STOCHASTIC KINEMATICS

Gregory S. Chirikjian

Johns Hopkins University

Tutorial, ICRA'09

Kobe, Japan

Based on "Stochastic Models, Information Theory, and Lie Groups", Birkhäuser, 2009

Contents

First Hour:

- Introduction and Overview of Problem Domains
- Review of Gaussian Distributions
- Basic Theories of Probability and Information
- Stochastic Differential Equations in Euclidean Space

Second Hour:

- Geometry of Curves, Surfaces and Manifolds
- Stochastic Differential Equations on Curved Spaces
- Lie Groups and Integral Geometry
- Fourier Analysis on Groups

Third Hour:

- Applications in Mobile Robotics
- Applications to Manipulators
- Applications in Molecular Biophysics

2

Introduction

2.1 What is Stochastic Kinematics ?

Problem 1: A random walker on a sphere starts at the north pole. What will the probability density function describing his position be based on the properties of his walk ?

Problem 2: The cart-like robot shown in Figure 2.1 moves around in the plane by turning each of its two wheels. This reference frame can be thought of as the time-dependent rigid-body motion

$$g = \begin{pmatrix} \cos\theta & -\sin\theta & x \\ \sin\theta & \cos\theta & y \\ 0 & 0 & 1 \end{pmatrix} \tag{2.1}$$

Let the two wheels each have radii r, and let the distance between the wheels (called the wheelbase) be denoted as L. Imagine that the angles through which the wheels turn around their axes are governed by “stochastic differential equations” of the form

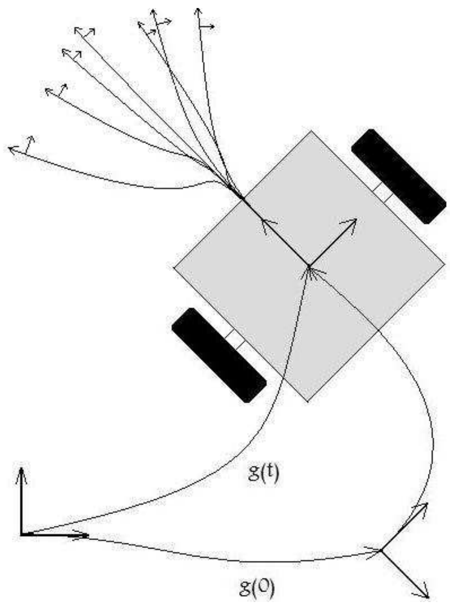

Fig. 2.1. A Kinematic Cart with an Uncertain Future Position and Orientation

$$d\phi_1 = \omega(t)dt + \sqrt{D}dw_1 \tag{2.2}$$
$$d\phi_2 = \omega(t)dt + \sqrt{D}dw_2 \tag{2.3}$$

where dw_i each represent "uncorrelated unit white noise," D scales the strength of the noise, and $r\omega(t)$ is what the forward speed of the robot would be if D were zero. Then a "stochastic trajectory" for $g(t)$ in (2.1) is defined by stochastic differential equations of the form

$$\begin{pmatrix} dx \\ dy \\ d\theta \end{pmatrix} = \begin{pmatrix} r\omega\cos\theta \\ r\omega\sin\theta \\ 0 \end{pmatrix} dt + \sqrt{D} \begin{pmatrix} \frac{r}{2}\cos\theta & \frac{r}{2}\cos\theta \\ \frac{r}{2}\sin\theta & \frac{r}{2}\sin\theta \\ \frac{r}{L} & -\frac{r}{L} \end{pmatrix} \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix} \tag{2.4}$$

The probability density $f(x, y, \theta; t)$ corresponding to this equation is of the form:

$$\begin{aligned} \frac{\partial f}{\partial t} =& -r\omega\cos\theta\frac{\partial f}{\partial x} - r\omega\sin\theta\frac{\partial f}{\partial y} \qquad (2.5) \\ &+ \frac{D}{2}\left(\frac{r^2}{2}\cos^2\theta\frac{\partial^2 f}{\partial x^2} + \frac{r^2}{2}\sin 2\theta\frac{\partial^2 f}{\partial x\partial y} + \frac{r^2}{2}\sin^2\theta\frac{\partial^2 f}{\partial y^2} + \frac{2r^2}{L^2}\frac{\partial^2 f}{\partial\theta^2}\right). \end{aligned}$$

Problem 3: A long and slender semi-flexible biological macromolecule, such as double-helical DNA composed of 300 stacked base pairs, is subjected to random Brownian motion bombardment by the surrounding solvent molecules. If reference frames are attached to both ends of the DNA, what will the distributions of rigid-body motions between these reference frames look like as a function of temperature and the stiffness of the molecule?

Problem 4: One rigid body is set at a fixed pose (or position and orientation) in a box, and a second rigid body is allowed to move uniformly at random in the box under the constraint that it cannot intersect the first body. How much free space is there for the second body to move?

Problem 5: A robot arm has errors in its joints. What is the corresponding error distribution at the end effector ?

Problem 6: A steerable needle is inserted into firm tissue many times, and the trajectories are not exactly repeatable. How can we describe the ensemble of trajectories ?

3

Gaussian Distributions and the Heat Equation

The main things to take away from this section are:

- To become familiar with the Gaussian distribution and its properties, and to be comfortable in performing integrals involving multi-dimensional Gaussians;
- To become acquainted with the concepts of mean, covariance, and information-theoretic entropy;
- To understand how to marginalize and convolve probability densities, to compute conditional densities, and to fold Gaussians;
- To observe that there is a relationship between Gaussian distributions and the heat equation;

3.1 The Gaussian Distribution on the Real Line

The Gaussian distribution with mean at μ and standard deviation σ is denoted

$$\boxed{\rho(x;\mu,\sigma^2) = \rho_{(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.} \tag{3.1}$$

Another common name for the Gaussian distribution is the *normal distribution*. Figure 3.1 shows a plot of the Gaussian distribution with $\mu = 0$

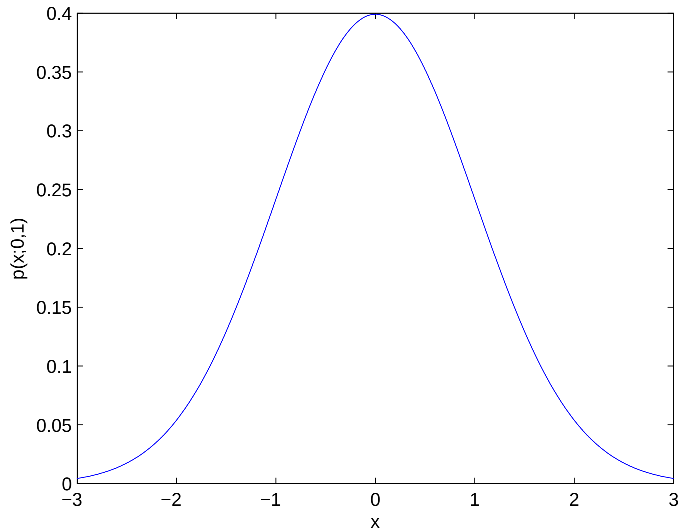

Fig. 3.1. The Gaussian Distribution $\rho(x; 0, 1)$ Plotted over $[-3, 3]$

and $\sigma = 1$ plotted over the range $[-3, 3]$.

The *cumulative distribution function* is

$$F(x; \mu, \sigma^2) = \int_{-\infty}^{x} \rho(\xi; \mu, \sigma^2) d\xi.$$

As $\sigma \to 0$ this is idealized with the *Heaviside step function*

$$H(x) \doteq \begin{cases} 1 \text{ for } x > 0 \\ 0 \text{ for } x \leq 0. \end{cases} \tag{3.2}$$

3.2 The Maximum Entropy Property

The entropy of a pdf $f(x)$ is defined by the integral:

$$\boxed{S(f) = -\int_{-\infty}^{\infty} f(x) \log f(x) dx} \tag{3.3}$$

where here $\log = \log_e = \ln$.

S is computed in closed-form for the Gaussian distribution as

$$S(\rho_{(\mu,\sigma^2)}) = \log(\sqrt{2\pi e}\,\sigma). \tag{3.4}$$

For any given value of variance, the Gaussian distribution is the pdf with maximal entropy:

$$\max_f S(f) \quad \text{subject to} \;\; f(x) \geq 0$$

and

$$\int_{-\infty}^{\infty} f(x)dx = 1\,, \quad \int_{-\infty}^{\infty} xf(x)dx = \mu\,, \quad \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \sigma^2. \tag{3.5}$$

Using Lagrange multipliers:

$$\frac{\partial C}{\partial f} = 0 \;\text{ where }\; C = -f\log f + \lambda_1 f + \lambda_2 x f + \lambda_3 (x-\mu)^2 f.$$

Solving gives $f(x) = \rho_{(\mu,\sigma^2)}(x)$. To show that it actually maximizes the entropy (at least in a local sense), it is possible to define a perturbed version of this pdf as

$$f(x) = \rho_{(\mu,\sigma^2)}(x) \cdot [1 + \epsilon(x)] \tag{3.6}$$

where $\epsilon(x)$ is arbitrary except for the fact that[1] $|\epsilon(x)| << 1$ and it is defined such that $f(x)$ satisfies (3.5). In other words,

$$\int_{-\infty}^{\infty} \rho_{(\mu,\sigma^2)}(x)\epsilon(x)dx = \int_{-\infty}^{\infty} x\rho_{(\mu,\sigma^2)}(x)\epsilon(x)dx$$
$$= \int_{-\infty}^{\infty} (x-\mu)^2\rho_{(\mu,\sigma^2)}(x)\epsilon(x)dx = 0.$$

Substituting (3.6) into (3.3) and using the Taylor series approximation $\log(1+\epsilon) \approx \epsilon - \epsilon^2/2$,

$$\begin{aligned} S(f) &= -\int_{-\infty}^{\infty} \rho_{(\mu,\sigma^2)}(x) \cdot [1+\epsilon(x)] \log(\rho_{(\mu,\sigma^2)}(x) \cdot [1+\epsilon(x)])dx \\ &= -\int_{-\infty}^{\infty} \rho_{(\mu,\sigma^2)}(x) \cdot [1+\epsilon(x)] \cdot [\log(\rho_{(\mu,\sigma^2)}(x)) + \log(1+\epsilon(x))]dx \\ &= S(\rho_{(\mu,\sigma^2)}) - F(\epsilon^2) + O(\epsilon^3) \end{aligned}$$

where the functional F is always positive.

[1]To be concrete, $\epsilon = 0.01 << 1$. Then $\epsilon^3 = 10^{-6}$ is certainly negligible in comparison to quantities that are on the order of 1.

3.3 The Convolution of Gaussians

The *convolution* of two pdfs on the real line is defined as

$$(f_1 * f_2)(x) \doteq \int_{-\infty}^{\infty} f_1(\xi) f_2(x-\xi) d\xi. \tag{3.7}$$

It can be shown that the convolution integral will always exist for "nice" functions, and furthermore

$$f_i \in \mathcal{N}(\mathbb{R}) \implies f_1 * f_2 \in \mathcal{N}(\mathbb{R}).$$

The Gaussian distribution has the property that the convolution of two Gaussians is a Gaussian:

$$\rho(x;\mu_1,\sigma_1^2) * \rho(x;\mu_2,\sigma_2^2) = \rho(x;\mu_1+\mu_2,\sigma_1^2+\sigma_2^2) \tag{3.8}$$

The Dirac δ-function can be viewed as the limit

$$\delta(x) = \lim_{\sigma \to 0} \rho(x;0,\sigma^2). \tag{3.9}$$

It then follows from (3.8) that

$$\rho(x;\mu_1,\sigma_1^2) * \delta(x) = \rho(x;\mu_1,\sigma_1^2).$$

3.4 The Fourier Transform of the Gaussian Distribution

The Fourier transform of a "nice" function $f \in \mathcal{N}(\mathbb{R})$ is defined as

$$[\mathcal{F}(f)](\omega) \doteq \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx. \tag{3.10}$$

The shorthand $\hat{f}(\omega) \doteq [\mathcal{F}(f)](\omega)$ will be used frequently.

From the definition of the Fourier transform, it can be shown that

$$(\widehat{f_1 * f_2})(\omega) = \hat{f}_1(\omega)\hat{f}_2(\omega) \tag{3.11}$$

(i.e., the Fourier transform of the convolution is the product of Fourier transforms) and

$$f(x) = [\mathcal{F}^{-1}(\hat{f})](x) \doteq \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega x} d\omega. \tag{3.12}$$

This is called the *inverse Fourier transform* or *Fourier reconstruction formula*

3.5 Diffusion Equations

A one-dimensional linear diffusion equation with constant coefficients has the form

$$\frac{\partial u}{\partial t} = a\frac{\partial u}{\partial x} + b\frac{\partial^2 u}{\partial x^2} \tag{3.13}$$

where $a \in \mathbb{R}$ is called the drift coefficient and $b \in \mathbb{R}_{>0}$ is called the diffusion coefficient.

Taking the Fourier transform of $u(x,t)$ for each value of t gives

$$\frac{d\hat{u}}{dt} = (ia\omega - b\omega^2)\hat{u} \quad \text{with} \quad \hat{u}(\omega, 0) = \hat{f}(\omega).$$

The solution to this initial value problem is of the form

$$\hat{u}(\omega, t) = \hat{f}(\omega)e^{(ia\omega - b\omega^2)t}.$$

Application of the inverse Fourier transform yields a solution. The above expression for $\hat{u}(\omega, t)$ is a Gaussian with phase factor, and on inversion this becomes a shifted Gaussian:

$$[\mathcal{F}^{-1}(e^{iat\omega}e^{-b\omega^2 t})](x) = \frac{1}{\sqrt{4\pi bt}} \exp\left(-\frac{(x+at)^2}{4bt}\right).$$

Using the convolution theorem in reverse then gives

$$u(x,t) = \frac{1}{\sqrt{4\pi bt}} \int_{-\infty}^{\infty} f(\xi) \exp\left(-\frac{(x+at-\xi)^2}{4bt}\right) d\xi. \tag{3.14}$$

3.6 The Multivariate Gaussian Distribution

The multivariate Gaussian distribution on $\mathbb{R}^n$ is defined as

$$\boxed{\rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma) \doteq \frac{1}{(2\pi)^{n/2}|\det \Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}.} \tag{3.15}$$

This is the maximum entropy distribution subject to the constraints

$$\int_{\mathbb{R}^n} \rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma)\, d\mathbf{x} = 1;$$

$$\int_{\mathbb{R}^n} \mathbf{x}\,\rho(\mathbf{x};\boldsymbol{\mu},\Sigma)\,d\mathbf{x} = \boldsymbol{\mu};$$

$$\int_{\mathbb{R}^n} (\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\rho(\mathbf{x};\boldsymbol{\mu},\Sigma)\,d\mathbf{x} = \Sigma.$$

3.6.1 Conditional and Marginal Densities

A vector $\mathbf{x} \in \mathbb{R}^n$ can be partitioned as

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T \in \mathbb{R}^{n_1+n_2}$$

where $\mathbf{x}_1 \in \mathbb{R}^{n_1}$ and $\mathbf{x}_2 \in \mathbb{R}^{n_2}$.

If $f(\mathbf{x}) = f([\mathbf{x}_1^T, \mathbf{x}_2^T]^T)$ is any pdf on $\mathbb{R}^{n_1+n_2}$, then

$$f_1(\mathbf{x}_1) = \int_{\mathbb{R}^{n_2}} f(\mathbf{x}_1, \mathbf{x}_2)\,d\mathbf{x}_2.$$

$f_2(\mathbf{x}_2)$ is obtained from $f(\mathbf{x}_1, \mathbf{x}_2)$ in a similar way by integrating over all values of $\mathbf{x}_1$.

The mean and variance of $f_1(\mathbf{x}_1)$ are obtained from the mean and variance of $f(\mathbf{x})$ by observing that

$$\begin{aligned}\boldsymbol{\mu}_1 &= \int_{\mathbb{R}^{n_1}} \mathbf{x}_1 f_1(\mathbf{x}_1)\, d\mathbf{x}_1 \\ &= \int_{\mathbb{R}^{n_1}} \mathbf{x}_1 \left(\int_{\mathbb{R}^{n_2}} f(\mathbf{x}_1, \mathbf{x}_2)\, d\mathbf{x}_2 \right) d\mathbf{x}_1 \\ &= \int_{\mathbb{R}^{n_1}} \int_{\mathbb{R}^{n_2}} \mathbf{x}_1 f(\mathbf{x}_1, \mathbf{x}_2)\, d\mathbf{x}_2\, d\mathbf{x}_1\end{aligned}$$

and

$$\begin{aligned}\Sigma_{11} &= \int_{\mathbb{R}^{n_1}} (\mathbf{x}_1 - \boldsymbol{\mu}_1)(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T f_1(\mathbf{x}_1)\, d\mathbf{x}_1 \\ &= \int_{\mathbb{R}^{n_1}} (\mathbf{x}_1 - \boldsymbol{\mu}_1)(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \left(\int_{\mathbb{R}^{n_2}} f(\mathbf{x}_1, \mathbf{x}_2)\, d\mathbf{x}_2 \right) d\mathbf{x}_1 \\ &= \int_{\mathbb{R}^{n_1}} \int_{\mathbb{R}^{n_2}} (\mathbf{x}_1 - \boldsymbol{\mu}_1)(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T f(\mathbf{x}_1, \mathbf{x}_2)\, d\mathbf{x}_2\, d\mathbf{x}_1\end{aligned}$$

In other words, the mean vector and covariance matrix for the marginal density are obtained directly from those of the full density. For example, $\boldsymbol{\mu} = [\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T]^T$.

Given a (multivariate) Gaussian distribution $\rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$, the associated covariance matrix can be written in terms of blocks as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where $\Sigma_{11} = \Sigma_{11}^T$, $\Sigma_{22} = \Sigma_{22}^T$, and $\Sigma_{21} = \Sigma_{12}^T$.

The *marginal density* that results from integrating the Gaussian distribution $\rho(\mathbf{x}, \boldsymbol{\mu}, \Sigma)$ over all values of $\mathbf{x}_2$ is:

$$\int_{\mathbb{R}^{n_2}} \rho([\mathbf{x}_1^T, \mathbf{x}_2^T]^T; \boldsymbol{\mu}, \Sigma) d\mathbf{x}_2 = \rho(\mathbf{x}_1; \boldsymbol{\mu}_1, \Sigma_{11}). \tag{3.16}$$

Given $f(\mathbf{x}_1, \mathbf{x}_2)$, the conditional density of $\mathbf{x}_1$ given $\mathbf{x}_2$ is

$$f(\mathbf{x}_1|\mathbf{x}_2) \doteq f(\mathbf{x}_1, \mathbf{x}_2)/f_2(\mathbf{x}_2). \tag{3.17}$$

Evaluating this expression using a Gaussian gives

$$\rho([\mathbf{x}_1^T, \mathbf{x}_2^T]^T; \boldsymbol{\mu}, \Sigma)/\rho(\mathbf{x}_2; \boldsymbol{\mu}_2, \Sigma_2) =$$

$$\rho(\mathbf{x}_1; \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}). \tag{3.18}$$

3.6.2 Multidimensional Integrals Involving Gaussians

First, it is well known that

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi} \quad \Longrightarrow \quad \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{x}\right) d\mathbf{x} = (2\pi)^{\frac{n}{2}}. \tag{3.19}$$

Here $\mathbf{x} \in \mathbb{R}^n$ and $d\mathbf{x} = dx_1 dx_2 \cdots dx_n$. Note also that

$$\int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}. \tag{3.20}$$

These identities are used below to prove:

$$\int_{\mathbb{R}^n} \exp(-\frac{1}{2}\mathbf{x}^T M\mathbf{x} - \mathbf{m}^T\mathbf{x})d\mathbf{x} = (2\pi)^{n/2}|\mathrm{det}M|^{-\frac{1}{2}} \exp\left(\frac{1}{2}\mathbf{m}^T M^{-1}\mathbf{m}\right) \tag{3.21}$$

and

$$\int_{\mathbb{R}^n} \mathbf{x}^T G\mathbf{x} \exp\left(-\frac{1}{2}\mathbf{x}^T A\mathbf{x}\right) d\mathbf{x} = (2\pi)^{n/2} \frac{\mathrm{tr}(GA^{-1})}{|\det A|^{\frac{1}{2}}}. \tag{3.22}$$

3.7 Folded, or Wrapped, Gaussians

In some applications, data on the circle is given, and a corresponding concept of Gaussian distribution is needed. The tails of a Gaussian can be "wrapped around" the circle as

$$\boxed{\rho_W(\theta; \mu, \sigma) \doteq \sum_{k=-\infty}^{\infty} \rho(\theta - 2\pi k; \mu, \sigma),.} \tag{3.23}$$

Recall that any 2π-periodic function, i.e., a "function on the unit circle" can be expanded in a Fourier series:

$$f(\theta) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{in\theta} \;\text{ where }\; \hat{f}(n) = \int_0^{2\pi} f(\theta) e^{-in\theta} d\theta. \tag{3.24}$$

where $e^{in\theta} = \cos n\theta + i \sin n\theta$ and $i = \sqrt{-1}$. This leads to the Fourier series representation of the folded Gaussian distribution:

$$\boxed{\rho_W(\theta;\mu,\sigma) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{\infty} e^{-\frac{\sigma^2}{2}n^2}\cos\left(n(\theta-\mu)\right).} \tag{3.25}$$

As σ becomes large, very close approximations can be achieved with the first couple of terms in the summation in (3.25). In contrast, as σ becomes very small, using very few of the terms in the series (3.23) will produce a very good approximation when $\mu = 0$.

3.8 The Heat Equation

3.8.1 The One-Dimensional Case

Consider the diffusion equation

$$\frac{\partial f}{\partial t} = \frac{1}{2}k(t)\frac{\partial^2 f}{\partial x^2} - a(t)\frac{\partial f}{\partial x}. \tag{3.26}$$

The initial condition is $f(x,0) = \delta(x)$. The solution $f(x,t)$ can be obtained in closed form, following essentially the same procedure as in Section 3.5, and then the mean and variance can be computed from this solution as

$$\mu(t) = \int_{-\infty}^{\infty} x f(x,t) dx \quad \text{and} \quad \sigma^2(t) = \int_{-\infty}^{\infty} [x - \mu(t)]^2 f(x,t) dx. \tag{3.27}$$

Alternatively, the mean and variance of $f(x,t)$ can be computed directly from (3.26) without actually knowing the solution $f(x,t)$.

$$\int_{-\infty}^{\infty} \frac{\partial f}{\partial x} dx = f(x,t)\Big|_{x=-\infty}^{\infty} \quad \text{and} \quad \int_{-\infty}^{\infty} \frac{\partial^2 f}{\partial x^2} dx = \left.\frac{\partial f}{\partial x}\right|_{x=-\infty}^{\infty}$$

and under the boundary conditions that $f(x,t)$ and $\partial f/\partial x$ decay rapidly to zero as $x \to \pm\infty$, these terms become zero. Since the initial conditions are a delta function in x, it follows that

$$\int_{-\infty}^{\infty} f(x,t) dx = 1.$$

In other words, (3.26) preserves the initial mass of the distribution over all values of time after $t = 0$.

To compute $\mu(t)$, multiply both sides of (3.26) by x and integrate. On the one hand,

$$\int_{-\infty}^{\infty} x\frac{\partial f}{\partial t}dx = \frac{d}{dt}\int_{-\infty}^{\infty} xf(x,t)dx = \frac{d\mu}{dt}.$$

On the other hand,

$$\int_{-\infty}^{\infty} x\frac{\partial f}{\partial t}dx = \frac{1}{2}k(t)\int_{-\infty}^{\infty} x\frac{\partial^2 f}{\partial x^2}dx - a(t)\int_{-\infty}^{\infty} x\frac{\partial f}{\partial x}dx.$$

Evaluating both integrals on the right side by integrating by parts and using the conditions that both $f(x,t)$ and $\partial f/\partial x$ decay rapidly to zero as $x \to \pm\infty$, it becomes clear that

$$\frac{d\mu}{dt} = a(t) \quad \text{or} \quad \mu(t) = \int_0^t a(s)ds. \tag{3.28}$$

A similar argument shows that

$$\frac{d}{dt}(\sigma^2) = k(t) \quad \text{or} \quad \sigma^2(t) = \int_0^t k(s)ds. \tag{3.29}$$

3.8.2 The Multi-Dimensional Case

Consider the following time-varying diffusion equation without drift:

$$\frac{\partial f}{\partial t} = \frac{1}{2}\sum_{i,j=1}^{n} D_{ij}(t)\frac{\partial^2 f}{\partial x_i \partial x_j}, \tag{3.30}$$

Multiplying both sides by $x_k x_l$ and integrating over $\mathbf{x} \in \mathbb{R}^n$ gives

$$\frac{d}{dt}(\sigma_{kl}) = \frac{1}{2}\sum_{i,j=1}^{n} D_{ij}(t)\int_{\mathbb{R}^n} x_k x_l \frac{\partial^2 f}{\partial x_i \partial x_j} d\mathbf{x}. \tag{3.31}$$

From integration by parts

$$\int_{\mathbb{R}^n} x_k x_l \frac{\partial^2 f}{\partial x_i \partial x_j} d\mathbf{x} = \int_{\mathbf{x}-x_i}\left[x_k x_l \left.\frac{\partial f}{\partial x_j}\right|_{x_i=-\infty}^{\infty} - \int_{-\infty}^{\infty}\frac{\partial}{\partial x_i}(x_k x_l)\frac{\partial f}{\partial x_j}dx_i \right] d\mathbf{x}/dx_i$$

The assumption that $f(\mathbf{x}, t)$ decays rapidly as $\|\mathbf{x}\| \to \infty$ for all values of t makes the first term in the brackets disappear. Using the fact that $\partial x_i/\partial x_j = \delta_{ij}$, and integrating by parts again (over x_j) reduces the above integral to

$$\int_{\mathbb{R}^n} x_k x_l \frac{\partial^2 f}{\partial x_i \partial x_j} d\mathbf{x} = \delta_{kj}\delta_{il} + \delta_{ik}\delta_{lj}.$$

Substituting this into (3.31) results in

$$\frac{d}{dt}(\sigma_{kl}) = D_{kl}(t) \quad \text{or} \quad \sigma_{kl}(t) = \int_0^t D_{kl}(s)ds. \tag{3.32}$$

3.8.3 The Heat Equation on the Unit Circle

Given

$$\frac{\partial f}{\partial t} = \frac{1}{2}k\frac{\partial^2 f}{\partial \theta^2} \quad \text{subject to} \quad f(\theta, 0) = \delta(\theta),$$

the Fourier solution is

$$f(\theta, t) = \sum_{k=-\infty}^{\infty} \rho(\theta - 2\pi k; 0, (kt)^{\frac{1}{2}}) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{\infty} e^{-ktn^2/2}\cos n\theta \tag{3.33}$$

This is the folded Gaussian in (3.25) with $\sigma^2 = kt$ and $\mu = 0$.

3.9 Gaussians and Multidimensional Diffusions

3.9.1 The Constant Diffusion Case

Consider the diffusion equation

$$\frac{\partial f}{\partial t} = \frac{1}{2} \sum_{i,j=1}^{n} D_{ij} \frac{\partial f^2}{\partial x_i \partial x_j} \tag{3.34}$$

subject to the initial conditions $f(\mathbf{x}, t) = \delta(\mathbf{x})$, where $D = [D_{ij}] = D^T$ is a constant matrix of diffusion constants.

Since diffusion equations preserve mass (see Section 3.8.2), it follows that

$$\int_{\mathbb{R}^n} f(\mathbf{x}, t) d\mathbf{x} = 1 \tag{3.35}$$

for all values of time, $t \in \mathbb{R}_{>0}$.

Try a solution of the form

$$f(\mathbf{x}, t) = c(t) \exp(-\frac{1}{2}\mathbf{x}^T A(t)\mathbf{x}) \tag{3.36}$$

where $A(t) = \phi(t)A_0$ and $A_0 = [\alpha_{ij}] = A_0^T$. Then, from (3.35) it follows that

$$c(t) = \left(\frac{\phi(t)}{2\pi}\right)^{n/2} |\det A_0|^{\frac{1}{2}}.$$

With this constraint in mind, substituting $f(\mathbf{x}, t)$ into (3.34) produces the following conditions on $\phi(t)$ and A_0:

$$n\phi' = -\phi^2 \sum_{i,j=1}^{n} D_{ij}\alpha_{ij}$$

$$\phi'\mathbf{x}^T A_0 \mathbf{x} = -\phi^2 \sum_{i,j=1}^{n} D_{ij} \left(\sum_{k=1}^{n} \alpha_{ik}x_k\right) \left(\sum_{l=1}^{n} \alpha_{jl}x_l\right)$$

where $\phi' = d\phi/dt$.

Both of the conditions (3.37) are satisfied if $A_0 = \alpha_0 D^{-1}$ and $\phi(t) = (\alpha_0 t)^{-1}$ for some arbitrary constant $\alpha_0 \in \mathbb{R}_{>0}$. But since $A(t) = \phi(t)A_0 = t^{-1}D^{-1}$, this constant does not matter.

Putting all of this together,

$$\boxed{f(\mathbf{x},t) = \frac{1}{(2\pi t)^{n/2}|\det D|^{\frac{1}{2}}} \exp(-\frac{1}{2t}\mathbf{x}^T D^{-1}\mathbf{x}).} \tag{3.37}$$

Stated in another way, the solution to (3.34) is a time-varying Gaussian distribution with $\Sigma(t) = tD$ when D is symmetric.

4

Probability and Information Theory

For those not familiar with probability and information theory the main things to take away from this section are:

- To know that the definitions of convolution, mean, covariance, and marginal and conditional densities, are fully general, and apply to a wide variety of probability density functions (not only Gaussians);
- To understand the definitions and properties of (continuous/differential) information-theoretic entropy, including how it scales and how it behaves under convolution;
- To understand the fundamental inequalities of information theory such as the entropy power inequality;

4.1 Marginalization, Conditioning and Convolution

Marginalization:

$$\rho(x_1, x_2, ..., x_m) = \int_{x_{m+1}=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} \rho(x_1, x_2, ..., x_n) dx_{m+1} \cdots dx_n.$$

Conditioning:

$$\rho(x_1, x_2, ..., x_m | x_{m+1}, x_{m+2}, ..., x_n) = \rho(x_1, x_2, ..., x_n)/\rho(x_{m+1}, x_{m+2}, ..., x_n)$$

Convolution:

$$\rho_{X+Y}(\mathbf{x}) = (\rho_X * \rho_Y)(\mathbf{x}) = \int_{\mathbb{R}^n} \rho_X(\boldsymbol{\xi})\rho_Y(\mathbf{x} - \boldsymbol{\xi})d\boldsymbol{\xi}.$$

4.1.1 Mean and Covariance

$$\boxed{\boldsymbol{\mu} = \langle \mathbf{x} \rangle = \int_{\mathbb{R}^n} \mathbf{x}\rho(\mathbf{x})d\mathbf{x}, \quad \text{or} \quad \langle \mathbf{x} - \boldsymbol{\mu} \rangle = \int_{\mathbb{R}^n} (\mathbf{x} - \boldsymbol{\mu})\rho(\mathbf{x})d\mathbf{x} = \mathbf{0}.} \tag{4.1}$$

Note that $\boldsymbol{\mu}$ minimizes the cost function

$$c(\mathbf{x}) = \int_{\mathbb{R}^n} \|\mathbf{x} - \mathbf{y}\|^2 f(\mathbf{y})\, d\mathbf{y} \tag{4.2}$$

where $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$ is the 2-norm in $\mathbb{R}^n$.

The covariance about the mean is the $n \times n$ matrix defined as

$$\boxed{\Sigma = \langle(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\rangle = \int_{\mathbb{R}^n}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\rho(\mathbf{x})d\mathbf{x}.} \tag{4.3}$$

It follows from this definition that

$$\int_{\mathbb{R}^n}\mathbf{x}\mathbf{x}^T\rho(\mathbf{x})\,d\mathbf{x} = \Sigma + \boldsymbol{\mu}\,\boldsymbol{\mu}^T. \tag{4.4}$$

If $\mathbf{z} = A\mathbf{x} + \mathbf{a}$,

$$\begin{aligned}\boldsymbol{\mu}_Z = \langle\mathbf{z}\rangle &= \int_{\mathbb{R}^n}(A\mathbf{x}+\mathbf{a})\rho(\mathbf{x})d\mathbf{x} \\ &= A\left(\int_{\mathbb{R}^n}\mathbf{x}\rho(\mathbf{x})d\mathbf{x}\right) + \mathbf{a}\left(\int_{\mathbb{R}^n}\rho(\mathbf{x})d\mathbf{x}\right) \\ &= A\boldsymbol{\mu}_X + \mathbf{a}\end{aligned}$$

and

$$
\begin{aligned}
\Sigma_Z &= \langle (\mathbf{z} - \boldsymbol{\mu}_Z)(\mathbf{z} - \boldsymbol{\mu}_Z)^T \rangle \\
&= \int_{\mathbb{R}^n} (A\mathbf{x} + \mathbf{a} - \boldsymbol{\mu}_Z)(A\mathbf{x} + \mathbf{a} - \boldsymbol{\mu}_Z)^T \rho(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}^n} (A[\mathbf{x} - \boldsymbol{\mu}_X])(A[\mathbf{x} - \boldsymbol{\mu}_X])^T \rho(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}^n} A[\mathbf{x} - \boldsymbol{\mu}_X][\mathbf{x} - \boldsymbol{\mu}_X]^T A^T \rho(\mathbf{x}) d\mathbf{x} \\
&= A \left(\int_{\mathbb{R}^n} [\mathbf{x} - \boldsymbol{\mu}_X][\mathbf{x} - \boldsymbol{\mu}_X]^T \rho(\mathbf{x}) d\mathbf{x} \right) A^T \\
&= A \Sigma_X A^T.
\end{aligned}
$$

Pdfs are often used to describe distributions of errors. If these errors are concatenated, they ‘add’ by convolution:

$$(\rho_1 * \rho_2)(\mathbf{x}) = \int_{\mathbb{R}^n} \rho_1(\boldsymbol{\xi})\rho_2(\mathbf{x} - \boldsymbol{\xi})\, d\boldsymbol{\xi}. \tag{4.5}$$

The mean and covariance of convolved distributions are found as

$$\boldsymbol{\mu}_{1*2} = \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 \text{ and } \Sigma_{1*2} = \Sigma_1 + \Sigma_2 \tag{4.6}$$

If the scalar random variables $X_1, X_2, ..., X_n$ are all *independent* of each other, then the corresponding probability density function is separable:

$$\rho(x_1, x_2, ..., x_n) = \rho_1(x_1)\rho_2(x_2)\cdots\rho_n(x_n). \tag{4.7}$$

When this happens, the covariance matrix will be diagonal.

4.1.2 Jensen's Inequality

If $\Phi(x)$ is a *convex function* on $\mathbb{R}$, i.e.,

$$\Phi(tx + (1-t)y) \le t\Phi(x) + (1-t)\Phi(y) \quad \forall\ t \in [0,1] \tag{4.8}$$

then Jensen's inequality [8] states

$$\Phi\left(\int_{-\infty}^{\infty}\phi(x)f(x)dx\right)\leq\int_{-\infty}^{\infty}\Phi(\phi(x))f(x)dx$$

If $\phi(x) = f_2(x)/f_1(x)$, $f(x) = f_1(x)$, and $\Phi(y) = -\log y$, the following property of the *Kullback-Leibler divergence* is observed:

$$\begin{aligned}D_{KL}(f_1\|f_2) &\doteq \int_{-\infty}^{\infty} f_1(x)\log\frac{f_1(x)}{f_2(x)}dx \\ &= -\int_{-\infty}^{\infty} f_1(x)\log\frac{f_2(x)}{f_1(x)}dx \\ &\geq -\log\int_{-\infty}^{\infty} f_1(x)\frac{f_2(x)}{f_1(x)}dx \\ &= -\log 1 = 0,\end{aligned}$$

and likewise for domains other than the real line.

4.2 Some Information Theory

Given a probability density function (pdf) $f(\mathbf{x})$ describing the distribution of states of a random vector $\mathbf{X} \in \mathbb{R}^n$, the information-theoretic entropy is defined as[1]

$$\boxed{S(f) \doteq - \int_{\mathbf{x}} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x}.} \tag{4.9}$$

This is a measure of dispersion of a pdf.

Note that the standard in the literature is to denote the entropy of the random variable $\mathbf{X}$ as $H(\mathbf{X})$. However, the notation $S(f)$ (which stands for the entropy of the pdf that fully describes the random variable $\mathbf{X}$) generalizes more easily to the Lie group setting addressed in Volume 2.

4.2.1 Entropy and Gaussian Distributions

The information-theoretic entropy of a one-dimensional and n-dimensional Gaussian distributions

[1] In information theory, this would be called *differential entropy*. It is referred to here as *continuous entropy* to denote the difference between this and the discrete case.

$$\rho_{(0,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \quad \text{and} \quad \rho_{(\mathbf{0},\Sigma)}(\mathbf{x}) = \frac{1}{(2\pi)^{\mathrm{n}/2}|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x})$$

are respectively [11]:

$$S(\rho_{(0,\sigma^2)}) = \log(\sqrt{2\pi e}\sigma)$$

and

$$S(\rho_{(\mathbf{0},\Sigma)}) = \log\{(2\pi e)^{n/2}|\Sigma|^{\frac{1}{2}}\} \tag{4.10}$$

where $\log = \log_e$.

4.2.2 Information-Theoretic Measures of Divergence

Given two probability density functions f_1 and f_2 on $\mathbb{R}^n$, the *Kullback-Leibler divergence* between them is defined as

$$\boxed{D_{KL}(f_1\|f_2) \doteq \int_{\mathbb{R}^n} f_1(\mathbf{x}) \log\left(\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}\right) d\mathbf{x}.} \tag{4.11}$$

Note that [10]:

$$\frac{1}{4}\left(\int_{\mathbb{R}^n} |f_1(\mathbf{x}) - f_2(\mathbf{x})| d\mathbf{x}\right)^2 \leq D_{KL}(f_1 \| f_2).$$

The Fisher information divergence between two pdfs is defined as

$$\boxed{D_{FI}(f_1 \| f_2) \doteq \int_{\mathbb{R}^n} \left\| \frac{1}{f_1} \nabla f_1 - \frac{1}{f_2} \nabla f_2 \right\|^2 f_1 d\mathbf{x}.} \tag{4.12}$$

This is also not a "distance" function in the sense that it is not symmetric in the arguments and does not satisfy the triangle inequality. In the one-dimensional case, this can be written as

$$D_{FI}(f_1 \| f_2) = \int_{-\infty}^{\infty} \left(\frac{1}{f_1}\frac{df_1}{dx} - \frac{1}{f_2}\frac{df_2}{dx}\right)^2 f_1 dx = 4\int_{-\infty}^{\infty} \left(\frac{d}{dx}\sqrt{\frac{f_1}{f_2}}\right)^2 f_2 dx.$$

The Multi-Dimensional Case

The multidimensional case proceeds in a similar way as in the one-dimensional case. Given a pdf $f(\mathbf{x})$, a shifted version is $f_{\mathbf{a}}(\mathbf{x}) = f(\mathbf{x} - \mathbf{a})$. And

$$S(f_{\mathbf{a}}) = -\int_{\mathbb{R}^n} f(\mathbf{x}-\mathbf{a}) \log f(\mathbf{x}-\mathbf{a}) d\mathbf{x} = -\int_{\mathbb{R}^n} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x} = S(f).$$

Now consider the scaled version of the pdf $f(\mathbf{x})$ defined as

$$f_A(\mathbf{x}) = \frac{1}{\det A} f(A^{-1}\mathbf{x}) \text{ where } \det A > 0.$$

If $\det A > 1$ this is a more "spread out" version of f, and if $\det A < 1$, then this is a more "concentrated" version of f. It can be verified easily that $f_A(\mathbf{x})$ is indeed a pdf by making the change of coordinates $y = A^{-1}\mathbf{x}$ and replacing the integral over $\mathbf{x}$ with that over $\mathbf{y}$.

The entropy of $f_A(\mathbf{x})$ is calculated as

$$\begin{aligned} S(f_A) &= -\int_{\mathbb{R}^n} \frac{1}{\det A} f(A^{-1}\mathbf{x}) \log \left[\frac{1}{\det A} f(A^{-1}\mathbf{x}) \right] d\mathbf{x} \\ &= -\int_{\mathbb{R}^n} f(\mathbf{y}) \log \left[\frac{1}{\det A} f(\mathbf{y}) \right] d\mathbf{y} \\ &= S(f) + \log \det A. \end{aligned}$$

4.2.3 The Entropy Power Inequality

The statement of the entropy power inequality dates back to Shannon's original paper, though complete and rigorous proofs came later [12, 2].

Shannon defined the entropy power of a pdf $p(x)$ on $\mathbb{R}^n$ as

$$N(p) = \exp(2S(p)/n)/2\pi e$$

where $S(p)$ is the entropy of p. The entropy power inequality then states

$$\boxed{N(p * q) \geq N(p) + N(q)} \tag{4.13}$$

with equality if and only if p and q are both Gaussian distributions with covariance matrices that are a scalar multiple of each other.

References

1. Applebaum, D., *Probability and Information: An Integrated Approach*, 2^{nd} ed., Cambridge University Press, Cambridge, England, 2008.
2. Blachman, N.M., "The convolution inequality for entropy powers," *IEEE Trans. Inform. Theory*, 11(2): 267271, 1965.
3. Braunstein, S.L., Caves, C.M., "Statistical distance and the geometry of quantum states," *Phys. Rev. Lett.* 72, 3439 - 3443, 1994.
4. Cover, T.M., Thomas, J.A., *Elements of Information Theory*, Wiley-Interscience, 2^{nd} ed., Hoboken, NJ, 2006.
5. Fisher, R.A., "On the mathematical foundations of theoretical statistics," *Philos. Trans. Roy. Soc. London, A*, 222: 309-368, 1922
6. Fisher, R.A., "Theory of statistical estimation" Proc. Cambridge Phil. Soc. 22, 700-725, 1925.
7. Frieden, B.R., "Science from Fisher Information," Cambridge University Press, New York, 2004.

8. Jensen, J.L.W.V., “Sur les fonctions convexes et les inégalités entre les valeurs moyennes,” *Acta Mathematica* 30: 175-193, 1906.
9. Johnson, O., *Information Theory and the Central Limit Theorem*, Imperial College Press, London, England, 2004.
10. Kullback, S., *Information Theory and Statistics*, Dover, 1997 (originally published in 1958).
11. Shannon, C. E., Weaver, W., *The Mathematical Theory of Communication*, Univ. Illinois Press, Urbana, Illinois, 1949.
12. Stam, A.J., “Some inequalities satisfied by the quantities of information of Fisher and Shannon,” *Information and Control*, 2(2): 101-112, June 1959.

5

Stochastic Differential Equations

The main points to take away from this section are:

- Whereas a deterministic system of ordinary differential equations that satisfies certain conditions (i.e., the Lipschitz conditions) are guaranteed to have a unique solution for any given initial conditions, when random noise is introduced the resulting "stochastic differential equation" will not produce repeatable solutions;
- It is the ensemble behavior of the sample paths obtained from numerically solving a stochastic differential equation many times that is important;
- This ensemble behavior can be described either as a stochastic integral (of which there are two main types, called Itô and Stratonovich), or by using a partial differential equation called the Fokker-Planck (or forward Kolmogorov) equation.
- Two different forms of the Fokker-Planck equation exist, corresponding to the interpretation of the solution of a given SDE as being either an Itô or Stratonovich integral, and an analytical apparatus exists for converting between these forms.

- Multi-dimensional SDEs in $\mathbb{R}^n$ can be written in Cartesian or curvilinear coordinates, but care must be taken when converting between coordinate systems because the usual rules of multi-variable Calculus do not apply in some situations.

5.1 Continuous-Time Brownian Motion in Continuous Space

A one-dimensional SDE will more generally be thought of as the limiting case of an equation of the form

$$x(t+\Delta t) - x(t) = a(x,t)\Delta t + b(x,t)n(t)\Delta t \quad \text{where} \quad x(0) = x_0 \tag{5.1}$$

as $\Delta t \to 0$. How do we define noise, $n(t)$? Basically sample from a Gaussian distribution.

5.1.1 Formal Properties of Weiner Processes

The vector $\mathbf{w}(t) = [w_1, ..., w_m]^T$ denotes an m-dimensional *Wiener process* (also called a Brownian motion process) with the following properties:

$$
\begin{aligned}
\langle w_j(t)\rangle &= 0 \;\forall\;\; t \geq 0;\\
w_j(0) &= 0;\\
\langle [w_j(t_1+t) - w_j(t_2+t)]^2\rangle &= \langle [w_j(t_1) - w_j(t_2)]^2\rangle\\
&\quad \forall \;\;\; t_1, t_2, t_1+t, t_2+t \geq 0;\\
\langle [w(t_i) - w(t_j)][w(t_k) - w(t_l)]\rangle &= 0 \;\;\forall\;\; t_i > t_j \geq t_k > t_l \geq 0.
\end{aligned}
$$

From these defining properties, it is clear that for the Wiener process, $w_j(t)$,

$$
\langle [w_j(t_1+t_2)]^2\rangle = \langle [w_j(t_1+t_2) - w_j(t_1) + w_j(t_1) - w_j(0)]^2\rangle
$$

$$
= \langle [w_j(t_1+t_2) - w_j(t_1)]^2 + [w_j(t_1) - w_j(0)]^2\rangle = \langle [w_j(t_1)]^2\rangle + \langle [w_j(t_2)]^2\rangle.
$$

For the equality

$$
\langle [w_j(t_1+t_2)]^2\rangle = \langle [w_j(t_1)]^2\rangle + \langle [w_j(t_2)]^2\rangle \tag{5.2}
$$

to hold for all values of time t_1, t_2, it must be the case that [9]

$$
\langle [w_j(t-s)]^2\rangle = \sigma_j^2 |t-s|. \tag{5.3}
$$

for some positive real number σ_j^2. The notation dw_j is defined by

$$dw_j(t) \doteq w_j(t+dt) - w_j(t). \tag{5.4}$$

From the definitions and discussion above,

$$\langle dw_j(t) \rangle = \langle w_j(t+dt) \rangle - \langle w_j(t) \rangle = 0$$

and

$$\begin{aligned} \langle [dw_j(t)]^2 \rangle &= \langle (w_j(t+dt) - w_j(t))(w_j(t+dt) - w_j(t)) \rangle \\ &= \langle [w_j(t+dt)]^2 \rangle - 2\langle w_j(t) w_j(t+dt) \rangle + \langle [w_j(t)]^2 \rangle \\ &= \sigma_j^2 (t + dt - 2t + t) = \sigma_j^2 dt. \end{aligned}$$

And

$$\boxed{\langle w_i(s) w_j(t) \rangle = \sigma_j^2 \delta_{ij} \min(s,t) \quad \text{and} \quad \langle dw_i(t_i) dw_j(t_j) \rangle = \sigma_j^2 \delta_{ij} dt_j.} \tag{5.5}$$

The *unit strength* Wiener process has $\sigma_j^2 = 1$.

5.2 The Itô Stochastic Calculus

In the usual Calculus, the *Riemann integral* of a continuous function $f : [a, b] \to \mathbb{R}$ is obtained as a limit of the form

$$\int_a^b f(x)dx \doteq \lim_{n\to\infty} \sum_{i=1}^{n} f(y_i(x_i, x_{i-1}))(x_i - x_{i-1}) \tag{5.6}$$

where

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b.$$

Similarly, given two continuous functions, f and g, with g being monotonically increasing, the *Riemann-Stieltjes integral* can be defined as

$$\int_a^b f(x)dg(x) \doteq \lim_{n\to\infty} \sum_{i=1}^{n} f(y_i(x_i, x_{i-1}))(g(x_i) - g(x_{i-1})) \tag{5.7}$$

If $g(x)$ is continuously differentiable, this can be evaluated as

$$\int_a^b f(x)dg(x) = \int_a^b f(x)g'(x)dx.$$

The *Itô integral* is defined analogously as [4]:

$$\boxed{\int_{t_0}^{t} f(\tau)dw(\tau) \doteq \lim_{n\to\infty}\sum_{i=1}^{n} f(t_{i-1})[w(t_i)-w(t_{i-1})]}. \tag{5.8}$$

It has some counter-intuitive features. For example, in usual Calculus

$$\int_a^b x dx = \frac{1}{2}(b^2-a^2) \ \text{ and } \ \int_a^b f(x)df(x) = \frac{1}{2}(f(b)^2 - f(a)^2),$$

and more generally

$$\int_a^b [f(x)]^n df(x) = \frac{1}{n+1}([f(b)]^{n+1} - [f(a)]^{n+1}). \tag{5.9}$$

But

$$\int_{t_0}^{t} w(\tau)dw(\tau) = \frac{1}{2}[w(t)^2 - w(t_0)^2 - (t-t_0)].$$

And more generally [4]:

$$\int_{t_0}^{t}[w(\tau)]^n dw(\tau) = \frac{1}{n+1}([w(t)]^{n+1} - [w(t_0)]^{n+1}) - \frac{n}{2}\int_{t_0}^{t}[w(t)]^{n-1}dt. \tag{5.10}$$

5.2.1 Itô Stochastic Differential Equations in $\mathbb{R}^d$

Consider the system of d stochastic differential equations (SDEs):

$$\boxed{dx_i(t) = h_i(x_1(t), ..., x_d(t), t)dt + \sum_{j=1}^{m} H_{ij}(x_1(t), ..., x_d(t), t)dw_j(t).} \tag{5.11}$$

The "solution" is

$$x_i(t) - x_i(0) = \int_0^t h_i(x_1(\tau), ..., x_d(\tau), \tau)d\tau \tag{5.12}$$

$$\tag{5.13}$$

$$+ \sum_{j=1}^{m} \int_0^t H_{ij}(x_1(\tau), ..., x_d(\tau), \tau)dw_j(\tau), \tag{5.14}$$

is interpreted as in (5.8).

5.2.2 Numerical Approximations

Sample paths are generated from $t = 0$ to a particular end time $t = T$, and the values t_k are taken to be $t_k = Tk/n$:

$$\begin{aligned}\hat{x}_i(T) - x_i(0) = &\frac{1}{n}\sum_{k=1}^{n} h_i(\hat{x}_1(t_{k-1}), ..., \hat{x}_d(t_{k-1}), t_{k-1}) \\ &+ \sum_{j=1}^{m}\sum_{k=1}^{n} H_{ij}(\hat{x}_1(t_{k-1}), ..., \hat{x}_d(t_{k-1}), t_{k-1})[w_j(t_k) - w_j(t_{k-1})].\end{aligned} \tag{5.15}$$

In practice, not only the end value $\hat{x}_i(T)$ is of interest, but rather all values $\hat{x}_i(t_k)$, and so (5.15) is calculated along a whole *sample path* using the Euler-Maruyama approach by observing that the increments follow the rule

$$\begin{aligned}\hat{x}_i(t_k) - \hat{x}_i(t_{k-1}) = &\frac{1}{n} h_i(\hat{x}_1(t_{k-1}), ..., \hat{x}_d(t_{k-1}), t_{k-1}) \\ &+ \sum_{j=1}^{m} H_{ij}(\hat{x}_1(t_{k-1}), ..., \hat{x}_d(t_{k-1}), t_{k-1})[w_j(t_k) - w_j(t_{k-1})],\end{aligned} \tag{5.16}$$

which is basically a localized version of Itô 's rule, and provides a numerical way to evaluate (5.11) at discrete values of time.

Figure 5.1 shows six sample paths of a Wiener process over the period of time $0 \leq t \leq 1$ generated using the Matlab™ code provided in [5].

5.2.3 Mathematical Properties of the Itô Integral

Returning now to the "exact" mathematical treatment of SDEs interpreted by Itô 's rule, recall that all equalities are interpreted as being true in the mean-squared sense. In other words, the statement

$$\int_0^t F(\tau)dw_j(\tau) = \lim_{n\to\infty} \sum_{k=1}^{n} F(t_{k-1})[w_j(t_k) - w_j(t_{k-1})] \tag{5.17}$$

is not strictly true. But if we understand this to be shorthand for

$$\lim_{n\to\infty} \left\langle \left[\int_0^t F(\tau)dw_j(\tau) - \sum_{k=1}^{n} F(t_{k-1})[w_j(t_k) - w_j(t_{k-1})] \right]^2 \right\rangle = 0, \tag{5.18}$$

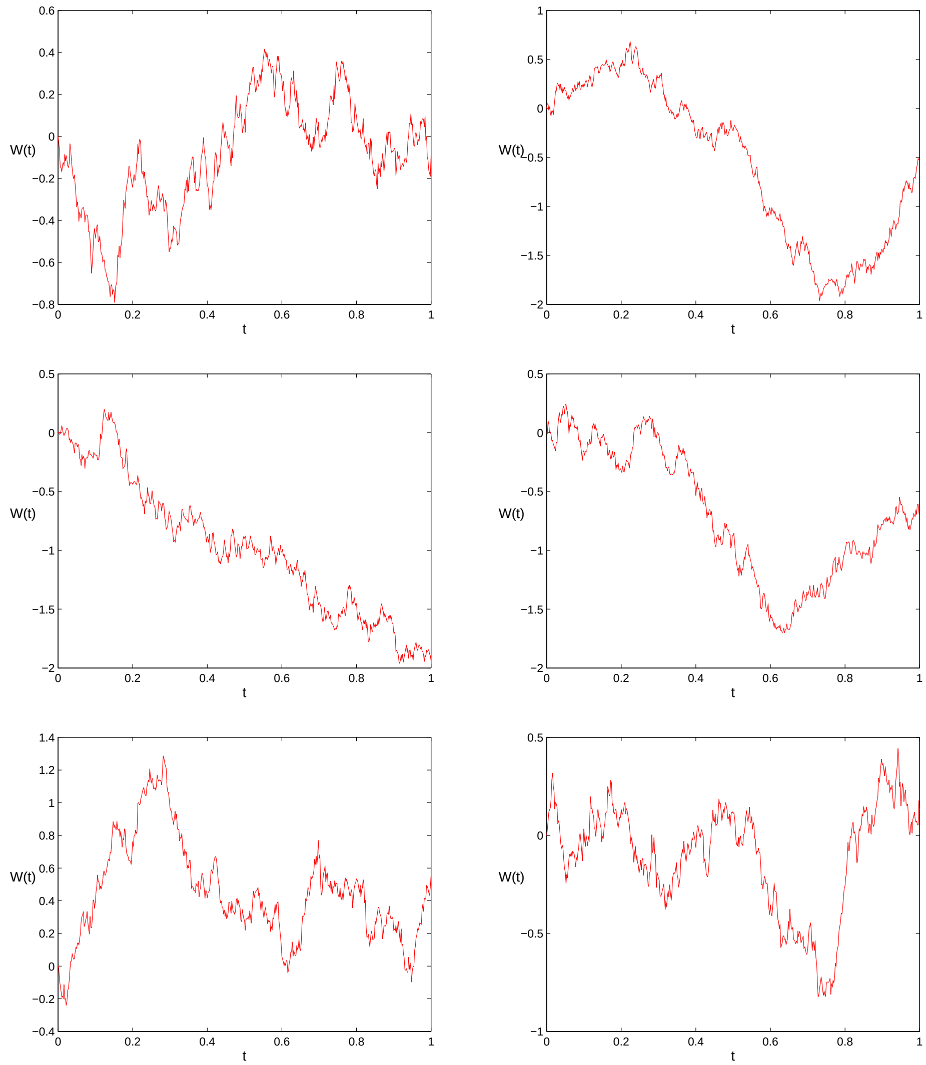

Fig. 5.1. Sample Paths of a Wiener Process

then a number of "equalities" will follow (in the same sense that (5.17) itself is an "equality").

5.2.4 Itô's Rule

Given the sample paths, $\mathbf{x}(t)$ and a smooth function $\mathbf{f}(\mathbf{x})$, then $d\mathbf{y} = \mathbf{f}(\mathbf{x} + d\mathbf{x}) - \mathbf{f}(\mathbf{x})$ can be calculated by expanding $\mathbf{f}(\mathbf{x} + d\mathbf{x})$ in a Taylor series around $\mathbf{x}$:

$$dy_i = \sum_j \frac{\partial f_i}{\partial x_j} dx_j + \frac{1}{2} \sum_{k,l} \frac{\partial f_i^2}{\partial x_k \partial x_l} dx_k dx_l + \text{h.o.t's.} \tag{5.19}$$

The higher order terms (h.o.t.'s) are third order and higher in the increments dx_i. Substituting an SDE of the form (5.11) into (5.19) gives *Itô 's rule*:

$$
\begin{aligned}
dy_i = &\left(\sum_j \frac{\partial f_i}{\partial x_j} h_j(\mathbf{x}, t) + \frac{1}{2} \sum_{k,l} \frac{\partial f_i^2}{\partial x_k \partial x_l} [H(\mathbf{x}, t) H^T(\mathbf{x}, t)]_{kl} \right) dt \\
&+ \sum_{k,l} \frac{\partial f_i}{\partial x_k} H_{kl}(\mathbf{x}, t) dw_l \qquad (5.20)
\end{aligned}
$$

The reason why the higher order terms disappear is that the sense of equality used here is that of *equality under expectation*. In other words, $a = b$ is shorthand for $\langle ac \rangle = \langle bc \rangle$ for any deterministic c. And taking expectations using the results of the previous subsection means that all terms that involve third-order and higher powers of dw_i as well as products such as $dtdw_i$ will vanish.

5.2.5 The Fokker-Planck Equation (Itô Version)

The goal of this section is to review the Fokker-Planck equation, which governs the evolution of the pdf $f(\mathbf{x}, t)$ for a given stationary Markov process, e.g., for a system of the form in (5.11) which is forced by a Wiener process.

$$\frac{\partial f(\mathbf{x},t)}{\partial t} + \sum_{i=1}^{d} \frac{\partial}{\partial x_i} \left(h_i(\mathbf{x},t) f(\mathbf{x},t)\right) - \frac{1}{2} \sum_{k=1}^{m} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} \left(H_{ik}(\mathbf{x},t) H_{kj}^T(\mathbf{x},t) f(\mathbf{x},t)\right) = 0 \tag{5.21}$$

This can also be written as

$$\frac{\partial f}{\partial t} = -\nabla_{\mathbf{x}} \cdot (\mathbf{h} f) + \frac{1}{2} \mathrm{tr} \left[(\nabla_{\mathbf{x}} \nabla_{\mathbf{x}}^T)(H H^T f) \right] \tag{5.22}$$

where $(\nabla_{\mathbf{x}} \nabla_{\mathbf{x}}^T)_{ij} = \partial^2 / \partial x_i \partial x_j$.

5.3 The Stratonovich Stochastic Calculus

The *Stratonovich stochastic integral* is defined as [19, 4]

$$\boxed{\int_{t_0}^{t} f(\tau) \circledS\, dw(\tau) \doteq \lim_{n\to\infty} \sum_{i=1}^{n} f((t_{i-1}+t_i)/2)[w(t_i) - w(t_{i-1})]} \tag{5.23}$$

here the function $f(t)$ can be of the form $f(t) = F(\mathbf{x}(t), t)$ where $\mathbf{x}(t)$ is governed by a stochastic differential equation which itself is defined by an integral like the one in (5.23).

The inclusion of the symbol $\circledS$ inside the integral is to distinguish it from the Itô integral, because in general

$$\int_{t_0}^{t} f(\tau) \circledS\, dw(\tau) \neq \int_{t_0}^{t} f(\tau)\, dw(\tau).$$

Though these two integrals are generally not equal, it is always possible to convert one into the other.

Consider the system of d stochastic differential equations (SDEs):

$$dx_i(t) = h_i^s(x_1(t), ..., x_d(t), t)dt + \sum_{j=1}^{m} H_{ij}^s(x_1(t), ..., x_d(t), t) \circledS \, dw_j(t).$$

This is called a *Stratonovich SDE* if its solution is interpreted as the integral

$$\begin{aligned} x_i(t) - x_i(0) = & \int_0^t h_i^s(x_1(\tau), ..., x_d(\tau), \tau)d\tau \\ & + \sum_{j=1}^{m} \int_0^t H_{ij}^s(x_1(\tau), ..., x_d(\tau), \tau) \circledS \, dw_j(\tau) \end{aligned} \tag{5.24}$$

In vector form this is written as

$$\mathbf{x}(t) - \mathbf{x}(0) = \int_0^t \mathbf{h}^s(\mathbf{x}(\tau), \tau)d\tau + \int_0^t H^s(\mathbf{x}(\tau), \tau) \circledS \, d\mathbf{w}(\tau).$$

Expanding everything out in a multi-dimensional Taylor series and using Itô 's rule then establishes the following equivalence between Itô and Stratonovich integrals:

$$\int_0^t H^s(\mathbf{x}(\tau),\tau) \circledS\, d\mathbf{w}(\tau) = \int_0^t H^s(\mathbf{x}(\tau),\tau)\, d\mathbf{w}(\tau) \tag{5.25}$$
$$+ \frac{1}{2}\sum_{i=1}^{d} \mathbf{e}_i \sum_{j=1}^{m}\sum_{k=1}^{d} \int_0^t \frac{\partial H^s_{ij}}{\partial x_k} H_{kj} d\tau.$$

If we *choose* to set $H_{kj} = H^s_{ij}$, then $\mathbf{x}(t)$ as defined in the Itô and Stratonovich forms will be equal if the drift terms are chosen appropriately.

In general if $\{x_1, ..., x_d\}$ is a set of Cartesian coordinates, given the Stratonovich equation (5.24), the corresponding Itô equation will be (5.11) where

$$\boxed{h_i(\mathbf{x},t) = h^s_i(\mathbf{x},t) + \frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{d} \frac{\partial H^s_{ij}}{\partial x_k} H^s_{kj} \quad \text{and} \quad H_{ij} = H^s_{ij}.} \tag{5.26}$$

This important relationship allows for the conversion between Itô and Stratonovich forms of an SDE. Using it in the reverse direction is trivial once (5.26) is known:

$$\boxed{h_i^s(\mathbf{x},t) = h_i(\mathbf{x},t) - \frac{1}{2}\sum_{j=1}^{m}\sum_{k=1}^{d}\frac{\partial H_{ij}}{\partial x_k}H_{kj} \quad \text{and} \quad H_{ij}^s = H_{ij}.} \tag{5.27}$$

Starting with the Stratonovich SDE (5.24), and using (5.26) to obtain the equivalent Itô SDE, the the Fokker-Planck equation resulting from the derivation of the Itô version can be used as an indirect way of obtaining the Stratonovich version of the Fokker-Planck equation:

$$\boxed{\frac{\partial f}{\partial t} = -\sum_{i=1}^{d}\frac{\partial}{\partial x_i}\left(h_i^s f\right) + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial}{\partial x_i}\left[\sum_{k=1}^{m} H_{ik}^s \frac{\partial}{\partial x_j}(H_{jk}^s f)\right]} \tag{5.28}$$

In the next section, a special kind of SDE is reviewed, which happens to be the same in both the Itô and Stratonovich forms.

5.3.1 Brownian Motion in the Plane

From the presentation earlier in this section, it should be clear that the following two-dimensional SDE and Fokker-Planck equation describe the same process:

$$d\mathbf{x} = d\mathbf{w} \quad \Longleftrightarrow \quad \frac{\partial f}{\partial t} = \frac{1}{2}\left(\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2}\right).$$

Coordinate Changes and the Fokker-Planck Equation

Let $\tilde{f}(r, \phi; t) = f(r\cos\phi, r\sin\phi; t)$. Then it is clear from the classical chain rule that

$$\frac{\partial \tilde{f}}{\partial r} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial r} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial r}$$

and

$$\frac{\partial \tilde{f}}{\partial \phi} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial \phi} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial \phi}.$$

If the Jacobian of the coordinate change is defined as

$$J(r,\phi) = \begin{pmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \phi} \\ \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \phi} \end{pmatrix} = \begin{pmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{pmatrix},$$

then the Jacobian determinant is $|J| = r$.

It is clear from the above equations that

$$\begin{pmatrix} \frac{\partial \tilde{f}}{\partial r} \\ \\ \frac{\partial \tilde{f}}{\partial \phi} \end{pmatrix} = J^T(r,\phi) \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \\ \frac{\partial f}{\partial x_2} \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = J^{-T}(r,\phi) \begin{pmatrix} \frac{\partial \tilde{f}}{\partial r} \\ \\ \frac{\partial \tilde{f}}{\partial \phi} \end{pmatrix}.$$

In component form this means that

$$\frac{\partial f}{\partial x_1} = \cos\phi \frac{\partial \tilde{f}}{\partial r} - \frac{\sin\phi}{r}\frac{\partial \tilde{f}}{\partial \phi}$$

and

$$\frac{\partial f}{\partial x_2} = \sin\phi \frac{\partial \tilde{f}}{\partial r} + \frac{\cos\phi}{r}\frac{\partial \tilde{f}}{\partial \phi}.$$

Applying this rule twice,

$$
\begin{aligned}
\frac{\partial^2 f}{\partial x_1^2} &= \cos\phi \frac{\partial}{\partial r}\left(\cos\phi \frac{\partial \tilde{f}}{\partial r} - \frac{\sin\phi}{r}\frac{\partial \tilde{f}}{\partial \phi}\right) - \frac{\sin\phi}{r}\frac{\partial}{\partial \phi}\left(\cos\phi \frac{\partial \tilde{f}}{\partial r} - \frac{\sin\phi}{r}\frac{\partial \tilde{f}}{\partial \phi}\right) \\
&= \cos^2\phi \frac{\partial^2 \tilde{f}}{\partial r^2} - \sin\phi\cos\phi \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial \tilde{f}}{\partial \phi}\right) + \frac{\sin^2\phi}{r}\frac{\partial \tilde{f}}{\partial r} - \\
&\quad \frac{\sin\phi\cos\phi}{r}\frac{\partial^2 \tilde{f}}{\partial \phi \partial r} + \frac{\sin\phi\cos\phi}{r^2}\frac{\partial \tilde{f}}{\partial \phi} + \frac{\sin^2\phi}{r^2}\frac{\partial^2 \tilde{f}}{\partial \phi^2}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2 f}{\partial x_2^2} &= \sin\phi \frac{\partial}{\partial r}\left(\sin\phi \frac{\partial \tilde{f}}{\partial r} + \frac{\cos\phi}{r}\frac{\partial \tilde{f}}{\partial \phi}\right) + \frac{\cos\phi}{r}\frac{\partial}{\partial \phi}\left(\sin\phi \frac{\partial \tilde{f}}{\partial r} + \frac{\cos\phi}{r}\frac{\partial \tilde{f}}{\partial \phi}\right) \\
&= \sin^2\phi \frac{\partial^2 \tilde{f}}{\partial r^2} + \sin\phi\cos\phi \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial \tilde{f}}{\partial \phi}\right) + \frac{\cos^2\phi}{r}\frac{\partial \tilde{f}}{\partial r} + \\
&\quad \frac{\sin\phi\cos\phi}{r}\frac{\partial^2 \tilde{f}}{\partial \phi \partial r} - \frac{\sin\phi\cos\phi}{r^2}\frac{\partial \tilde{f}}{\partial \phi} + \frac{\cos^2\phi}{r^2}\frac{\partial^2 \tilde{f}}{\partial \phi^2}.
\end{aligned}
$$

Therefore,

$$\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} = \frac{\partial^2 \tilde{f}}{\partial r^2} + \frac{1}{r}\frac{\partial \tilde{f}}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \tilde{f}}{\partial \phi^2},$$

and so

$$\frac{\partial f}{\partial t} = \frac{1}{2}\left(\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2}\right) \quad \Longleftrightarrow \quad \frac{\partial \tilde{f}}{\partial t} = \frac{1}{2}\left(\frac{\partial^2 \tilde{f}}{\partial r^2} + \frac{1}{r}\frac{\partial \tilde{f}}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \tilde{f}}{\partial \phi^2}\right). \tag{5.29}$$

The next question is, if $d\mathbf{x} = d\mathbf{w}$ is interpreted as a Stratonovich or Itô SDE, what will the corresponding SDEs in polar coordinates look like?

Coordinate Conversion and the Stratonovich SDE

The Stratonovich case is straightforward, since it obeys the usual Newton-Leibnitz Calculus, and so $d\mathbf{x} = J(r,\phi)[dr, d\phi]^T$. This then means that $[dr, d\phi]^T = J^{-1}(r,\phi)d\mathbf{w}$, which is written in component form as

$$
\begin{aligned}
dr &= \cos\phi \circledS dw_1 + \sin\phi \circledS dw_2 \\
d\phi &= -\frac{\sin\phi}{r} \circledS dw_1 + \frac{1}{r}\cos\phi \circledS dw_2
\end{aligned}
\tag{5.30}
$$

Coordinate Conversion and the Itô SDE

This same problem can be approached in a different way. Inverting the transformation of coordinates so that polar coordinates are written in terms of Cartesian coordinates,

$$r = [x_1^2 + x_2^2]^{\frac{1}{2}} \quad \text{and} \quad \phi = \tan^{-1}\left(\frac{x_2}{x_1}\right).$$

It follows that

$$dr = [(x_1 + dx_1)^2 + (x_2 + dx_2)^2]^{\frac{1}{2}} - [x_1^2 + x_2^2]^{\frac{1}{2}}$$

and

$$d\phi = \tan^{-1}\left(\frac{x_2 + dx_2}{x_1 + dx_1}\right) - \tan^{-1}\left(\frac{x_2}{x_1}\right).$$

Expanding the above in a Taylor series to second order in dx_i (knowing that higher order terms will vanish) gives

$$dr = \frac{1}{2}\frac{[2x_1dx_1 + (dx_1)^2 + 2x_2dx_2 + (dx_2)^2]}{[x_1^2 + x_2^2]^{\frac{1}{2}}} - \frac{1}{8}\frac{[4x_1^2(dx_1)^2 + 4x_2^2(dx_2)^2]}{[x_1^2 + x_2^2]^{\frac{3}{2}}}$$

and

$$d\phi = \frac{x_1dx_2 - x_2dx_1 + \frac{x_2}{x_1}(dx_1)^2}{x_1^2 + x_2^2} - \frac{x_2x_1^3(x_1^{-1}(dx_2)^2 + x_2^2x_1^{-4}(dx_1)^2}{(x_1^2 + x_2^2)^2}.$$

Now making the substitutions $x_1 = r\cos\phi$, $x_2 = r\sin\phi$, $dx_1 = dw_1$, $dx_2 = dw_2$, and using the usual properties of the Wiener process, this reduces (after some trigonometric simplifications) to

$$\begin{aligned} dr &= \frac{1}{2}r^{-1}dt + \cos\phi dw_1 + \sin\phi dw_2 \\ d\phi &= -r^{-1}\sin\phi dw_1 + r^{-1}\cos\phi dw_2 \end{aligned} \tag{5.31}$$

5.3.2 General Conversion Rules

Formulas were given in Section 5.3 for converting between Itô and Stratonovich versions of the same underlying process described in Cartesian coordinates. The same rules hold for this conversion in curvilinear coordinates.

In general if $\{q_1, ..., q_d\}$ is a set of generalized coordinates, given the Stratonovich equation

$$dq_i = h_i^s(\mathbf{q}, t)dt + \sum_{j=1}^{m} H_{ij}^s(\mathbf{q}, t) \circledS \, dw_j$$

for $i = 1, ..., d$ the corresponding Itô equation will be

$$dq_i = h_i(\mathbf{q}, t)dt + \sum_{j=1}^{m} H_{ij}(\mathbf{q}, t)dw_j$$

where

$$h_i(\mathbf{q}, t) = h_i^s(\mathbf{q}, t) + \frac{1}{2} \sum_{j=1}^{m} \sum_{k=1}^{d} \frac{\partial H_{ij}^s}{\partial q_k} H_{kj}^s. \tag{5.32}$$

In the above example of Brownian motion in the plane, the Stratonovich equation (5.30) has no drift, and the corresponding Itô equation (5.31) does have a drift, which is consistent with $h_i(\mathbf{q}, t) \neq h_i^s(\mathbf{q}, t)$.

Now consider the Stratonovich equivalent of the Itô equation (5.31). Using (5.32), it becomes clear that

$$\begin{pmatrix} dr \\ d\phi \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1/r \\ 0 \end{pmatrix} dt + \begin{pmatrix} 1 & 0 \\ 0 & 1/r \end{pmatrix} \circledS \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix} \tag{5.33}$$

$$
\begin{tikzcd}
 & \boxed{\begin{array}{c}\text{Strat. FP}\\ \text{Eq., Cart.}\end{array}} \arrow[rr, leftrightarrow] \arrow[dl, equal] \arrow[dd, leftrightarrow] & & \boxed{\begin{array}{c}\text{Strat. SDE}\\ \text{Cart.}\end{array}} \arrow[dl, equal] \arrow[dd, leftrightarrow] \\
\boxed{\begin{array}{c}\text{Itô FP}\\ \text{Eq., Cart.}\end{array}} \arrow[rr] \arrow[dd, leftrightarrow] & & \boxed{\begin{array}{c}\text{Itô SDE}\\ \text{Cart.}\end{array}} \arrow[dd, leftrightarrow] & \\
 & \boxed{\begin{array}{c}\text{Strat. FP}\\ \text{Eq., Polar}\end{array}} \arrow[rr, leftrightarrow] \arrow[dl, equal] & & \boxed{\begin{array}{c}\text{Strat. SDE}\\ \text{Polar}\end{array}} \arrow[dl, leftrightarrow] \\
\boxed{\begin{array}{c}\text{Itô FP}\\ \text{Eq., Polar}\end{array}} \arrow[rr, leftrightarrow] & & \boxed{\begin{array}{c}\text{Itô SDE}\\ \text{Polar}\end{array}} &
\end{tikzcd}
\tag{5.34}
$$

5.3.3 Coordinate Changes and Fokker-Planck Equations

Itô Version

The Itô version of the Fokker-Planck equation in generalized coordinates is

$$\frac{\partial f}{\partial t} = -|G|^{-\frac{1}{2}} \sum_{i} \frac{\partial}{\partial q_i} \left(a_i |G|^{\frac{1}{2}} f \right) + \frac{1}{2} |G|^{-\frac{1}{2}} \sum_{i,j} \frac{\partial^2}{\partial q_i \partial q_j} \left[(BB^T)_{ij} |G|^{\frac{1}{2}} f \right]. \tag{5.35}$$

Given $f(\mathbf{q}, 0)$ this generates $f(\mathbf{q}, t)$ for the Itô SDE

$$d\mathbf{q} = \mathbf{a}(\mathbf{q}, t) + B(\mathbf{q}, t) d\mathbf{w}.$$

When $B(\mathbf{q}, t) = [J(\mathbf{q})]^{-1}$, (5.35) will be the heat equation under special conditions on $\mathbf{a}(\mathbf{q}, t)$.

Stratonovich Version

$$\frac{\partial f}{\partial t} = -|G|^{-\frac{1}{2}} \sum_i \frac{\partial}{\partial q_i} \left(a_i^s |G|^{\frac{1}{2}} f \right) + \frac{1}{2} |G|^{-\frac{1}{2}} \sum_{i,j,k} \frac{\partial}{\partial q_i} \left[B_{ik}^s \frac{\partial}{\partial q_j} (B_{jk}^s |G|^{\frac{1}{2}} f) \right] \tag{5.36}$$

Given $f(\mathbf{q}, 0)$ this generates $f(\mathbf{q}, t)$ for the Stratonovich SDE

$$d\mathbf{q} = \mathbf{a}^s(\mathbf{q}, t) + B^s(\mathbf{q}, t) \, \text{Ⓢ} \, d\mathbf{w}.$$

References

1. Einstein, A., *Investigations on the Theory of the Brownian Movement*, Dover, New York, 1956.
2. Fokker, A.D., "Die Mittlere Energie rotierender elektrischer Dipole in Strahlungs Feld," *Ann. Physik.*, Vol. 43, pp. 810-820, 1914.
3. Gard, T.C., *Introduction to Stochastic Differential Equations*, Marcel Dekker, 1988.
4. Gardiner, C.W., *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences* 3^{rd} ed., Springer-Verlag, Berlin, 2004.
5. Higham, D.J., "An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations," *SIAM Review*, Vol. 43, No. 3, pp. 525-546, 2001
6. Itô, K., McKean, H.P., *Diffusion Processes and their Sample Paths*, Springer-Verlag, 1974.
7. Kloedon, P.E., Platen, E., *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin, 1992.
8. Knight, F.B., *Essentials of Brownian Motion*, *Math. Survey* 18, American Mathematical Society, Providence, RI, 1981.

9. Kuo, H.-H., *Introduction to Stochastic Integration*, Springer, New York, 2006.
10. Langevin, P., “Sur la théorie du mouvement brownien,” *Comptes. Rendues Acad. Sci. Paris*, Vol. 146, pp. 530- 533, 1908.
11. Maruyama, G., “Continuous Markov processes and Stochastic equations,” *Rend. Circ. Mat. Palermo*, 4, 48-90 (1955).
12. McKean, H.P. Jr., *Stochastic Integrals*, Academic Press, New York, 1969.
13. Millstein, G.N., Tretyakov, M.V., *Stochastic Numerics for Mathematical Physics*, Springer-Verlag, Berlin, 2004.
14. Øksendal, B., *Stochastic Differential Equations, An Introduction with Applications*, 5^{th} ed., Springer, Berlin, 1998.
15. Onsager, L., “Reciprocal Relations in Irreversible Processes. I.,II,” *Physical Review*, 37:405-426, 38:2265-2280. 1931.
16. Planck, M., “Uber einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie,” *Sitz. ber. Berlin A Akad. Wiss.*, 1917, pp. 324-341, 1917.
17. Protter, P., *Stochastic Integration and Differential Equations*, Springer 1990.
18. Risken, H., *The Fokker-Planck Equation, Methods of Solution and Applications*, 2^{nd} ed., Springer-Verlag, Berlin, 1989.
19. Stratonovich, R.L., *Topics in the Theory of Random Noise* (vol. I., II.), (translated by R. A. Silverman) Gordon and Breach, New York; London, 1963.

6

Geometry of Curves and Surfaces

The main points to take away from this section are:

- Curves and surfaces in two- and three-dimensional space can be described parametrically or implicitly, and the local geometry is described by intrinsic quantities that are independent of the particular description.
- The global topological features of these geometric objects can be related to integrals of curvature; In particular the Euler characteristic describes how many "holes" there are in an object, and the integrals of certain kinds of curvature over a tubular surface can help to determine whether it is knotted or not.

6.1 Differential Geometry of Curves

Differential geometry is concerned with characterizing the local shape of curves and surfaces using the tools of differential Calculus, and relating these local shape properties at each point on the object of interest to its global characteristics.

6.1.1 Local Theory of Curves

The *arc length* of a differentiable space curve, $\mathbf{x}(t)$ is

$$s(t_2) - s(t_1) = \int_{t_1}^{t_2} (\mathbf{x}'(t), \mathbf{x}'(t))^{\frac{1}{2}} dt \tag{6.1}$$

where $\mathbf{x}' = d\mathbf{x}/dt$.

A *unit tangent* is

$$\mathbf{u}(t) \doteq \frac{1}{\left\| \frac{d\mathbf{x}}{dt} \right\|} \frac{d\mathbf{x}}{dt}.$$

When $t = s$, this reduces to

$$\mathbf{u}(s) = \frac{d\mathbf{x}}{ds}.$$

Since $\mathbf{u}(s)$ is a unit vector $\mathbf{u}(s) \cdot \mathbf{u}(s) = 1$, and so

$$\frac{d}{ds}(\mathbf{u} \cdot \mathbf{u}) = 0 \quad \Longrightarrow \quad \mathbf{u} \cdot \frac{d\mathbf{u}}{ds} = 0. \tag{6.2}$$

The (unsigned) *curvature* of an arc-length-parameterized curve (planar or spatial) is defined as

$$\kappa(s) \doteq \left(\frac{d\mathbf{u}}{ds} \cdot \frac{d\mathbf{u}}{ds}\right)^{\frac{1}{2}} = \left(\frac{d^2\mathbf{x}}{ds^2} \cdot \frac{d^2\mathbf{x}}{ds^2}\right)^{\frac{1}{2}}, \tag{6.3}$$

which is a measure of the amount of change in tangent direction at each value of arc length.

The *signed curvature* of a planar curve is denoted as $k(s)$.

By defining the *(principal) normal* vector as

$$\mathbf{n}_1(s) \doteq \frac{1}{\kappa(s)} \frac{d\mathbf{u}}{ds} \tag{6.4}$$

when $\kappa(s) = \|d\mathbf{u}/ds\| \neq 0$, it follows from (6.2) that

$$\mathbf{u}(s) \cdot \mathbf{n_1}(s) = 0.$$

A second normal vector (called the binormal) is defined as:

$$\mathbf{n}_2(s) \doteq \mathbf{u}(s) \times \mathbf{n}_1(s). \tag{6.5}$$

The *torsion* of the curve is defined as

$$\tau(s) \doteq -\frac{d\mathbf{n}_2(s)}{ds} \cdot \mathbf{n}_1(s)$$

and is a measure of how much the curve bends out of the $(\mathbf{u}, \mathbf{n}_1)$- plane at each s.

The Frenet-Serret apparatus published independently by Frenet (1852) and Serret (1851) states:

$$\boxed{\frac{d}{ds}\begin{pmatrix} \mathbf{u}(s) \\ \mathbf{n}_1(s) \\ \mathbf{n}_2(s) \end{pmatrix} = \begin{pmatrix} 0 & \kappa(s) & 0 \\ -\kappa(s) & 0 & \tau(s) \\ 0 & -\tau(s) & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}(s) \\ \mathbf{n}_1(s) \\ \mathbf{n}_2(s) \end{pmatrix}.} \tag{6.6}$$

Given a space curve $\mathbf{x}(t) \in \mathbb{R}^3$ where t is not necessarily arc length, the (unsigned) curvature is computed as

$$\kappa(t) = \frac{\|\mathbf{x}'(t) \times \mathbf{x}''(t)\|}{\|\mathbf{x}'(t)\|^3} \tag{6.7}$$

and the torsion is

$$\tau(t) = \frac{\det[\mathbf{x}'(t), \mathbf{x}''(t), \mathbf{x}'''(t)]}{\|\mathbf{x}'(t) \times \mathbf{x}''(t)\|^2} \tag{6.8}$$

6.1.2 Global Theory of Curves

Theorem 6.1. *(Fenchel [9]): For smooth closed curves,*

$$\oint \kappa(s)ds \geq 2\pi \tag{6.9}$$

with equality holding only for some kinds of planar ($\tau(s) = 0$) curves.

Theorem 6.2. *(Fary-Milnor [8, 20]): For closed space curves forming a knot*

$$\oint \kappa(s)ds \geq 4\pi. \tag{6.10}$$

6.1.3 Signed Curvature and the Topology of Planar Regions

$$\oint k(s)ds = 2\pi\,, \tag{6.11}$$

where $k(s)$ is the signed curvature of the curve. The sign is given such that $|k(s)| = \kappa(s)$ with $k(s) > 0$ for counterclockwise bending and $k(s) < 0$ for clockwise bending.

The *Euler characteristic* of B, denoted as $\chi(B)$, is obtained by subdividing, or tessellating, the body into disjoint polygonal regions, the union of which is the body, counting the number of polygonal faces, f, edges, e, and vertices, v, and using the formula

$$\chi(B) = v(B) - e(B) + f(B). \tag{6.12}$$

Interestingly, for a planar body

$$\chi(B) = 1 - \gamma(B). \tag{6.13}$$

Whereas $\gamma(B)$ is the number of holes in the body.

It can be shown that

$$\int_{\partial B} k(s)ds = 2\pi\chi(B) \tag{6.14}$$

where ∂B denotes the union of all boundary curves of B. This is shown in Figure 6.1.

For a planar object, it is also possible to define the Euler characteristic of the boundary as

$$\chi(\partial B) = v(\partial B) - e(\partial B). \tag{6.15}$$

6.2 Differential Geometry of Surfaces in $\mathbb{R}^3$

For a *spatial body* B (i.e., a region in $\mathbb{R}^3$ with finite nonzero volume), the surface area over the boundary of B is

$$F = \int_{\partial B} dS,$$

and volume is

$$V = \int_B dV.$$

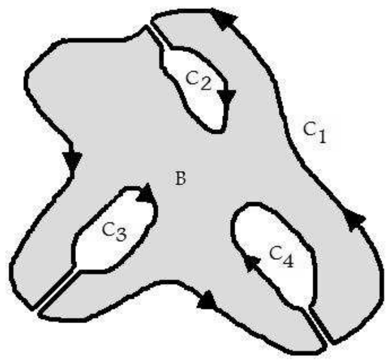

Fig. 6.1. Global Topological Features of a Planar Body are Dictated By Signed Curvature

6.2.1 The First and Second Fundamental Forms

Consider a two-dimensional surface parameterized as $\mathbf{x}(\mathbf{q})$ where $\mathbf{x} \in \mathbb{R}^3$ and $\mathbf{q} \in \mathbb{R}^2$. Define

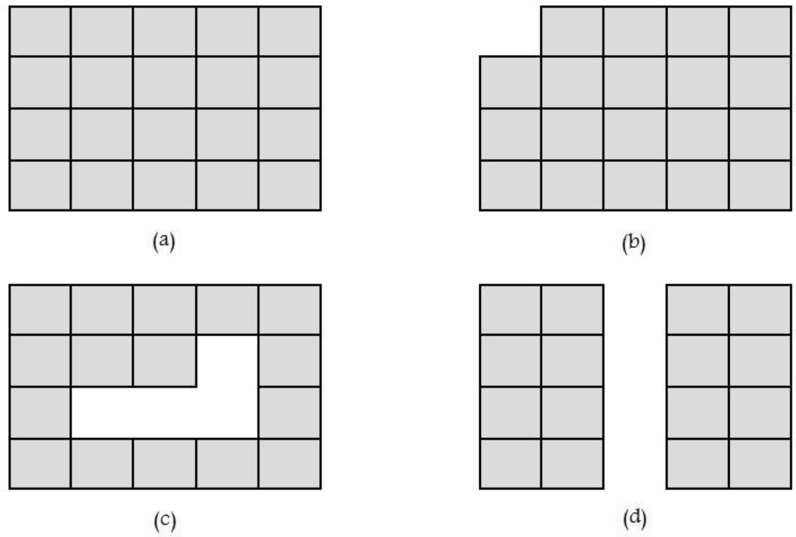

Fig. 6.2. Topological Operations on Body Divided into Squares: (a) An Initial Rectangular Grid; (b) Removal of One Square from the Perimeter; (c) Creation of an L-Shaped Void; (d) Cutting the Body into Two Disjoint Pieces.

$$g_{ij} = \frac{\partial \mathbf{x}}{\partial q_i} \cdot \frac{\partial \mathbf{x}}{\partial q_j} \tag{6.16}$$

for $i, j \in \{1, 2\}$.

The *first fundamental form* of the surface is defined as

$$\mathcal{F}^{(1)}(d\mathbf{q}, d\mathbf{q}) \doteq d\mathbf{q}^T G(\mathbf{q})\, d\mathbf{q}. \tag{6.17}$$

For two-dimensional surfaces in three-dimensional space,

$$|G(q_1, q_2)| = \left\| \frac{\partial \mathbf{x}}{\partial q_1} \times \frac{\partial \mathbf{x}}{\partial q_2} \right\|^2 \tag{6.18}$$

where $\times$ denotes the vector cross product.

The *second fundamental form* of a surface is defined as

$$\mathcal{F}^{(2)}(d\mathbf{q}, d\mathbf{q}) = -d\mathbf{x} \cdot d\mathbf{n},$$

where the vectors $\mathbf{x}$ and $\mathbf{n}$ are the position and normal at any point on the surface.

Let the matrix L be defined by its entries:

$$L_{ij} = \frac{\partial^2 \mathbf{x}}{\partial q_i \partial q_j} \cdot \mathbf{n}. \tag{6.19}$$

The matrix $L = [L_{ij}]$ contains information about how curved the surface is. For example, for a plane $L_{ij} = 0$.

It can be shown that

$$\mathcal{F}^{(2)}(d\mathbf{q}, d\mathbf{q}) = d\mathbf{q}^T L(\mathbf{q})\, d\mathbf{q}. \tag{6.20}$$

6.2.2 Curvature

Christoffel symbols:

$$\Gamma_{ij}^k = \frac{1}{2} \sum_l \left[\frac{\partial g_{il}}{\partial q_j} - \frac{\partial g_{ij}}{\partial q_l} + \frac{\partial g_{lj}}{\partial q_i} \right] g^{lk}. \tag{6.21}$$

The *Riemannian curvature* is the four-index tensor given in component form as [19]

$$\boxed{R^l_{ijk} \doteq \frac{\partial \Gamma^l_{ik}}{\partial q_j} - \frac{\partial \Gamma^l_{ij}}{\partial q_k} + \sum_m (\Gamma^m_{ik}\Gamma^l_{mj} - \Gamma^m_{ij}\Gamma^l_{mk}).} \tag{6.22}$$

This can be expressed in terms of the coefficients of the second fundamental form and inverse of the metric tensor as [19]:

$$R^l_{ijk} = L_{ik} \sum_m g^{lm} L_{mj} - L_{ij} \sum_m g^{lm} L_{mk}. \tag{6.23}$$

From this, the *Gaussian curvature*, $k(q_1, q_2)$, is computed as

$$k = \det(G^{-1}L) = |G|^{-\frac{1}{2}} R_{1212}. \tag{6.24}$$

The *mean sectional curvature* (or simply *mean curvature*) is defined as

$$\boxed{m \doteq \frac{1}{2}\mathrm{trace}(G^{-1}L).} \tag{6.25}$$

The integrals of the Gaussian and mean curvature over the entirety of a a closed surface figure are

$$K \doteq \int_S k\, dS \tag{6.26}$$

$$M \doteq \int_S m\, dS \tag{6.27}$$

These are respectively called the *total Gaussian curvature* and *total mean curvature*

6.2.3 Example 1: The Sphere

A sphere of radius R can be parameterized as

$$\mathbf{x}(\phi, \theta) = \begin{pmatrix} R\cos\phi\sin\theta \\ R\sin\phi\sin\theta \\ R\cos\theta \end{pmatrix} \tag{6.28}$$

where $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$.

The corresponding metric tensor is

$$G(\phi, \theta) = \begin{pmatrix} g_{\phi,\phi} & g_{\phi,\theta} \\ g_{\theta,\phi} & g_{\theta,\theta} \end{pmatrix} = \begin{pmatrix} R^2 \sin^2\theta & 0 \\ 0 & R^2 \end{pmatrix}.$$

Clearly, $\sqrt{\det G(\phi,\theta)} = R^2 \sin\theta$ (there is no need for absolute value signs since $\sin\theta \geq 0$ for $\theta \in [0,\pi]$). The element of surface area is therefore

$$dS = R^2 \sin\theta \, d\phi \, d\theta.$$

Surface area of the sphere is computed as

$$F = \int_0^{\pi} \int_0^{2\pi} \sin\theta \, d\phi \, d\theta = 4\pi R^2.$$

The volume of the ball of radius R can be computed in spherical coordinates in $\mathbb{R}^3$ (i.e., treating $\mathbb{R}^3$ as the surface of interest) and restricting the range of parameters defined by the interior of the ball. The volume of the ball enclosed by the sphere of radius R, and surface area of the sphere are summarized, respectively, as

$$V = \frac{4}{3}\pi R^3; \quad F = 4\pi R^2. \tag{6.29}$$

The inward-pointing normal for the sphere is simply $\mathbf{n} = -\mathbf{x}/R$, and

$$L(\phi,\theta) = \begin{pmatrix} R\sin^2\theta & 0 \\ 0 & R \end{pmatrix}.$$

Therefore,

$$G^{-1}L = \begin{pmatrix} 1/R & 0 \\ 0 & 1/R \end{pmatrix}.$$

It follows that

$$m = \frac{1}{2}\mathrm{tr}(G^{-1}L) = 1/R$$

and

$$k = \det(G^{-1}L) = 1/R^2.$$

Since these are both constant, it follows that integrating each of them over the sphere of radius R is the same as their product with the surface area:

$$M = 4\pi R; \quad K = 4\pi.$$

6.2.4 Example 2: The Ellipsoid of Revolution

Consider an *ellipsoid of revolution* parameterized as

$$\mathbf{x}(\phi, \theta) = \begin{pmatrix} a\cos\phi\sin\theta \\ a\sin\phi\sin\theta \\ b\cos\theta \end{pmatrix} \tag{6.30}$$

where $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$, and a, b are positive constants.

The corresponding metric tensor is

$$G(\phi, \theta) = \begin{pmatrix} a^2\sin^2\theta & 0 \\ 0 & a^2\cos^2\theta + b^2\sin^2\theta \end{pmatrix}.$$

$$L(\phi, \theta) = |G(\phi, \theta)|^{-\frac{1}{2}} \begin{pmatrix} a^2 b\sin^2\theta & 0 \\ 0 & a^2 b\sin\theta \end{pmatrix}.$$

Therefore,

$$K = 4\pi \quad \text{and} \quad V = \frac{4}{3}\pi a^2 b.$$

The values of F and M for prolate and oblate ellipsoids have been reported in [15], along with a variety of other solids of revolution. In particular, if $a = R$ and $b = \lambda R$ with $0 < \lambda < 1$, then

$$F = 2\pi R^2 \left[1 + \frac{\lambda^2}{\sqrt{1-\lambda^2}} \log\left(\frac{1+\sqrt{1-\lambda^2}}{\lambda}\right)\right]; \quad M = 2\pi R \left[\lambda + \frac{\arccos\lambda}{\sqrt{1-\lambda^2}}\right].$$

In contrast, when $\lambda > 1$,

$$F = 2\pi R^2 \left[1 + \frac{\lambda^2 \arccos(1/\lambda)}{\sqrt{\lambda^2-1}}\right]; \quad M = 2\pi R \left[\lambda + \frac{\log(\lambda + \sqrt{\lambda^2-1})}{\sqrt{\lambda^2-1}}\right].$$

6.2.5 Example 3: The Torus

The 2-torus can be parameterized as

$$\mathbf{x}(\theta, \phi) = \begin{pmatrix} (R + r\cos\theta)\cos\phi \\ (R + r\cos\theta)\sin\phi \\ r\sin\theta \end{pmatrix} \tag{6.31}$$

where $R > 2r$ and $0 \leq \theta, \phi \leq 2\pi$.

The metric tensor for the torus is written in this parametrization as

$$G(\phi, \theta) = \begin{pmatrix} (R + r\cos\theta)^2 & 0 \\ 0 & r^2 \end{pmatrix}.$$

The surface area is computed directly as

$$F = \int_0^{2\pi} \int_0^{2\pi} r(R + r\cos\theta)d\phi d\theta = 2\pi \int_0^{2\pi} r(R + r\cos\theta)d\theta = 4\pi^2 rR.$$

The matrix L is:

$$L(\phi, \theta) = \begin{pmatrix} (R + r\cos\theta)\cos\theta & 0 \\ 0 & r \end{pmatrix},$$

and

$$G^{-1}L = \begin{pmatrix} (R + r\cos\theta)^{-1}\cos\theta & 0 \\ 0 & 1/r \end{pmatrix}.$$

The total Gaussian curvature is then computed as

$$K = \int_0^{2\pi} \int_0^{2\pi} [(R + r\cos\theta)^{-1}\cos\theta/r][r(R + r\cos\theta)]d\phi d\theta = 0.$$

The mean curvature is $m = (R + r\cos\theta)^{-1}\cos\theta + 1/r$. The total mean curvature is

$$M = \frac{F}{2r} = 2\pi^2 R.$$

6.2.6 The Gauss-Bonnet Theorem and Related Inequalities

It is no coincidence that the total Gaussian curvature, K, is equal to 4π for the sphere and ellipsoid, and equal to zero for a torus.

Theorem 6.3. *(Gauss-Bonnet) Let k be the Gaussian curvature of a closed surface S. Then*

$$\boxed{\int_S k\, dS = 2\pi\chi(S),} \tag{6.32}$$

where $\chi(S)$ is the Euler characteristic of the closed surface S.

The Euler characteristic of a two-dimensional surface is equal to

$$\boxed{\chi(S) = 2(1 - \gamma(S))} \tag{6.33}$$

where $\gamma(S)$ is the genus (or "number of donut holes") of the surface and

$$\chi(S) = v - e + f$$

where v is the number of vertices, e is the number of edges, and f is the number of faces of the polygons.

Other global theorems (Voss): [25]:

$$\int_S \max(k, 0) dS \geq 4\pi. \tag{6.34}$$

$$\int_S |k| dS \geq \int_S \max(k, 0) dS \geq 4\pi. \tag{6.35}$$

Moreover, B.-Y. Chen [2] states the *Chern-Lashof* inequality

$$\int_S |k| dS \geq 4\pi(4 - \chi(S)) = 8\pi(1 + \gamma(S)). \tag{6.36}$$

Integrals of the square of mean curvature have resulted in several inequalities. For example, Wilmore (see e.g., [27, 28] and references therein) proved that

$$\int_S m^2 dS \geq 4\pi \tag{6.37}$$

with equality holding only for the usual sphere in $\mathbb{R}^3$. Shiohama and Takagi proved that for smoothly distorted 2-tori

$$\int_{T^2} m^2 dS \geq 2\pi^2, \tag{6.38}$$

Ros's Theorem: Let D be a bounded domain in $\mathbb{R}^3$ with finite volume and compact boundary ∂D. If $m > 0$ everywhere on this boundary then

$$\int_{\partial D} \frac{1}{m} dS \geq 3 \cdot Vol(D). \tag{6.39}$$

6.2.7 Tubes/Offsets of Surfaces in $\mathbb{R}^3$

Given a smooth parameterized surface, $\mathbf{x}(t_1, t_2)$, a unit normal can be defined to the surface at each point as

$$\mathbf{u}(t_1, t_2) = \frac{\partial \mathbf{x}}{\partial t_1} \times \frac{\partial \mathbf{x}}{\partial t_2} / \left\| \frac{\partial \mathbf{x}}{\partial t_1} \times \frac{\partial \mathbf{x}}{\partial t_2} \right\|.$$

Define

$$\mathbf{o}(t_1, t_2; r) = \mathbf{x}(t_1, t_2) + r\mathbf{u}(t_1, t_2). \tag{6.40}$$

The element of surface area for this *offset surface* can be shown to be of the form

$$dS = \left\| \frac{\partial \mathbf{o}}{\partial t_1} \times \frac{\partial \mathbf{o}}{\partial t_2} \right\| dt_1 dt_2$$

where [13]

$$\left\| \frac{\partial \mathbf{o}}{\partial t_1} \times \frac{\partial \mathbf{o}}{\partial t_2} \right\| = [1 - 2rm(t_1, t_2) + r^2 k(t_1, t_2)] \left\| \frac{\partial \mathbf{x}}{\partial t_1} \times \frac{\partial \mathbf{x}}{\partial t_2} \right\|. \tag{6.41}$$

the area of the offset surface will be:

$$\boxed{A = F - 2rM + r^2 K.} \tag{6.42}$$

Steiner's formula for the volume enclosed by the surface offset is

$$\boxed{V(B_r) = V(B) + rF(\partial B) + \frac{r^2}{2} M(\partial B) + \frac{r^3}{3} K(\partial B).} \tag{6.43}$$

The volume contained within the two offset surfaces defined by $r \in [-r_0, r_0]$ is ("Weyl's tube theorem"):

$$V_o = \int_{-r_0}^{r_0} \int_S \left[\frac{\partial \mathbf{o}}{\partial r}, \frac{\partial \mathbf{o}}{\partial t_1}, \frac{\partial \mathbf{o}}{\partial t_2} \right] dt_1 dt_2 dr = 2rF + \frac{2r^3}{3} \int_S k dS = 2rF + \frac{4}{3} \pi r^3 \chi(S). \tag{6.44}$$

References

1. Bloomenthal, J., (ed.), *Introduction to Implicit Surfaces*, Morgan Kaufmann Publishers, Inc., San Francisco, CA, 1997.
2. Chen, B.-Y., *Total mean curvature and submanifolds of finite type*, World Scientific, Singapore, 1984.
3. Chen, B.-Y., "On the total curvature of immersed manifolds," *Amer. J. Math.* 93, pp. 148-162, 1971; 94, pp. 899-907, 1972; 95, pp. 636-642, 1973.
4. Chen, B.-Y., "On an inequality of mean curvature," *J. London Math. Soc.* 4, pp. 647-650, 1972.
5. do Carmo, M., *Differential Geometry of Curves and Surfaces*, Prentice-Hall, Englewood Cliffs, NJ, 1976.
6. Farouki, R.T., Neff, C.A., "Analytical Properties of Plane Offset Curves," *Computer Aided Geometric Design*, Vol. 7, pp. 83-99, 1990.
7. Farouki, R.T., Neff, C.A., "Algebraic Properties of Plane Offset Curves," *Computer Aided Geometric Design*, Vol. 7, pp. 101-127, 1990.
8. Fary, I., "Sur la courbure totale d'une courbe gauche faisant un noeud," *Bulletin de la Société Mathématique de France*, Vol. 77, pp. 128-138, 1949.

9. Fenchel, W., "Uber Krümmung und Windung geschlossenen Raumkurven," *Mathematische Annalen*, Vol. 101, pp. 238-252, 1929.
10. Fox, R.H., "On the total curvature of some tame knots," *Ann. of Math.*, 52, pp. 258-261, 1950.
11. Goldman, R., "Curvature formulas for implicit curves and surfaces," *Computer Aided Geometric Design* 22(7) pp. 632 - 658, October 2005
12. Gray, A., Abbena, E., Salamon, S., *Modern Differential Geometry of Curves and Surfaces with MATHEMATICA*, Chapman & Hall/CRC, Boca Raton, FL, 2006.
13. Gray, A., *Tubes*, 2^{nd} ed., Birkhäuser, 2004.
14. Guggenheimer, H.W., *Differential Geometry*, Dover Publications, New York, 1977.
15. Hadwiger, H., *Altes und Neues über Konvexe Körper,* Birkhäuser Verlag, Basel, 1955.
16. Kuiper, N.H., Meeks, W.H., "The Total Curvature of a Knotted Torus," *J. Differential Geometry* 26, pp. 371-384, 1987.
17. Langevin, R., Rosenburg, H., "On Curvature Integrals and Knots," *Topology*, Vol. 15, pp. 405-416, 1976.
18. Manocha, D., Canny, J., "Efficient inverse kinematics for general 6R manipulators," *IEEE Trans. Robot. Automat.*, 10:648-657, 1994.
19. Millman, R.S., Parker, G.D., *Elements of Differential Geometry*, Prentice-Hall, Englewood Cliffs, NJ 1977.
20. Milnor, J., "On the Total Curvature of Knots," *Annals of Mathematics*, Vol. 52, pp. 248-257, 1950.
21. Raghavan, M., Roth, B., "Inverse kinematics of the general 6R manipulator and related linkages," *ASME J. Mech. Design*, 115:502-508, 1993.
22. Ros, A., "Compact Surfaces with Constant Scalar Curvature and a Congruence Theorem," *J. Diff. Geom*, 27, pp. 215-220, 1988.
23. San Jose Estepar, R., Haker, S., Westin, C.F., "Riemannian mean curvature flow," in *Lecture Notes in Computer Science: ISVC05*, vol. 3804, pp. 613-620, Springer, 2005
24. Sommese, A.J., Wampler C.W., *The Numerical Solution of Systems of Polynomials Arising in Engineering And Science*, World Scientific Publishing, Singapore, 2005.
25. Voss, K., "Eine Bemerkung über die Totalkrümmung geschlossener Raumkurven," *Arch. Math.*, 6, pp. 259-263, 1955.
26. Weyl, H., "On the volume of tubes," *Amer. J. Math.* 61, 461-472, 1939.

27. Willmore, T.J., "Mean Curvature of Riemannian Immersions," *J. London Math. Soc.*, 3, pp. 307-310, 1971.
28. Willmore, T.J., "Tight Immersions and Total Absolute Curvature," *Bull. London Math. Soc.*, 3, pp. 129-151, 1971.

7

Manifolds

The main points to take away from this section are:

- The concepts of simple planar or spatial curves and simply connected surfaces in $\mathbb{R}^3$ extend to higher dimensions and are examples of more general mathematical structures called manifolds;
- Sometimes it is natural to treat these geometric objects as "living in" a higher dimensional Euclidean space, and sometimes it is more natural to use purely intrinsic approaches;

7.1 Examples of Manifolds

Example 1: The Sphere S^3 Embedded in $\mathbb{R}^4$

A simple example of a manifold resulting from a constraint equation is the unit sphere in $\mathbb{R}^4$, which is denoted as S^3, and is described in terms of Cartesian coordinates as

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1.$$

Since $\mathbb{R}^4$ is a four-dimensional space, and this is a single constraint equation, we conclude that S^3 is a $4-1=3$-dimensional manifold. Parametric equations that satisfy this constraint and "reach" every point on S^3 (as well as S^n) were given in Section 2.3.

Example 2: The Group of Motions of the Euclidean Plane

The group of planar rigid-body motions has been encountered several times earlier in this volume. Elements of this group are described using matrices of the form

$$g = \begin{pmatrix} \cos\theta & -\sin\theta & x \\ \sin\theta & \cos\theta & y \\ 0 & 0 & 1 \end{pmatrix} \quad \text{with} \quad x, y \in \mathbb{R} \quad \text{and} \quad \theta \in [0, 2\pi). \tag{7.1}$$

The set of all such matrices is called the *special Euclidean group* of the plane, and is denoted as $SE(2)$, where the "2" corresponds to the dimension of the plane. The group operation is matrix multiplication. In fact, any Lie group with elements that are matrices and which has a group operation of matrix multiplication is called a *matrix Lie group*. Therefore, when referring to a matrix Lie group, there is no need to mention the group operation, since it is understood in advance to be matrix multiplication.

Example 3: The Group of Rotations of Three-Dimensional Euclidean Space

Each element of this group is written in terms of columns as

$$R = [\mathbf{a}, \mathbf{b}, \mathbf{c}]$$

with

$$\mathbf{a} \cdot \mathbf{a} = \mathbf{b} \cdot \mathbf{b} = \mathbf{c} \cdot \mathbf{c} = 1$$

and

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} = 0.$$

Furthermore,

$$\mathbf{c} = \mathbf{a} \times \mathbf{b} \iff \det R = +1$$

This means that

$$R = [\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b}] \in SO(3).$$

3×3 rotation matrix can be parameterized using the *Euler parameters* as:

$$R(u_1, u_2, u_3, u_4) = \begin{pmatrix} u_1^2 - u_2^2 - u_3^2 + u_4^2 & 2(u_1u_2 - u_3u_4) & 2(u_3u_1 + u_2u_4) \\ 2(u_1u_2 + u_3u_4) & u_2^2 - u_3^2 - u_1^2 + u_4^2 & 2(u_2u_3 - u_1u_4) \\ 2(u_3u_1 - u_2u_4) & 2(u_2u_3 + u_1u_4) & u_3^2 - u_1^2 - u_2^2 + u_4^2 \end{pmatrix} \tag{7.2}$$

where

$$u_1^2 + u_2^2 + u_3^2 + u_4^2 = 1.$$

Example 4: Polytopes with a Twist and Crystallography

The Klein Bottle and Real Projective Plane depicted as gluings in Figures 7.2, 7.2 and 7.4 are both nonorientable two-dimensional surfaces that cannot be embedded in $\mathbb{R}^3$. They can be displayed as planar gluings, but this should not be confused with planar embeddings.

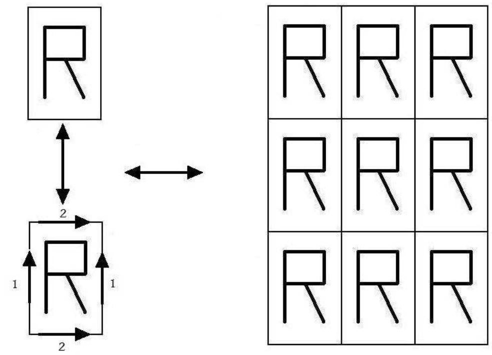

Fig. 7.1. A Pattern on the Torus Transferred to the Euclidean Plane

7.2 What We Care About

If the manifold is defined by an embedding so that $\mathbf{x} = \mathbf{x}(\mathbf{q})$, then we can compute

$$G(\mathbf{q}) = [g_{ij}(\mathbf{q})] \text{ where } g_{ij}(\mathbf{q}) = \frac{\partial \mathbf{x}}{\partial q_i} \cdot \frac{\partial \mathbf{x}}{\partial q_j}.$$

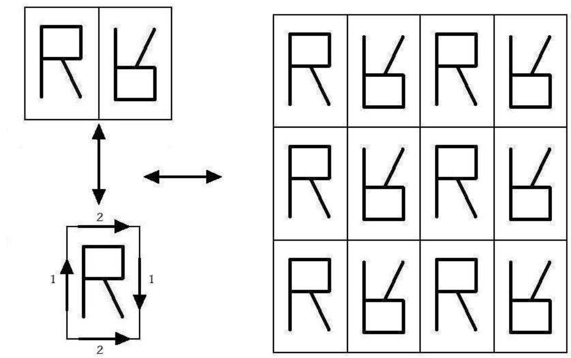

Fig. 7.2. A Pattern on the Klein Bottle Transferred to the Euclidean Plane

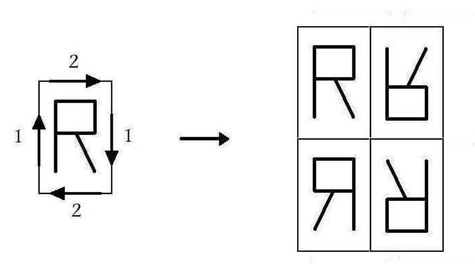

Fig. 7.3. A Pattern on the Real Projective Plane Transferred to the Euclidean Plane

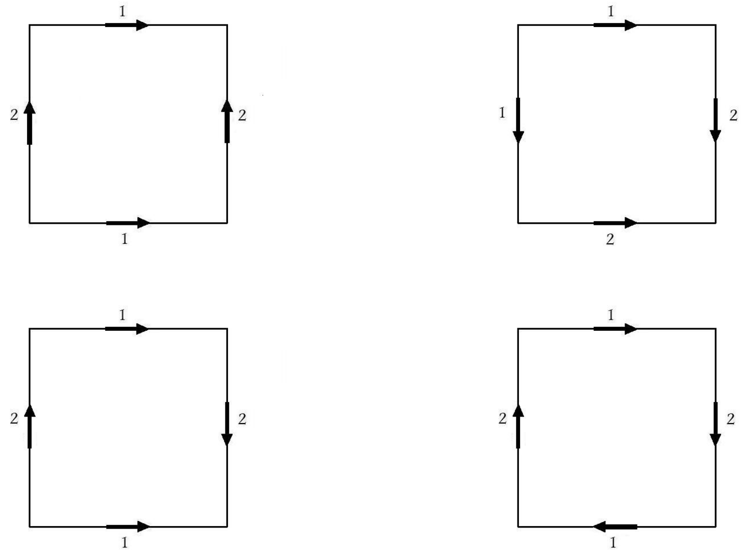

Fig. 7.4. Various Squares with Glued Edges: (upper left) The Torus; (upper right) The Sphere; (lower left) The Klein Bottle; (lower right) The Real Projective Plane

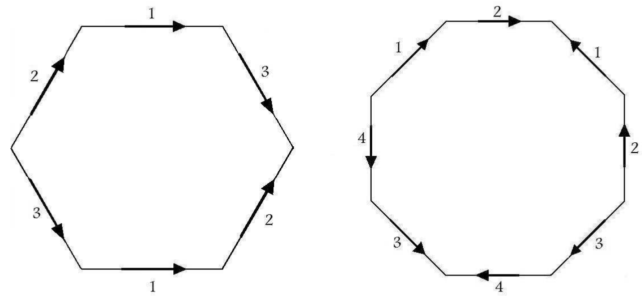

Fig. 7.5. Two-Dimensional Manifolds Represented as Polygons with Glued Edges: (left) The Torus as a Glued Hexagon; (right) The Two-Holed Torus as a Glued Octagon.

References

1. Abraham, R., Marsden, J.E., Ratiu, T., *Manifolds, Tensor Analysis, and Applications* 2^{nd} ed., Applied Mathematical Sciences 75, Springer, New York, 1988.
2. Berger, M., *A Panoramic View of Riemannian Geometry*, Springer, New York, 2003.
3. Bishop, R.L., Goldberg, S.I., *Tensor Analysis on Manifolds*, Dover, 1980. (originally MacMillan, 1968).
4. Boothby, W.M., *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press, New York, 1975.
5. Chern, S.-S., "A Simple intrinsic proof of the Gauss-Bonnet formula for closed Riemannian manifolds," *Ann. Math.* 45, pp. 747-752, 1944.
6. Fenchel, W., "On total curvatures of Riemannian manifolds: I," *J. London Math, Soc.*, 15, pp. 15-22, 1940.
7. Nash, J., "The embedding theorem for Riemannian manifolds," *Ann. of Math.* 63, pp. 20-63 (1956).
8. Thurston, W.P., *Three-Dimensional Geometry and Topology*, (edited by Silvio Levy), Princeton University Press, Princeton, New Jersey, 1997.
9. Whitney, H., "Differentiable Manifolds," *Ann. of Math.*, 37, pp. 645-680 (1936)

10. Yano, K., Bochner, S., *Curvature and Betti Numbers*, Annals of Mathematics Studies 32, Princeton University Press, Princeton, N.J., 1953.

8

Stochastic Processes on Manifolds

The main points to take away from this section are:

- SDEs and Fokker-Planck equations can be formulated for stochastic processes in any coordinate patch of a manifold in a way that is very similar to the case of $\mathbb{R}^n$;
- Stochastic processes on embedded manifolds can also be formulated extrinsically, i.e., using an implicit description of the manifold as a system of constraint equations;
- In some cases Fokker-Planck equations can be solved using separation of variables;
- Practical examples of this theory include Brownian motion on the sphere and the kinematic cart with noise.

8.1 The Fokker-Planck Equation for an Itô SDE on a Manifold

Consider the Itô SDE

$$d\mathbf{q} = \mathbf{h}(\mathbf{q}, t) + H(\mathbf{q}, t)d\mathbf{w} \tag{8.1}$$

where $\mathbf{h}, \mathbf{q} \in \mathbb{R}^d$ and $\mathbf{w} \in \mathbb{R}^m$.

The following Fokker-Planck equation results:

$$\boxed{\frac{\partial f}{\partial t} + |G|^{-\frac{1}{2}} \sum_{i=1}^{d} \frac{\partial}{\partial q_i}\left(|G|^{\frac{1}{2}} h_i f\right) = \frac{1}{2}\,|G|^{-\frac{1}{2}} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial q_i \partial q_j}\left(|G|^{\frac{1}{2}} \sum_{k=1}^{m} H_{ik} H_{kj}^T f\right)} \tag{8.2}$$

In many cases, of interest, the matrices $H_{ik}(\mathbf{q}, t)$ will be the inverse of the Jacobian matrix, and hence in these cases $\sum_k H_{ik}(\mathbf{q}, t) H_{kj}^T(\mathbf{q}, t) = \sum_k ((J_{ik})^{-1}((J_{kj})^{-1})^T) = (g_{ij}(\mathbf{q}))^{-1} = (g^{ij}(\mathbf{q}))$. Therefore, the Fokker-Planck equation on M becomes

$$\frac{\partial f(\mathbf{q}, t)}{\partial t} + |G(\mathbf{q})|^{-\frac{1}{2}} \sum_{i=1}^{d} \frac{\partial}{\partial q_i}\left(|G(\mathbf{q})|^{\frac{1}{2}} h_i(\mathbf{q}, t) f(\mathbf{q}, t)\right) = \tag{8.3}$$

$$\frac{1}{2}\,|G(\mathbf{q})|^{-\frac{1}{2}} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial q_i \partial q_j}\left(|G(\mathbf{q})|^{\frac{1}{2}} (g^{ij}(\mathbf{q})) f(\mathbf{q}, t)\right).$$

8.2 Stratonovich SDEs and Fokker-Planck Equations on Manifolds

The Stratonovich SDE corresponding to (8.1) is:

$$d\mathbf{q} = \mathbf{h}^s(\mathbf{q}, t) + H^s(\mathbf{q}, t) ⓢ\, d\mathbf{w} \tag{8.4}$$

where ⓢ is used to denote the Stratonovich interpretation of an SDE, and

$$h_i^s = h_i - \frac{1}{2} \sum_{j=1}^{m} \sum_{k=1}^{d} H_{kj} \frac{\partial H_{ij}}{\partial q_k} \quad \text{and} \quad H_{ij}^s = H_{ij}. \tag{8.5}$$

If instead the SDE (8.4) is given and the corresponding Itô equation (8.1) is sought, then (8.5) is used in reverse to yield:

$$h_i = h_i^s + \frac{1}{2} \sum_{j=1}^{m} \sum_{k=1}^{d} H_{kj}^s \frac{\partial H_{ij}^s}{\partial q_k} \quad \text{and} \quad H_{ij} = H_{ij}^s. \tag{8.6}$$

Therefore, it follows from substitution of (8.6) into (8.2) that the Stratonovich version of a Fokker-Planck equation describing a process on a manifold is

$$\frac{\partial f}{\partial t} + |G|^{-\frac{1}{2}} \sum_{i=1}^{d} \frac{\partial}{\partial q_i} \left[\left(h_i^s + \frac{1}{2} \sum_{j=1}^{m} \sum_{k=1}^{d} H_{kj}^s \frac{\partial H_{ij}^s}{\partial q_k} \right) f |G|^{\frac{1}{2}} \right] =$$
$$\frac{1}{2} |G|^{-\frac{1}{2}} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial q_i \partial q_j} \left(\sum_{k=1}^{m} H_{ik}^s H_{jk}^s f |G|^{\frac{1}{2}} \right) \tag{8.7}$$

This is important because in many physical modeling problems, the following sort of Stratonovich SDE is presented:

$$J(\mathbf{q}) d\mathbf{q} = \mathbf{b}(t) + B_0 \, \text{Ⓢ} \, d\mathbf{w} \tag{8.8}$$

where B_0 is a constant coupling matrix. For example, if $g(t)$ represents a rotational or full rigid-body motion, then infinitesimal motions are described in terms of a Jacobian matrix as

$$(g^{-1} \dot{g})^{\vee} dt = J(\mathbf{q}) d\mathbf{q}, \tag{8.9}$$

where $\vee$ is an operation that extracts the nonredundant information in $g^{-1}\dot{g}$ and collects it in the form of a column vector. The Jacobian matrix is related to the metric tensor as $G = J^T J$.

And (8.8) is written as

$$d\mathbf{q} = [J(\mathbf{q})]^{-1}\mathbf{b}(t) + [J(\mathbf{q})]^{-1}B_0 \,\text{Ⓢ}\, d\mathbf{w}. \tag{8.10}$$

The interpretation of (8.8) is what allows for the simple expression in (8.9), rather than the extra terms that would be required when using Itô 's rule. Clearly the final result in (8.10) now has a coupling matrix that is not constant, and so even if (8.8) could be interpreted as either Itô or Stratonovich, the result after the Stratonovich interpretation in (8.9) must thereafter be interpreted as a Stratonovich equation.

8.3 Entropy and Fokker-Planck Equations on Manifolds

The entropy of a probability density function on a manifold can be defined as

$$S(f) \doteq -\int_M f(x)\log f(x)dV = -\int_{\mathbf{q}\in D} f(\mathbf{q})\log f(\mathbf{q})|G(\mathbf{q})|^{\frac{1}{2}}d(\mathbf{q}). \tag{8.11}$$

where in the second equality $f(\mathbf{q})$ is shorthand for $f(x(\mathbf{q}))$ and $D \subset \mathbb{R}^n$ is the coordinate domain (assuming that the whole manifold minus a set of measure zero can be parameterized by one such domain).

A natural issue to address is how the entropy $S(f)$ behaves as a function of time when $f(x;t)$ satisfies a Fokker-Planck equation. Differentiating (8.11) with respect to time gives:

$$\frac{dS}{dt} = -\int_M \left\{ \frac{\partial f}{\partial t} \log f + \frac{\partial f}{\partial t} \right\} dV.$$

It is easy to show that

$$\int_M \frac{\partial f}{\partial t}\, dV = \frac{d}{dt} \int_M f\, dV = 0$$

because probability density is preserved by the Fokker-Planck equation.

Taking a coordinate-dependent view, the remaining term is written as

$$\frac{dS}{dt} = -\int_{\mathbf{q} \in D} \frac{\partial f}{\partial t} \log f \, |G|^{\frac{1}{2}} d(\mathbf{q})$$

$$= \int_{\mathbf{q}\in D} \left\{ \sum_{i=1}^{d} \frac{\partial}{\partial q_i} \left(|G|^{\frac{1}{2}} h_i f \right) - \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial q_i \partial q_j} \left(|G|^{\frac{1}{2}} \sum_{k=1}^{m} H_{ik} H_{kj}^T f \right) \right\} \log f \, d(\mathbf{q}).$$

Integrating by parts, and ignoring the boundary terms, gives dS/dt equal to

$$- \int_{\mathbf{q}\in D} \left\{ \sum_{i=1}^{d} \frac{\partial f}{\partial q_i} h_i + \frac{1}{2} \sum_{i,j=1}^{d} \left[-\frac{1}{f} \sum_{k=1}^{m} H_{ik} H_{kj}^T \frac{\partial f}{\partial q_i} \frac{\partial f}{\partial q_j} + \frac{\partial^2 f}{\partial q_i \partial q_j} \sum_{k=1}^{m} H_{ik} H_{kj}^T \right] \right\} |G|^{\frac{1}{2}} \, d(\mathbf{q}). \tag{8.12}$$

If some constraints on the coefficient functions $\{h_i(\mathbf{q}, t)\}$ and $\{H_{ij}(\mathbf{q}, t)\}$ are preserved, then entropy can be shown to be non-decreasing. In particular, in cases when the first and third term vanish, the entropy will be non-decreasing because

$$\frac{1}{f} \sum_{i,j,k} \frac{\partial f}{\partial q_i} H_{ik} H_{kj}^T \frac{\partial f}{\partial q_j} \geq 0.$$

8.4 Examples

8.4.1 Stochastic Motion on the Unit Circle

Consider the SDE

$$
\begin{aligned}
dx_1 &= -\frac{1}{2}x_1 dt - x_2 dw \\
dx_2 &= -\frac{1}{2}x_2 dt + x_1 dw.
\end{aligned}
\tag{8.13}
$$

Let

$$x_1 = x_1(r,\theta) = r\cos\theta \quad \text{and} \quad x_2 = x_2(r,\theta) = r\sin\theta.$$

In this problem the parametric coordinates $\mathbf{q} = [q_1, q_2]^T$ are $q_1 = r$ and $q_2 = \theta$. And so,

$$\begin{pmatrix} \frac{\partial x_1}{\partial q_1} \\ \\ \frac{\partial x_1}{\partial q_2} \end{pmatrix} = \begin{pmatrix} \cos\theta \\ -r\sin\theta \end{pmatrix} \text{ and } \begin{pmatrix} \frac{\partial x_2}{\partial q_1} \\ \\ \frac{\partial x_2}{\partial q_2} \end{pmatrix} = \begin{pmatrix} \sin\theta \\ r\cos\theta \end{pmatrix}.$$

Likewise,

$$\begin{pmatrix} \frac{\partial^2 x_1}{\partial q_1 \partial q_1} & \frac{\partial^2 x_1}{\partial q_1 \partial q_2} \\ \\ \frac{\partial^2 x_1}{\partial q_2 \partial q_1} & \frac{\partial^2 x_1}{\partial q_2 \partial q_2} \end{pmatrix} = \begin{pmatrix} 0 & -\sin\theta \\ -\sin\theta & -r\cos\theta \end{pmatrix}$$

and

$$\begin{pmatrix} \frac{\partial^2 x_2}{\partial q_1 \partial q_1} & \frac{\partial^2 x_2}{\partial q_1 \partial q_2} \\ \\ \frac{\partial^2 x_2}{\partial q_2 \partial q_1} & \frac{\partial^2 x_2}{\partial q_2 \partial q_2} \end{pmatrix} = \begin{pmatrix} 0 & \cos\theta \\ \cos\theta & -r\sin\theta \end{pmatrix}.$$

Substitution into Itô 's rule, which holds regardless of the SDE in (8.13), gives

$$dx_1 = \cos\theta dr - r\sin\theta d\theta - \sin\theta dr d\theta - \frac{1}{2} r\cos\theta (d\theta)^2$$
$$dx_2 = \sin\theta dr + r\cos\theta d\theta + \cos\theta dr d\theta - \frac{1}{2} r\sin\theta (d\theta)^2.$$

Now, assume that an SDE in these new variables exists and can be written as

$$dr = a_1 dt + b_1 dw$$
$$d\theta = a_2 dt + b_2 dw$$

where $a_i = a_i(r,\theta)$ and $b_i = b_i(r,\theta)$.

Substitution of the above expressions, and using the stochastic calculus rules $dw^2 = dt$ and $dt^2 = dtdw = 0$ gives

$$dx_1 = \left[a_1 \cos\theta - a_2 r \sin\theta - b_1 b_2 \sin\theta - \frac{1}{2} b_2^2 r \cos\theta\right] dt + (b_1 \cos\theta - b_2 r \sin\theta) dw$$

and

$$dx_2 = \left[a_1 \sin\theta + a_2 r \cos\theta + b_1 b_2 \cos\theta - \frac{1}{2} b_2^2 r \sin\theta\right] dt + (b_1 \sin\theta + b_2 r \cos\theta) dw$$

Then substituting these into (8.13), forces $a_1 = a_2 = b_1 = 0$ and $b_2 = 1$, resulting in the SDE

$$d\theta = dw$$

This shows that (8.13) are stochastic differential equations for a process that evolves only in θ, with r remaining constant. In other words, this is a kind of stochastic motion on the circle.

8.4.2 The Unit Sphere in $\mathbb{R}^3$

Let the position of any point on the unit sphere, S^2, be parameterized as

$$\mathbf{u}(\phi,\theta) \doteq \begin{pmatrix} \cos\phi\sin\theta \\ \sin\phi\sin\theta \\ \cos\theta \end{pmatrix} \tag{8.14}$$

It follows from the fact that $\mathbf{u}\cdot\mathbf{u} = 1$ that taking the derivative of both sides yields $\mathbf{u}\cdot d\mathbf{u} = 0$ where

$$d\mathbf{u} = \frac{\partial \mathbf{u}}{\partial \theta}d\theta + \frac{\partial \mathbf{u}}{\partial \phi}d\phi. \tag{8.15}$$

And since $d\theta$ and $d\phi$ are independent,

$$\mathbf{u}\cdot\frac{\partial \mathbf{u}}{\partial \theta} = \mathbf{u}\cdot\frac{\partial \mathbf{u}}{\partial \phi} = 0. \tag{8.16}$$

Of course, this can be verified by direct calculation. Furthermore, since

$$\frac{\partial \mathbf{u}}{\partial \theta} \cdot \frac{\partial \mathbf{u}}{\partial \theta} = 1 \quad \text{and} \quad \frac{\partial \mathbf{u}}{\partial \phi} \cdot \frac{\partial \mathbf{u}}{\partial \phi} = \sin^2 \theta,$$

the vectors

$$\mathbf{v}_1 \doteq \frac{\partial \mathbf{u}}{\partial \theta} \quad \text{and} \quad \mathbf{v}_2 \doteq \frac{1}{\sin \theta} \frac{\partial \mathbf{u}}{\partial \phi}$$

form an orthonormal basis for the tangent plane to the sphere at the point $\mathbf{u}(\phi, \theta)$, with $\mathbf{v}_1 \times \mathbf{v}_2 = \mathbf{u}$.

Indeed, any version of this coordinate system rotated around the vector $\mathbf{u}$ of the form

$$\begin{aligned} \mathbf{v}'_1 &= \mathbf{v}_1 \cos \alpha - \mathbf{v}_2 \sin \alpha \\ \mathbf{v}'_2 &= \mathbf{v}_1 \sin \alpha + \mathbf{v}_2 \cos \alpha \end{aligned} \tag{8.17}$$

will also form an orthonormal basis for this tangent plane, where $\alpha = \alpha(\phi, \theta)$ is an arbitrary smooth function. This will be relevant later, but for now the focus will be the basis $\{\mathbf{v}_1, \mathbf{v}_2\}$.

Consider the Stratonovich equation

$$d\mathbf{u} = \mathbf{v}_1 \circledS\, dw_1 + \mathbf{v}_2 \circledS\, dw_2,$$

which would seem like a reasonable definition of Brownian motion on the sphere. Taking the dot product of both sides with respect to $\mathbf{v}_1$ and $\mathbf{v}_2$, and observing (8.15), the resulting two scalar equations can be written as

$$\begin{pmatrix} d\theta \\ d\phi \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1/\sin\theta \end{pmatrix} \circledS \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix} \tag{8.18}$$

The corresponding Fokker-Planck equation is

$$\frac{\partial f}{\partial t} = \frac{1}{2}\left[\frac{\partial^2 f}{\partial \theta^2} + 2\cot\theta \frac{\partial f}{\partial \theta} - f + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial \phi^2}\right],$$

which is clearly not the heat equation.

A Stratonovich SDE that does correspond to the heat equation,

$$\frac{\partial f}{\partial t} = \frac{1}{2}\left[\frac{\partial^2 f}{\partial \theta^2} + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial \phi^2}\right],$$

is

$$\begin{pmatrix} d\theta \\ d\phi \end{pmatrix} = \frac{1}{2}\cot\theta\,\mathbf{e}_i + \begin{pmatrix} 1 & 0 \\ 0 & 1/\sin\theta \end{pmatrix} ⓢ \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix} \tag{8.19}$$

The Itô SDE corresponding to (8.19) is of exactly the same form.

8.4.3 The SDE and Fokker-Planck Equation for the Kinematic Cart

Each matrix $g(x, y, \theta)$ of the form in (2.1) for $\theta \in [0, 2\pi)$ and $x, y \in \mathbb{R}$ can be identified with a point on the manifold $M = \mathbb{R}^2 \times S^1$. In addition, the product of such matrices produces a matrix of the same kind. Explicitly, if

$$g_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i & x_i \\ \sin\theta_i & \cos\theta_i & y_i \\ 0 & 0 & 1 \end{pmatrix}$$

for $i = 1, 2$, then

$$g_1 g_2 = \begin{pmatrix} \cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2) & x_1 + x_2\cos\theta_1 - y_2\sin\theta_1 \\ \sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2) & y_1 + x_2\sin\theta_2 + y_2\cos\theta_2 \\ 0 & 0 & 1 \end{pmatrix}.$$

This product is an analytic function from $M \times M \to M$, which makes M (together with the operation of matrix multiplication) a Lie group (called the Special Euclidean group, or motion group, of the plane, and denoted as $SE(2)$). Lie groups are not addressed formally in this volume, and M is treated simply as a manifold. The added structure provided by Lie groups makes the formulation of problems *easier* rather than harder. Lie groups are addressed in detail in Volume 2. For now, the manifold structure of M is sufficient to formulate the problem of the stochastic cart.

Consider the following variant on the SDE stated in (2.4) that describes the scenario in Figure 2.1:

$$\begin{pmatrix} dx \\ dy \\ d\theta \end{pmatrix} = \begin{pmatrix} r\omega\cos\theta \\ r\omega\sin\theta \\ 0 \end{pmatrix} dt + \sqrt{D} \begin{pmatrix} \frac{r}{2}\cos\theta & \frac{r}{2}\cos\theta \\ \frac{r}{2}\sin\theta & \frac{r}{2}\sin\theta \\ \frac{r}{L} & -\frac{r}{L} \end{pmatrix} \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix} \tag{8.20}$$

Using the general formulation in (8.2), the Fokker-Planck equation becomes:

$$\begin{aligned}\frac{\partial f}{\partial t} =& - r\omega\cos\theta\frac{\partial f}{\partial x} - r\omega\sin\theta\frac{\partial f}{\partial y}\\ &+ \frac{D}{2}\left(\frac{r^2}{2}\cos^2\theta\frac{\partial^2 f}{\partial x^2} + \frac{r^2}{2}\sin 2\theta\frac{\partial^2 f}{\partial x\partial y} + \frac{r^2}{2}\sin^2\theta\frac{\partial^2 f}{\partial y^2} + \frac{2r^2}{L^2}\frac{\partial^2 f}{\partial\theta^2}\right).\end{aligned}$$

References

1. Atiyah, M., Bott, R., Patodi, V.K., "On the heat equation and the Index theorem," *Invent. Math.* Vol. 19, pp. 279-330, 1973.
2. Brockett, R.W., "Lie algebras and Lie groups in control theory," in *Geometric Methods in System Theory*, (D.Q. Mayne and R.W. Brockett, eds.), Reidel Publishing Company, Dordrecht-Holland, 1973.
3. Brockett, R.W., "Notes on Stochastic Processes on Manifolds," in *Systems and Control in the Twenty-First Century* (C.I. Byrnes et al eds.), Birkhäuser, Boston, 1997.
4. Elworthy, K.D., *Stochastic Differential Equations on Manifolds*, Cambridge University Press, 1982.
5. Emery, M., *Stochastic Calculus in Manifolds*, Springer-Verlag, Berlin, 1989.
6. Hsu, E.P., *Stochastic Analysis on Manifolds*, Graduate Studies in Mathematics, Vol. 38, American Mathematical Society, Providence, RI, 2002.
7. Itô, K., McKean, H.P.Jr., *Diffusion Processes and their Sample Paths*, Springer, 1996.
8. Kunita, H., *Stochastic flows and stochastic differential equations*, Cambridge University Press, 1997.

9. Lewis, J., “Brownian motion on a submanifold of Euclidean Space,” *Bull. London Math. Soc.* 18, pp. 616-620, 1986.
10. McKean, H.,Jr., Singer, I.M., “Curvature and the eigenvalues of the Laplacian,” *J. Diff. Geom.* Vol. 1, pp. 43-69, 1967.
11. McKean, H.P.,Jr., “Brownian motions on the 3-dimensional rotation group,” *Memoirs of the College of Science, University of Kyoto, Series A*, Vol. 33, No. 1, pp. 2538, 1960.
12. McLachlan, N. W., “Theory and application of Mathieu functions,” Oxford, Clarendon Press, 1951.
13. Øksendal, B., *Stochastic Differential Equations, An Introduction with Applications*, 5^{th} ed., Springer, Berlin, 1998.
14. Orsingher, E., “Stochastic Motions on the 3-Sphere Governed by Wave and Heat Equations,” *J. Appl. Prob.*, Vol. 24, pp. 315-327, 1987.
15. Perrin, P.F., ”Mouvement Brownien D’un Ellipsoide (I). Dispersion Diélectrique Pour des Molécules Ellipsoidales,” *Le Journal de Physique et Le Radium*, Vol. 7, No. 10, pp. 497-511, Oct. 1934.
16. Perrin, P.F., ”Mouvement Brownien D’un Ellipsoide (II). Rotation Libre et Dépolarisation des Fluorescences. Translation et Diffusion de Molécules Ellipsoidales,” *Le Journal de Physique et Le Radium*, Vol. 7, No. 1, pp. 1-11, Jan. 1936.
17. Perrin, P.F., ” Étude Mathématique du Mouvement Brownien de Rotation,” *Annales Scientifiques de L’ École Normale Supérieure*, Vol. 45, pp. 1-51, 1928.
18. Pinsky, M., “Isotropic transport process on a Riemannian manifold,” *TAMS* 218, 353-360 (1976)
19. Stroock, D.W., *An Introduction to the Analysis of Paths on a Riemannian Manifold*, Mathematical Surveys and Monographs, Vol. 74, American Mathematical Society, Providence, RI, 2000.
20. Yosida, K., “Integration of Fokker-Planck’s equation in a compact Riemannian space,” *Arkiv für Matematik*, Band 1, Nr. 9, pp. 71-75, 1949.

9

Rigid-Body Motions: A Lie-Theoretic View

Homogenous transformation matrices are closed under multiplication and inversion, and have an identity element. This makes them a group. They also form a manifold. They are an example of a Lie group, which means that they are "easier" to handle than other manifolds. We can integrate, differentiate, and convolve functions in a natural way.

9.1 Integration over Rigid-Body Motions in the plane

$$g(a_1, a_2, \theta) = \begin{pmatrix} \cos\theta & -\sin\theta & a_1 \\ \sin\theta & \cos\theta & a_2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Such matrices form a group under the operation of matrix multiplication, and the bi-invariant integration measure on this group is defined using the following procedure. First, observe that the basis elements of the Lie algebra $\mathcal{G} = se(2)$ are:

$$X_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \implies \exp tX_1 = \begin{pmatrix} 1 & 0 & t \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$X_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \implies \exp tX_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}$$

$$X_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \implies \exp tX_3 = \begin{pmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is easy to see that

$$g(a_1, a_2, \theta) = \exp(a_1 X_1) \exp(a_2 X_2) \exp(\theta X_3)$$

where $\exp(\cdot)$ is the matrix exponential. Furthermore, it can be shown by direct calculation that

$$g^{-1}\partial g = v_1^r X_1 + v_2^r X_2 + \omega X_3$$

and

$$(\partial g)g^{-1} = v_1^l X_1 + v_2^l X_2 + \omega X_3$$

where ∂g denotes any of the partials $\partial g/\partial a_i$ or $\partial g/\partial \theta$. This means that we can extract the relevant information from the above expressions into a vector as:

$$(g^{-1}\partial g)^\vee = \begin{pmatrix} v_1^r \\ v_2^r \\ \omega \end{pmatrix} \quad \text{or} \quad ((\partial g)g^{-1})^\vee = \begin{pmatrix} v_1^l \\ v_2^l \\ \omega \end{pmatrix}.$$

The bi-invariant volume element is then obtained as

$$dg = |\det J_r| da_1 da_2 d\theta = |\det J_l| da_1 da_2 d\theta$$

where

$$J_r(g) = \left[\left(g^{-1} \frac{\partial g}{\partial a_1} \right)^{\vee}, \left(g^{-1} \frac{\partial g}{\partial a_2} \right)^{\vee}, \left(g^{-1} \frac{\partial g}{\partial \theta} \right)^{\vee} \right]$$

and

$$J_l(g) = \left[\left(\frac{\partial g}{\partial a_1} g^{-1} \right)^{\vee}, \left(\frac{\partial g}{\partial a_2} g^{-1} \right)^{\vee}, \left(\frac{\partial g}{\partial \theta} g^{-1} \right)^{\vee} \right].$$

The fact that $J_r(g_0 g) = J_r(g)$ and $J_l(g g_0) = J_l(g)$ is obvious from their definition. The bi-invariance therefore follows follows from the fact that $|\mathrm{det} J_r| = |\mathrm{det} J_l|$, which in this particular case is equal to the number 1.

9.2 Integration over Rigid-Body Motions in Space

For the spatial case, we see the invariance of the volume element as follows. Right invariance follows from the fact that for any constant homogeneous transform

$$h_0 = \begin{pmatrix} R_0 & \mathbf{b}_0 \\ \mathbf{0}^T & 1 \end{pmatrix},$$

$$J_l(gh_0) = \left[(\frac{\partial g}{\partial q_1} h_0 (gh_0)^{-1})^\vee \cdots (\frac{\partial g}{\partial q_6} h_0 (gh_0)^{-1})^\vee \right].$$

Since $(gh_0)^{-1} = h_0^{-1} g^{-1}$, and $h_0 h_0^{-1} = 1$, we have that $J_l(gh_0) = J_l(g)$.

The left invariance follows from the fact that

$$J_l(h_0 g) = \left[(h_0 \frac{\partial g}{\partial q_1} g^{-1} h_0^{-1})^\vee \cdots (h_0 \frac{\partial g}{\partial q_6} g^{-1} h_0^{-1})^\vee \right],$$

where

$$(h_0 \frac{\partial g}{\partial q_i} g^{-1} h_0^{-1})^\vee = [Ad(h_0)] (\frac{\partial g}{\partial q_i} g^{-1})^\vee.$$

Therefore,

$$J_l(h_0 g) = [Ad(h_0)] \mathcal{J}_l(g).$$

But since $\det(Ad(h_0)) = 1$,

$$\det(J_l(h_0 g)) = \det(J_l(g)).$$

Explicitly, the volume element in the case of $SE(3)$ is found when the rotation matrix is parametrized using ZXZ or ZYZ Euler-Angles (α, β, γ) as

$$d(g(x_1, x_2, x_3, \alpha, \beta, \gamma) = \frac{1}{8\pi^2} \sin\beta d\alpha d\beta d\gamma dx_1 dx_2 dx_3,$$

which is the product of the volume elements for $\mathbb{R}^3$ ($d\mathbf{x} = dx_1dx_2dx_3$), and for $SO(3)$ ($dR = \sin\beta d\alpha d\beta d\gamma$). Since $\beta \in [0, \pi]$, this is positive, except at the two points $\beta = 0$ and $\beta = \pi$, which constitute a set of zero measure and therefore does not contribute to the integral of singularity free functions.

$$\int_G f(g^{-1})dg = \int_G f(h \circ g)dg = \int_G f(g \circ h)dg = \int_G f(g)dg.$$

10

Parts Entropy and The Principal Kinematic Formula

Sanderson quantified this with the concept of "parts entropy," which is a statistical measure of the ensemble of all possible positions and orientations of a single part confined to move in a finite domain. Here it is shown that the rapid computation of excluded-volume effects using the "Principal Kinematic Formula" from the field of Integral Geometry is illustrated as a way to potentially avoid the massive computations associated with brute-force calculation of parts entropy when many interacting parts are present.

10.1 Problem Formulation

10.1.1 A Continuous Version of Sanderson's Parts Entropy

Information-theoretic entropy has been used by Sanderson to characterize parts for use in assembly operations [10]

$$S_f(t) = -\int_G f(g;t)\log f(g;t)dg \tag{10.1}$$

10.2 Multiple Parts And The Principal Kinematic Formula

Suppose that we have two convex bodies, H and K, viewed as subsets of $\mathbb{R}^n$.

The *indicator function* on any measurable body, B, (not necessarily convex and perhaps not even connected) is defined by:

$$i(B) = \begin{cases} 1 & \text{if } B \neq \emptyset \\ 0 & \text{for } B = \emptyset \end{cases}$$

Note that $i(g \cdot B) = i(B)$.

Let H be stationary, and let K be mobile. Let us denote $g \in G$ (where $G = SE(n)$ is the group of rigid-body motions for bodies contained in $\mathbb{R}^n$). Then, by definition,

$$gK = \{g \cdot \mathbf{x} | \mathbf{x} \in K\},$$

where if $g = (A, \mathbf{a})$ is the rigid-body motion with rotational part $A \in SO(n)$ and translational part $\mathbf{a} \in \mathbb{R}^n$, then the action of $G = SE(n)$ on $\mathbb{R}^n$ is $g \cdot \mathbf{x} = A\mathbf{x} + \mathbf{a}$.

The intersection of two convex bodies is a convex body [7]. $H \cap gK$ will be a convex body, and $f_{H,K}(g) = i(H \cap gK)$ will be a compactly supported function on G that takes the value of 1 when H and the moved version of K (denoted as gK) intersect, and it will be zero otherwise. The function $f_{H,K}(g)$ has some interesting properties. Namely, if we shift the whole picture by an amount g_0, then this does not change the value of

$f_{H,K}(g)$. In other words,

$$i(g_0(H \cap gK)) = i((g_0 H) \cap (g_0 g K)).$$

This means that if we choose $g_0 = g^{-1}$, then

$$f_{H,K}(g) = i((g^{-1}H) \cap K) = i(K \cap g^{-1}H) = f_{K,H}(g^{-1}).$$

In the special case when $H = K$ (i.e., they are two copies of the same body) then $f_{H,H}(g) = f_{H,H}(g^{-1})$, which is called a symmetric function.

More generally, "counting up" all values of g for which an intersection occurs is then equivalent to computing the integral

$$J = \int_G i(H \cap gK) dg. \tag{10.2}$$

10.2.1 The Planar Case

A closed arc-length-parameterized curve of length L in the plane can be described (up to rigid-body motion) using the equation:

$$\mathbf{x}(s) = \begin{pmatrix} \int_0^s \cos\theta(\sigma)d\sigma \\ \\ \int_0^s \sin\theta(\sigma)d\sigma \end{pmatrix}$$

where

$$\theta(s) = \int_0^s \kappa(\sigma)d\sigma$$

is the counterclockwise-measured angle that the tangent to the curve makes with respect to the x-axis and $s \in [0, L]$.

For a simple, convex, closed curve,

$$\chi = \theta(L)/2\pi,$$

is equal to one.

In the planar case, we can write (10.2) explicitly as

$$J = \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} i(H \cap g(a_1, a_2, \theta)K)da_1 da_2 d\theta \tag{10.3}$$

Theorem 1 (Blaschke, [2]): Given planar convex bodies H and K, then (10.3) evaluates as:

$$\int_{SE(2)} i(H \cap gK)dg = 2\pi[A(H) + A(K)] + L(H)L(K). \tag{10.4}$$

where $A(\cdot)$ is the area and $L(\cdot)$ is the perimeter of the body. Proof: See [2, 12].

In the nonconvex case, we can bound the integral of interest from below and above by inscribing and circumscribing convex bodies inside and outside of H and K. Then computing (10.4) with the convex inscribed/circumscribed bodies will give lower and upper bounds on (10.4) for nonconvex H and K.

10.2.2 The Spatial Case

It follows that if B has a continuous piecewise differentiable surface, ∂B, that we can compute

$$\int_{\partial B} dS = F(B)$$

(the total surface area). Furthermore, if κ denotes the Gaussian curvature at each point on the surface, we can compute (via the Gauss-Bonnet Theorem):

$$\int_{\partial B} \kappa dS = 2\pi\chi(B)$$

where $\chi(B)$ is the Euler characteristic. In the case of a convex spatial body, which necessarily is bounded by a surface of genus zero, $\chi(H) = 2 \cdot i(B)$.

In differential geometry a second kind of curvature is defined at every point on a surface. This is the *mean curvature*, m. The total mean sectional curvature is defined as

$$M(B) = \int_{\partial B} mdS.$$

In contrast to the indicator function, if we define

$$v_B(\mathbf{x}) = \begin{cases} 1 \text{ for } \mathbf{x} \in B \\ 0 \text{ for } \mathbf{x} \notin B \end{cases}$$

then,

$$\int_{\mathbb{R}^n} v_B(\mathbf{x})d\mathbf{x} = \int_B d\mathbf{x} = V(B)$$

(the volume of B).

Theorem 2 (Blaschke, [2]): Given convex bodies H and K in $\mathbb{R}^3$, then

$$\int_{SE(3)} i(H \cap gK)dg = 8\pi^2[V(H) + V(K)] \qquad (10.5)$$
$$+ 2\pi[A(H)M(K) + A(K)M(K)]$$

where $A(\cdot)$ and $M(\cdot)$ are respectively the area and integral of mean curvature of the surface enclosing the body, and $V(\cdot)$ is the volume of the body.
Proof: See [2, 12].

10.3 Examples

If part 1 is fixed in space, the second has a volume of possible motions in $SE(n)$ given by

$$V = Vol(B^n(R)) \cdot Vol(SO(n))$$

where $Vol(B^n(R))$ is the volume of the ball defined by the interior of a sphere of radius R in n-dimensional space (which is πR^2 in $\mathbb{R}^2$ and $4\pi R^3/3$ in $\mathbb{R}^3$) and $Vol(SO(n))$ is the volume of the rotation group in n-dimensional space (which is 2π for $SO(2)$ and $8\pi^2$ for $SO(3)$) [5].

Therefore the the positional and orientational distribution of part # 2 computed in the absence of part # 1 would be

$$f(g) = \frac{1}{V}$$

for $g = (A, \mathbf{a}) \in SE(n)$ with $\|\mathbf{a}\| < R$, and $f(g) = 0$ otherwise.

The entropy of a single isolated part under these conditions is then

$$S_f = \log V.$$

In contrast, the total volume in $SE(n)$ that is available for part #2 to move if part # 1 is fixed in the environment, thereby limiting the range of possible motions of part #2, will be

$$V' = V - \int_{SE(n)} i(H \cap gK)dg$$

as long as R is larger than half of the sum of the maximal dimensions of the two parts. Otherwise, the effects of part #1 on limiting the motion may be even greater. With that caveat,

$$S_{f'} = \log V'. \tag{10.6}$$

10.3.1 Example 1: The Planar Case: Circular Disks in Planar Motion

Let part # 1 be a circular disk of radius r_1 fixed at the origin, and let part # 2 be a circular disk of radius r_2. If part # 2 were completely free to rotate, and free to translate such that its center stays anywhere in the large circle defined by radius R, then the part entropy would be

$$S = \log(2\pi^2 R^2).$$

In contrast, if all conditions are the same except that the constraint of no interpenetration is imposed, then

$$S' = \log(2\pi^2[R^2 - (r_1 + r_2)^2]),$$

which just removes the disallowed translations defined by the distance of the center of part # 2 from the origin in the range $[0, r_1 + r_2]$. This is a simple example that does not require any numerical computation of integrals of motion, or even the evaluation of the principal kinematic formula. But it serves to verify the methodology, since in this case

$$2\pi[A(H) + A(K)] + L(H)L(K) =$$
$$2\pi[\pi r_1^2 + \pi r_2^2] + (2\pi r_1)(2\pi r_2) =$$
$$2\pi^2(r_1 + r_2)^2,$$

which means that the adjustment to the computation of parts entropy from the principal kinematic formula (10.4) will be exactly the same as expected.

10.3.2 Example 2: Ellipsoids of Revolution in Spatial Motion

Consider an ellipsoid of revolution with dimensions of length a, a and b. The volume can be computed as:

$$V = \frac{4}{3}\pi a^2 b.$$

The values of surface area, F, and mean sectional curvature, M, for prolate and oblate ellipsoids have been reported in [8], along with a variety of other solids of revolution. In particular, if $a = R$ and $b = \lambda r$ with $0 < \lambda < 1$, then

$$F = 2\pi r^2 \left[1 + \frac{\lambda^2}{\sqrt{1-\lambda^2}} \log \left(\frac{1+\sqrt{1-\lambda^2}}{\lambda} \right) \right]$$

and

$$M = 2\pi r \left[\lambda + \frac{\arccos\lambda}{\sqrt{1-\lambda^2}} \right].$$

In contrast, when $\lambda > 1$,

$$F = 2\pi r^2 \left[1 + \frac{\lambda^2 \arccos(1/\lambda)}{\sqrt{\lambda^2 - 1}} \right]$$

and

$$M = 2\pi r\left[\lambda + \frac{\log(\lambda + \sqrt{\lambda^2 - 1})}{\sqrt{\lambda^2 - 1}}\right].$$

In the case of a sphere ($\lambda = 1$),

$$V = \frac{4}{3}\pi r^3; \quad F = 4\pi r^2; \quad M = 4\pi r.$$

As a specific example to demonstrate this, consider the case of two spherical parts: part # 1 has radius r_1 and part # 2 has radius r_2. If part # 1 is fixed at the origin, and part #2 is free to move as long as its center does not go further than a distance R from the origin, then the volume of allowable motion of part #2 in $SE(3)$ will be

$$(8\pi^2)(4\pi/3)[R^3 - (r_1 + r_2)^3].$$

But (10.6) gives the amount of excluded volume in $SE(3)$ to be

$$8\pi^2[V(H) + V(K)] + 2\pi[A(H)M(K) + A(K)M(K)] =$$

$$8\pi^2[4\pi r_1^3/3 + 4\pi r_2^3/3] + 2\pi[(4\pi r_1^2)(4\pi r_2) + (4\pi r_2^2)(4\pi r_1)] =$$

$$(32\pi^3/3)(r_1^3 + r_2^3 + 3r_1^2 r_2 + 3r_1 r_2^2) =$$
$$(32\pi^3/3)(r_1 + r_2)^3.$$

And this too matches the direct analytical calculation for this simple example.

10.4 Extensions And Limitations

The principal kinematic formula has been used to compute integrals of the form

$$J = \int_G i(H \cap gK)dg.$$

that arise when calculating the entropy of convex parts that can be placed uniformly at random. In integral geometry, generalized integrals of the form

$$J_1 = \int_G \mu(H \cap gK)dg$$

can be computed in closed form for bodies that are not convex, where μ can be the volume, Euler characteristic, surface area, mean curvature, or

Gaussian curvature. This is not directly applicable to the current discussion, though it does open up intriguing possibilities.

A quantity that is not directly addressed in integral geometry is

$$J_2 = \int_G i(H \cap gK)\rho(g)dg$$

where $\rho(g)$ is a probability density function on G. This would be something that is useful for parts entropy calculations. The author is currently investigating this.

References

1. A.J. Baddeley, E.B.V. Jensen, *Stereology for Statisticians. Monographs on Statistics and Applied Probability 103.* Chapman & Hall/CRC, Boca Raton, 2005
2. Blaschke, Wilhelm Vorlesungen über Integralgeometrie. Berlin, Deutscher Verlag der Wissenschaften, 1955.
3. Boothroyd G.,AH Redford, *Mechanized Assembly: Fundamentals of parts feeding, orientation, and mechanized assembly*, McGraw-Hill, London, 1968.
4. Boothroyd G., *Assembly Automation and Product Design, Second Edition*, CRC Press, Boca Raton, FL, 2005
5. Chirikjian, G.S., Kyatkin, A.B., *Engineering Applications of Noncommutative Harmonic Analysis* CRC Press, Boca Raton, FL, 2001.
6. Erdmann M.A., Mason M.T., "An Exploration of Sensorless Manipulation", I.E.E.E. *Journal of Robotics and Automation*, Vol. 4, No. 4, pp. 369-379, August 1988.
7. Hadwiger, H., *Vorlesungen über Inhalt, Oberfläche und Isoperimetrie.* Springer Verlag, Berlin, 1957.
8. Hadwiger, H., *Altes und Neues über Konvexe Körper,* Birkhäuser Verlag, Basel, 1955.
9. Reza, F.M., *An Introduction to Information Theory*. Dover, New York, 1994.

10. A.C. Sanderson, “Parts Entropy Methods For Robotic Assembly System Design” , *Proceedings of the 1984 IEEE International Conference on Robotics and Automation (ICRA ’84)*, Vol. 1, pp. 600 - 608, March 1984
11. Shiing-shen Chern, “On the Kinematic Formula in the Euclidean Space of N Dimensions,” *American Journal of Mathematics*, Vol. 74, No. 1 (Jan., 1952), pp. 227-236
12. Santaló, L., *Integral Geometry and Geometric Probability*, Cambridge University Press, 2004 (originally published in 1976 by Addison-Wesley)