proposition 0.4.4 (Bochner-Weitzenböck formula) *For any smooth function f one has*

$$\frac{1}{2}\Delta|\nabla f|^2 - \langle \nabla\Delta f, \nabla f\rangle = \|\mathrm{Hess}_f\|_{HS}^2 + \mathrm{Ric}(\nabla f, \nabla f),$$

where $\|\mathrm{Hess}_f\|_{HS}^2(x) := \sum_{i,j=1}^d \mathrm{Hess}_f(E_i, E_j)^2 = \sum_{i,j=1}^d (E_iE_jf(x))^2$ *for an orthonormal basis* $\{E_i\}$ *of* T_xM *such that* $\nabla E_i = 0$ *at* x *for all* i*, which is called a normal frame at* x*.*

Now, we introduce some useful integral formulae for the above operators. For given $f \in C_b^\infty(M)$, by Sard's theorem the set of critical values in $f(M)$ has Lebesgue measure zero. In other words, $\{f = t\}$ is a $(d-1)$-dimensional submanifold of M for a.e. $t \in f(M)$. Let $\boldsymbol{A}$ denote the volume measure on a $(d-1)$-dimensional submanifold of M with the induced metric.

Theorem 0.4.5 (Coarea formula) *For any* $f \in C_b^\infty(M)$ *and any* $h \in L^1(\mathrm{d}x)$*,*

$$\int_M h|\nabla f|\mathrm{d}x = \int_0^\infty \mathrm{d}t \int_{\{f=t\}} h\mathrm{d}\boldsymbol{A}.$$

In particular, if $\mathrm{d}\mu := h\mathrm{d}x$ *is a finite measure and let* $\mathrm{d}\mu_\partial := h\mathrm{d}\boldsymbol{A}$ *then*

$$\mu(|\nabla f|) = \int_0^\infty \mu_\partial(\{f = t\})\mathrm{d}t.$$

Theorem 0.4.6 (Green formula or integration by parts formula) (1) *If* $X \in \Gamma^1(TM)$ *with compact support, then*

$$\int_M \mathrm{div}X(x)\mathrm{d}x = 0.$$

(2) *If* $f, g \in C_0^2(M)$*, then*

$$\int_M (f\Delta g)(x)\mathrm{d}x = \int_M (g\Delta f)(x)\mathrm{d}x = -\int_M \langle\nabla f, \nabla g\rangle(x)\mathrm{d}x.$$

(3) *Let* $X \in \Gamma^1(TM)$ *and* $D \subset M$ *a smooth open domain, i.e. an open domain with boundary a* $(d-1)$*-dimensional differential manifold. Then*

$$\int_D (\mathrm{div}X)(x)\mathrm{d}x = \int_{\partial D} \langle X, N\rangle \mathrm{d}\boldsymbol{A},$$

where N *is the outward unit normal vector field on* ∂D*.*

(4) *For a smooth open domain* D*,*

$$\int_D (f\Delta g + \langle\nabla f, \nabla g\rangle)(x)\mathrm{d}x = \int_{\partial D} f(Ng)\mathrm{d}\boldsymbol{A}, \qquad f \in C_0^1(\bar{D}), g \in C_0^2(\bar{D}).$$