Optimal Estimation Methods

(Lecture 22 – Introduction to Stochastic Processes)

Dr. John L. Crassidis

University at Buffalo – State University of New York
Department of Mechanical & Aerospace Engineering
Amherst, NY 14260-4400
johnc@buffalo.edu
http://www.buffalo.edu/~johnc

University at Buffalo The State University of New York

Stochastic Processes (i)

- A *stochastic process* is simply a collection of random vectors defined on the same probability space
- Some basic definitions are now given
  - Let $\{\mathbf{x}(t_k)\}$ denote a sample function, which is a particular sequence of values taken as a result of an experiment
    - The variable $\mathbf{x}(t_k)$ is the random variable obtained at time t_k
  - Consider the pdf $p(\mathbf{x}(t_1), \mathbf{x}(t_2), \ldots, \mathbf{x}(t_m))$
    - The mean $\boldsymbol{\mu}(t_k)$ is denoted by $E\{\mathbf{x}(t_k)\}$
    - The *autocorrelation* is the set of quantities $E\{\mathbf{x}(t_i)\mathbf{x}^T(t_j)\}$
    - The covariance is defined by $E\{[\mathbf{x}(t_i) - \boldsymbol{\mu}(t_i)][\mathbf{x}(t_j) - \boldsymbol{\mu}(t_j)]^T\}$ for all t_i and t_j
    - Two processes, $\{\mathbf{x}(t_k)\}$ and $\{\mathbf{y}(t_k)\}$, are uncorrelated if $E\{\mathbf{x}(t_i)\,\mathbf{y}^T(t_j)\} = E\{\mathbf{x}(t_i)\}E\{\mathbf{y}^T(t_j)\}$ for all t_i and t_j
    - They are orthogonal if $E\{\mathbf{x}(t_i)\,\mathbf{y}^T(t_j)\} = 0$ for all t_i and t_j

Stochastic Processes (ii)

- A process is said to be *stationary* if its random variable statistics do not vary in time, i.e., for arbitrary N we have

$$p\left(\mathbf{x}(t_1), \mathbf{x}(t_2), \ldots, \mathbf{x}(t_m)\right) = p\left(\mathbf{x}(t_{1+N}), \mathbf{x}(t_{2+N}), \ldots, \mathbf{x}(t_{m+N})\right)$$

  - A process is *asymptotically stationary* if the following exists

$$\lim_{N \to \infty} p\left(\mathbf{x}(t_{1+N}) \ldots, \mathbf{x}(t_{m+N})\right)$$

  - A *wide-sense stationary process* exists if its first and second moments are invariant under time translation
  - An *ergodic process* is a stationary process where the time averages can be replaced by an expectation
    - For Gaussian processes ergodicity is simply given by the following sufficient condition

$$\sum_{t_k=-\infty}^{+\infty} ||R(t_k)|| < \infty$$

Stochastic Processes (iii)

- A *Markov process* is defined by the past having no influence on the present, which can be mathematically stated as

$$p\left(\mathbf{x}(t_1)|\mathbf{x}(t_2),\ \ldots,\ \mathbf{x}(t_m)\right) = p\left(\mathbf{x}(t_1)|\mathbf{x}(t_2)\right) \qquad (1)$$

  - A second order Markov process is one where the most recent two pieces of information are all that affect the present
    - Right-hand side of (1) is replaced by $p\left(\mathbf{x}(t_1)|\mathbf{x}(t_2), \mathbf{x}(t_3)\right)$
- A sequence of random variables $\mathbf{x}_k$ converges to a random variable $\mathbf{x}$ *with probability one* if every outcome is an event whose probability is one
- We now define convergence *in the mean sense*
  - A sequence of random variables $\mathbf{x}_k$ is said to converge to $\mathbf{x}$ in the *mean square* case if

$$\lim_{k\to\infty} E\left\{||\mathbf{x}_k - \mathbf{x}||^2\right\} = 0$$

  - The function $\mathbf{x}$ is called the limit in the mean and is often written as $\mathbf{x} = \text{l.i.m. } \mathbf{x}_k$, where l.i.m. is always defined as the *limit in the mean square* sense
  - A random variable is *mean square continuous* if

$$\underset{\tau\to t}{\text{l.i.m.}}\ \mathbf{x}(\tau) = \mathbf{x}(t) \quad \text{or} \quad \lim_{\tau\to t} \text{Tr}\left(E\left\{[\mathbf{x}(\tau) - \mathbf{x}(t)][\mathbf{x}(\tau) - \mathbf{x}(t)]^T\right\}\right) = 0$$

Moment Generating Function

- A transformation of the pdf is often done to help calculate the mean, covariance and higher moments
  - For a random variable $\mathbf{x}$ the *moment generating function* is

$$M_{\mathbf{x}}(\mathbf{s}) = E\left\{\exp(\mathbf{s}^T\mathbf{x})\right\}$$

  where $\mathbf{s} = [s_1 \;\; s_2 \;\; \cdots \;\; s_n]^T$ is a general vector
  - The cross moments are defined by $E\left\{x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}\right\}$, which can be computed by

$$E\left\{x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}\right\} = \left.\frac{\partial^k M_{\mathbf{x}}(\mathbf{s})}{\partial s_1^{k_1} \cdots \partial s_n^{k_n}}\right|_{s_1 = s_2 = \cdots = s_n = 0}$$

  where $k = k_1 + k_2 + \cdots + k_n$
    - For a scalar random variable x, the k^{th}-order moment is simply defined by

$$E\left\{x^k\right\} = \left[d^k M_x(s)/ds^k\right]_{s=0}$$

    - For a random vector $\mathbf{x}$, the second moment is simply

$$\nabla_{\mathbf{s}}^2 M_{\mathbf{x}}(\mathbf{s})\big|_{\mathbf{s}=\mathbf{0}}$$

Characteristic Function

- The *characteristic function*, denoted by $\phi_{\mathrm{x}}(\mathrm{s})$, is related to the moment generating function by

$$\begin{aligned}\phi_{\mathbf{x}}(\mathbf{s}) &= M_{\mathbf{x}}(j\,\mathbf{s}) \\ &= E\left\{j\exp(\mathbf{s}^T\mathbf{x})\right\} \\ &= \int_{-\infty}^{\infty} j\exp(\mathbf{s}^T\mathbf{x})\,p(\mathbf{x})\;d\mathbf{x}\end{aligned}$$

where j is the imaginary unit with $j^2 = -1$

  - Note that this function can also be viewed as the Fourier transform of the probability density function
  - The pdf can be computed using the inverse Fourier transform

$$\begin{aligned}p(\mathbf{x}) &= \frac{1}{(2\pi)^n}\int_{-\infty}^{\infty}\exp(-j\,\mathbf{s}^T\mathbf{x})\phi_{\mathbf{x}}(\mathbf{s})\;d\mathbf{s} \\ &= \int_{-\infty}^{\infty}\exp(-j2\pi\,\mathbf{s}^T\mathbf{x})\phi_{\mathbf{x}}(\mathbf{s})\;d\mathbf{s}\end{aligned}$$

Transformation of Variables

- If $\mathbf{x}$ is a random vector whose density function $p(\mathbf{x})$ is known, and if $\mathbf{y} = \mathbf{f}(\mathbf{x})$ is an arbitrary (generally nonlinear) one-to-one transformation, then it can be shown that the density function of $\mathbf{y}$ is given by

$$p(\mathbf{y}) = p(\mathbf{x}) \left| \det\left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right) \right|^{-1} \qquad (2)$$

with $\mathbf{x}$ on the right-hand side given by

$$\mathbf{x} = \mathbf{f}^{-1}(\mathbf{y})$$

where $\mathbf{f}^{-1}(\mathbf{y})$ denotes the “reverse” relationship

- Thus, to convert the density function of $\mathbf{x}$ to the density function of $\mathbf{y}$, simply write the density of $\mathbf{x}$ in terms of $\mathbf{y}$ and multiply by the inverse determinant of the Jacobian matrix
- Difficult part is to determine the inverse relationship
  - May not be possible

Transformation of Variables – Example

- Assume the following quadratic model

$$y = a\,x^2$$

  - Note that for each value of y there are two x-values
  - Assume that x has the following Gaussian density function

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$$

  - When applying (2) we need to account for all multiplicities for multiple roots (this essentially means that the right-hand side of (2) needs to be multiplied by 2), which leads to

$$p(y) = \frac{1}{\sigma\sqrt{2\pi\,a\,y}} \exp\left(\frac{-y}{2a\,\sigma^2}\right), \quad \text{for } y > 0$$

  and $p(y) = 0$ for $y < 0$
  - Note that $p(y)$ is not Gaussian

Gaussian Random Variables (i)

- Recall that the multidimensional case is given by

$$p(\mathbf{x}) = \frac{1}{[\det(2\pi R)]^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T R^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

- Let's prove that its integral is one
  - First we need to prove the following "Gaussian Integral"

$$g \equiv \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}x^2\right)\, dx = \sqrt{2\pi}$$

  - Use Poisson's approach by squaring g to simplify the integral

$$\begin{aligned}
\left[\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}x^2\right)\, dx\right]^2 &= \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}x^2\right)\, dx \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}y^2\right)\, dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(x^2+y^2)\right]\, dx\, dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(x^2+y^2)\right]\, dy\, dx
\end{aligned}$$

Gaussian Random Variables (ii)

- Two basic ways to compute the integral
  - Cartesian approach
  - Polar approach
    - This approach requires the "Jacobian determinant," which is easier than the Cartesian approach once the determinant is understood
    - This involves the transformation of variables approach
    - Choose to use the Cartesian approach here
- Cartesian approach goes back to Laplace (1812)
  - Let

$$y = x\,s$$
$$dy = x\,ds$$

- Since the limits on s as $y \to \pm\infty$ depend on the sign of x, use the fact that $\exp(-x^2/2)$ is an even function to simplify the calculation
  - Therefore, the integral over all real numbers is just twice the integral from zero to infinity

$$\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}x^2\right)\,dx = 2\int_{0}^{\infty} \exp\left(-\frac{1}{2}x^2\right)\,dx$$

Gaussian Random Variables (iii)

- Then

$$
\begin{aligned}
g^2 &= 4 \int_0^\infty \int_0^\infty \exp\left[-\frac{1}{2}(x^2+y^2)\right]\, dy\, dx \\
&= 4 \int_0^\infty \left(\int_0^\infty \exp\left[-\frac{1}{2}x^2(1+s^2)\right] x\ ds \right)\ dx \\
&= 4 \int_0^\infty \left(\int_0^\infty \exp\left[-\frac{1}{2}x^2(1+s^2)\right] x\ dx \right)\ ds \\
&= 4 \int_0^\infty \left\{ -\frac{1}{(1+s^2)} \exp\left[-\frac{1}{2}x^2(1+s^2)\right] \right\} \Bigg|_{x=0}^{x=\infty}\ ds \\
&= 4 \int_0^\infty -\frac{1}{(1+s^2)}(0-1)\ ds \\
&= 4 \arctan(s)\big|_{x=0}^{x=\infty} \\
&= 4\left(\frac{\pi}{2} - 0\right) \\
&= 2\pi
\end{aligned}
$$

Substituted

$$y = x\,s$$

$$dy = x\,ds$$

- So $g = \sqrt{2\pi}$ ✓

Gaussian Random Variables (iv)

- Take the eigenvalue/eigenvector decomposition of R^{-1}

$$R^{-1} = V\Lambda V^T, \quad \text{with} \quad \Lambda = \operatorname{diag}(\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_n)$$

- Define $\mathbf{z} \equiv \mathbf{x} - \boldsymbol{\mu}$ and $\mathbf{y} \equiv V^T\mathbf{z}$
- Then

$$\begin{aligned}(\mathbf{x}-\boldsymbol{\mu})^T R^{-1}(\mathbf{x}-\boldsymbol{\mu}) &\equiv \mathbf{z}^T R^{-1}\mathbf{z} \\ &= \mathbf{z}^T V\Lambda V^T\mathbf{z} \\ &\equiv \mathbf{y}^T\Lambda\,\mathbf{y} \\ &= \sum_{i=1}^{n}\lambda_i y_i^2\end{aligned}$$

- Also, using $\det(V\Lambda V^T) = \det(V)\det(\Lambda)\det(V^T)$, and since $\det(V) = \det(V^T) = \pm 1$, then it's obvious that

$$\det\left(R^{-1}\right) = \prod_{i=1}^{n}\lambda_i \quad \text{and} \quad \det\left(R\right) = \prod_{i=1}^{n}\lambda_i^{-1}$$

Gaussian Random Variables (v)

- Integrate the exponential term of the Gaussian pdf

$$\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T R^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] d\mathbf{x} = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\mathbf{z}^T R^{-1}\mathbf{z}\right) d\mathbf{z}$$

- Now look at the density function transformation from $\mathbf{z}$ to $\mathbf{y} = V^T\mathbf{z}$

$$p(\mathbf{y}) = p(\mathbf{z})\left|\det\left(\frac{\partial V^T\mathbf{z}}{\partial\mathbf{z}}\right)\right|^{-1} = p(\mathbf{z})\left|\det\left(V^T\right)\right|^{-1} = p(\mathbf{z})\left|\pm 1\right|^{-1} = p(\mathbf{z})$$

- Then

$$\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\mathbf{z}^T R^{-1}\mathbf{z}\right) d\mathbf{z} = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\lambda_i\, y_i^2\right) d\mathbf{y}$$

Used variable transformation

$$= \prod_{i=1}^{n}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\lambda_i\, y_i^2\right) dy_i$$

Used Gaussian Integral

Denominator of $p(\mathbf{x})$

$$= (2\pi)^{n/2}\prod_{i=1}^{n}\lambda_i^{-1/2} = [\det(2\pi R)]^{1/2} \quad \checkmark$$

Gaussian Random Variables (vi)

- Find the characteristic function
  - Begin with scalar normal distribution with $z_i \sim N(0,1)$
  - Then

$$M_{z_i}(s_i) = E\left\{e^{s_i z_i}\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{s_i z_i} e^{-z_i^2/2}\, dz_i$$

$$= e^{s_i^2/2} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(z_i - s_i)^2/2}\, dz_i}_{N(s_i,1)} = e^{s_i^2/2}$$

$$= 1$$

  - If $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ is a random sample with each z_i from $N(0,1)$, then

$$M_{\mathbf{z}}(\mathbf{s}) = E\left\{\exp\left(\sum_{i=1}^{n} s_i z_i\right)\right\} = E\left\{\prod_{i=1}^{n} \exp\left(s_i z_i\right)\right\} \overset{\text{ind}}{=} \prod_{i=1}^{n} E\left\{\exp\left(s_i z_i\right)\right\}$$

$$= \prod_{i=1}^{n} M_{z_i}(s_i) = \exp\left\{\sum_{i=1}^{n} s_i^2/2\right\} = \exp\left\{\mathbf{s}^T \mathbf{s}/2\right\}$$

Gaussian Random Variables (vii)

- From $R = \mathcal{L}\mathcal{L}^T$, then $\mathbf{x} = \mathcal{L}\mathbf{z} + \boldsymbol{\mu}$, so $E\{\mathbf{x}\} = \boldsymbol{\mu}$ and $\operatorname{cov}\{\mathbf{x}\} = R$
- Then the moment generating function for a Gaussian is given by

$$
\begin{aligned}
M_{\mathbf{x}}(\mathbf{s}) = E\left\{\exp(\mathbf{s}^T\mathbf{x})\right\} &= E\left\{\exp\left[\mathbf{s}^T(\mathcal{L}\,\mathbf{z} + \boldsymbol{\mu})\right]\right\} \\
&= \exp(\mathbf{s}^T\boldsymbol{\mu})\, E\left\{\exp\left(\mathbf{s}^T\mathcal{L}\,\mathbf{z}\right)\right\} \\
&= \exp(\mathbf{s}^T\boldsymbol{\mu})\, M_{\mathbf{z}}(\mathcal{L}^T\mathbf{s}) \\
&= \exp(\mathbf{s}^T\boldsymbol{\mu})\ \exp\left[\frac{1}{2}(\mathcal{L}^T\mathbf{s})^T(\mathcal{L}^T\mathbf{s})\right] \\
&= \exp(\mathbf{s}^T\boldsymbol{\mu})\ \exp\left(\frac{1}{2}\mathbf{s}^T\mathcal{L}\,\mathcal{L}^T\mathbf{s}\right) \\
&= \exp(\mathbf{s}^T\boldsymbol{\mu})\ \exp\left(\frac{1}{2}\mathbf{s}^T R\,\mathbf{s}\right) \\
&= \exp\left(\mathbf{s}^T\boldsymbol{\mu} + \frac{1}{2}\mathbf{s}^T R\,\mathbf{s}\right)
\end{aligned}
$$

- Characteristic function is then given by

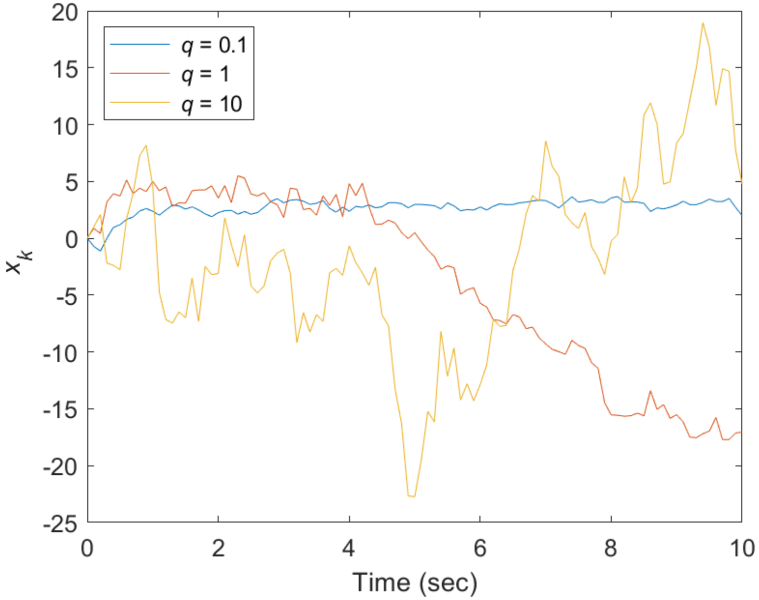
$$
\phi_{\mathbf{x}}(\mathbf{s}) = M_{\mathbf{x}}(j\,\mathbf{s}) = \exp\left(j\mathbf{s}^T\boldsymbol{\mu} - \frac{1}{2}\mathbf{s}^T R\,\mathbf{s}\right)
$$

Random Walk Process

- A *random walk process* is defined as a process where the current value of a variable is composed of the past value plus an error term defined by a white-noise process
  - Consider the following discrete-time process

$$x_{k+1} = x_k + w_k, \quad w_k \sim N(0, q)$$

    where w_k is a zero-mean Gaussian white-noise process
  - This indicates that the change $x_{k+1} - x_k = w_k$ is a random process



  - Thus, the best prediction of x for the next period is the current value
  - It can be shown that the mean of a random walk process is constant but its variance is not
  - Therefore, a random walk process is non-stationary and its variance increases with k

Wiener Process (i)

- The *Wiener process* is the limiting form of the random walk
  - This process, denoted by $\beta(t)$ for a single variable, has the following pdf $p(\beta(t)) = N(0, qt)$, where q is a constant
  - It relates to a zero-mean white noise, denoted by $w(t)$, through

$$\beta(t) = \int_0^t w(\tau)\, d\tau$$

    where $E\{w(t_1)\, w(t_2)\} = q\,\delta(t_1 - t_2)$
  - Also, it can be shown that $E\{\beta(t_1)\,\beta(t_2)\} = q \min(t_1, t_2)$
- Assume that $q = 1$, and study the behavior of the Wiener process using the Fokker-Planck equation (more on this later)
  - The drift coefficient is zero, and the diffusion coefficient is 1
  - So the Fokker-Planck equation is given by

$$\frac{\partial}{\partial t} p(\beta(t)|\beta(t_0)) = \frac{1}{2}\frac{\partial^2}{\partial \beta^2} p(\beta(t)|\beta(t_0))$$

    with initial condition $p(\beta(t_0)|\beta(t_0)) = \delta(\beta(t) - \beta(t_0))$

Wiener Process (ii)

- This can be solved using the following characteristic function

$$\phi_\beta(s,\,t) = \int_{-\infty}^{\infty} \exp(js\beta)p(\beta(t)|\beta(t_0))\;d\beta$$

which satisfies

$$\frac{\partial\phi}{\partial t} = -\frac{1}{2}s^2\phi$$

so that

$$\phi_\beta(s,\,t) = \exp\left[-\frac{1}{2}s^2(t-t_0)\right]\phi_\beta(s,\,t_0)$$

- From the initial condition we have $\phi_\beta(s,\,t_0) = \exp(js\beta(t_0))$ so that

$$\phi_\beta(s,\,t) = \exp\left[js\beta(t_0) - \frac{1}{2}s^2(t-t_0)\right]$$

- Using this equation the following pdf can be derived

$$p(\beta(t)|\beta(t_0)) = \frac{1}{\sqrt{2\pi(t-t_0)}}\exp\left[-\frac{1}{2}\frac{(\beta(t)-\beta(t_0))^2}{t-t_0}\right]$$

Wiener Process (iii)

- This represents a Gaussian variable with

$$E\{\beta(t)\} = \beta(t_0)$$
$$E\left\{[\beta(t) - \beta(t_0)]^2\right\} = t - t_0$$

- This indicates that an initially sharp distribution spreads in time

- The vector case with $\boldsymbol{\beta}(t)$ has the following properties

$$E\{\boldsymbol{\beta}(t)\} = \boldsymbol{\beta}(t_0)$$
$$E\{[\beta_i(t) - \beta_i(t_0)][\beta_j(t) - \beta_j(t_0)]\} = (t - t_0)\delta_{ij}$$

- The one-variable Wiener process is often called *Brownian motion.* The Wiener process is a Markov process. This follows from the fact that it is the integral of white noise

$$\beta(t) = \beta(t_1) + \int_{t_1}^{t} w(\tau)\, d\tau$$

and $w(\tau)$, $\tau \in [t_1,\, t]$, is independent of $\beta(t_1)$

- Furthermore, the state $\mathbf{x}(t)$ of a time-varying dynamic system driven by white noise is also a Markov process

Wiener Process (iv)

- The general vector case has independent Gaussian increments with

$$E\left\{\boldsymbol{\beta}(t_2)-\boldsymbol{\beta}(t_1)\right\}=\mathbf{0}$$

$$E\left\{[\boldsymbol{\beta}(t_2)-\boldsymbol{\beta}(t_1)][\boldsymbol{\beta}(t_2)-\boldsymbol{\beta}(t_1)]^T\right\}=\int_{t_1}^{t_2} Q(t)\;dt \quad (3)$$

where $Q(t)$ is a matrix

- Note that a Wiener process is continuous, but it is not differentiable
  - This is easily seen by attempting to find the limit of $(x_{k+1}-x_k)/\Delta t = w_k/\Delta t$ as $\Delta t \to 0$ for the random walk process
    - Clearly, this is not possible because w_k is a random process
  - Thus, writing the equation

$$\dot{\mathbf{x}}(t)=\mathbf{f}(\mathbf{x}(t),\,t)+G(\mathbf{x}(t),\,t)\,\mathbf{w}(t)$$

is not rigorously correct because $\mathbf{w}(t)$ can only be considered as the hypothetical derivative of $\boldsymbol{\beta}\,(t)$

Wiener Process (v)

- A stochastic integral can be defined through

$$\mathbf{I}(t) = \int_{t_0}^{t} A(\tau)\, d\boldsymbol{\beta}(\tau) \qquad (4)$$

  - Now attempt to define the integral as a mean square limit of

$$\mathbf{s}_N = \sum_{i=1}^{N} A(\tau_i)[\boldsymbol{\beta}(t_i) - \boldsymbol{\beta}(t_{i-1})] \qquad (5)$$

    for $t_{i-1} \leq \tau_i \leq t_{i+1}$ when the partition $t_0 < t_1 < \cdots < t_N$ of $[t_0, t]$ is refined so that $\Delta_N = \max_{1\leq i\leq N} (t_i - t_{i-1}) \to 0$, i.e.

$$\mathbf{I}(t) = \underset{\Delta_N \to 0}{\mathrm{l.i.m.}}\ \mathbf{s}_N$$

  - Before attempting to evaluate this limit, we discuss the *Levy oscillation property* of Brownian motion
    - This is also known as the *quadratic variation property*
  - Consider a unit-diffusion Brownian motion process, denoted by β, and the time partition $t_0 < t_1 < \cdots < t_N = t$

Wiener Process (vi)

- The sums

$$\xi_N = \sum_{i=1}^{N} [\beta(t_i) - \beta(t_{i-1})]^2 - (t - t_0)$$

are random variables with means zero and variances given by $2\sum_{i=1}^{N}(t_i - t_{i-1})^2$, which can be bounded by

$$2 \max_{1 \le i \le N} |t_i - t_{i-1}| \sum_{i=1}^{N} (t_i - t_{i-1}) = 2\Delta_N (t - t_0)$$

- Therefore, we now have $\underset{\Delta_N \to 0}{\text{l.i.m.}}\ \xi_N = 0$ and

$$\underset{\Delta_N \to 0}{\text{l.i.m.}} \sum_{i=1}^{N} [\beta(t_i) - \beta(t_{i-1})]^2 = t - t_0 \qquad (6)$$

- Another way to state this equation is the symbolic notation $[d\beta(t)]^2 = dt$ with probability one in the mean square sense

- For the non-unit diffusion case, we simply have $[d\beta(t)]^2 = q(t)\,dt$, and for the general vector case we have

$$[d\boldsymbol{\beta}(t)\,d\boldsymbol{\beta}^T(t)] = Q(t)\,dt$$

- Using (3) we also have $E\{d\boldsymbol{\beta}(t)\,d\boldsymbol{\beta}^T(t)\} = Q(t)\,dt$
  - This is an important result, which states that not only is $E\{d\boldsymbol{\beta}(t)\,d\boldsymbol{\beta}^T(t)\} = Q(t)\,dt$ true but $[d\boldsymbol{\beta}(t)\,d\boldsymbol{\beta}^T(t)] = Q(t)\,dt$ itself is true for all samples except possibly a set of total probability zero

- Turn our attention now to the scalar version of (5)
  - Evaluating this equation where τ_i is any point on the interval $[t_{i-1},\, t_i]$ will cause the value and properties of (4) to depend on the specific choice of τ_i
  - Thus, evaluation of truncated Taylor series expansions of nonlinear functions of $\beta(t)$ will invalidate the applicability of formal rules of differentials
    - To see this, assume that $A(\tau) = \beta(\tau)$ in (4) so that

$$\int_{t_0}^{t} \beta(\tau)\;d\beta(\tau) = \frac{1}{2}[\beta^2(t) - \beta^2(t_0)] \qquad (7)$$

Wiener Process (viii)

- This result is valid if the sums of the scalar version of (5), denoted by s_N, converge to a unique limit as $\Delta_N \to 0$ for any intermediate points τ_i
- Investigate a few useful choices for τ_i, letting $\tau_i = t_{i-1}$, then

$$
\begin{aligned}
s_N &= \sum_{i=1}^{N} \beta(t_{i-1})[\beta(t_i) - \beta(t_{i-1})] \\
&= \sum_{i=1}^{N} \left[\frac{\beta(t_i) + \beta(t_{i-1})}{2} - \frac{\beta(t_i) - \beta(t_{i-1})}{2} \right] [\beta(t_i) - \beta(t_{i-1})] \\
&= \frac{1}{2} \sum_{i=1}^{N} [\beta^2(t_i) - \beta^2(t_{i-1})] - \frac{1}{2} \sum_{i=1}^{N} [\beta(t_i) - \beta(t_{i-1})]^2 \\
&= \frac{1}{2} [\beta^2(t) - \beta^2(t_0)] - \frac{1}{2} \sum_{i=1}^{N} [\beta(t_i) - \beta(t_{i-1})]^2
\end{aligned}
$$

- From (6) we have

$$
\underset{\Delta_N \to 0}{\mathrm{l.i.m.}}\; s_N = \frac{1}{2} [\beta^2(t) - \beta^2(t_0)] - \frac{t - t_0}{2}
$$

Wiener Process (ix)

- Thus, using $\tau_i = t_{i-1}$ does not provide a convergence of the series to (7)
- Evaluating the series when $\tau_i = (1-\theta)t_{i-1} + \theta t_i$ for $0 \le \theta \le 1$ gives

$$\underset{\Delta_N \to 0}{\text{l.i.m.}}\ s_N = \frac{1}{2}[\beta^2(t) - \beta^2(t_0)] + \left(\theta - \frac{1}{2}\right)\left(\frac{t - t_0}{2}\right)$$

- Setting $\tau_i = (t_{i-1} + t_i)/2$ does provide a convergence of the series to (7)
- This choice corresponds to the *Stratonovich integral*, while the choice $\tau_i = t_{i-1}$ corresponds to the *Itô integral*
- The Stratonovich integral has the advantage that formal rules of integration can be applied
- Despite this attractive feature, though, the Stratonovich integral lacks some important properties possessed by the Itô integral that are essential to Markov process descriptions
  - For example, the Itô integral efficiently uses the property that the Wiener process has independent increments

Wiener Process (x)

- The *stochastic differential* $d\mathbf{I}(t)$ of $\mathbf{I}(t)$ is now simply

$$d\mathbf{I}(t) = A(\tau)\ d\boldsymbol{\beta}$$

- Gives us a mechanism to write a rigorously derived stochastic equation

$$\boxed{d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t),\, t)\, dt + G(\mathbf{x}(t),\, t)\, d\boldsymbol{\beta}(t)}$$

  where $\boldsymbol{\beta}(t)$ has diffusion strength $Q(t)$
  - This equation is often referred to as the *Itô differential equation*
  - Consider the scalar version

$$dx(t) = f(x(t),\, t)\, dt + g(x(t),\, t)\, d\beta(t)$$

  with diffusion strength $q(t)$
  - Stratonovich's stochastic integral, denoted by the subscript S, is related to Itô's integral by

$$\left\{ \int_{t_0}^{t} a(\tau)\ d\beta(\tau) \right\}_S = \int_{t_0}^{t} a(\tau)\ d\beta(\tau) + \frac{1}{2} \int_{t_0}^{t} q(\tau) \frac{\partial a(\tau)}{\partial \beta}\ dt$$

Wiener Process (xi)

- Defining $a(t) \equiv g(x(t),\, t)$ and since $\partial a/\partial\beta = (\partial g/\partial x)\ (\partial x/\partial\beta)$ we have the following equivalent *Stratonovich differential equation*

$$dx(t) = \left[f(x(t),\, t) - \frac{1}{2}\, q(t) g(x(t),\, t) \frac{\partial g(x(t),\, t)}{\partial x}\right] dt + g(x(t),\, t)\, d\beta(t)$$

- The multidimensional form can be written as

$$dx_i(t) = \left[f_i(\mathbf{x}(t),\, t) - \frac{1}{2}\, \mathbf{g}_i^T(\mathbf{x}(t),\, t)\, Q(t)\, \frac{\partial \mathbf{g}_i(\mathbf{x}(t),\, t)}{\partial x_i}\right] dt + \mathbf{g}_i^T(\mathbf{x}(t),\, t)\, d\boldsymbol{\beta}(t)$$

for $i = 1, 2, \ldots, n$, where $\mathbf{g}_i^T(\mathbf{x}(t),\, t)$ is the i^{th} row of $G(\mathbf{x}(t),\, t)$, and $f_i(\mathbf{x}(t),\, t)$ and $x_i(t)$ are the i^{th} elements of $\mathbf{f}(\mathbf{x}(t),\, t)$ and $\mathbf{x}(t)$, respectively

- Both the Itô and Stratonovich forms are equivalent when $G(\mathbf{x}(t),\, t)$ is not a function of $\mathbf{x}$

Itô Stochastic Differential Equations (i)

- Begin with the following stochastic differential equation

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t),\, t)\, dt + G(\mathbf{x}(t),\, t)\, d\boldsymbol{\beta}(t) \tag{8}$$

  - Again, the vector $\beta(t)$ represents Brownian motion of zero mean and diffusion $Q(t)$ so that

$$E\left\{d\boldsymbol{\beta}(t)\, d\boldsymbol{\beta}^T(t)\right\} = Q(t)\, dt$$

$$E\left\{[\boldsymbol{\beta}(t) - \boldsymbol{\beta}(\tau)][\boldsymbol{\beta}(t) - \boldsymbol{\beta}(\tau)]^T\right\} = \int_\tau^t Q(t)\ dt$$

  - The solution can now be characterized in a form given by

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \int_{t_0}^t \mathbf{f}(\mathbf{x}(\tau),\, \tau)\, d\tau + \int_{t_0}^t G(\mathbf{x}(\tau),\, \tau)\, \frac{d\boldsymbol{\beta}(\tau)}{d\tau} d\tau \tag{9}$$

    - The first integral is easily understood, but the second one is in the Itô form whose formal rules of integration and differentiation no longer apply, as established previously
    - Itô calculus develops a more rigorous approach for differentiation and integration of moments for stochastic processes

Itô Stochastic Differential Equations (ii)

- A good summary of the sufficient conditions for the existence and uniqueness of the solutions to (9) in the mean square sense is given by (similar to the ones for ordinary differential equations)

1. The functions $\mathbf{f}(\mathbf{x}(t),\, t)$ and $G(\mathbf{x}(t),\, t)$ are real functions that are uniformly *Lipschitz*. This means that a scalar k that is independent of time can be found such that

$$||\mathbf{f}(\mathbf{x} + \Delta\mathbf{x},\, t) - \mathbf{f}(\mathbf{x},\, t)|| \leq k||\Delta\mathbf{x}||$$
$$||G(\mathbf{x} + \Delta\mathbf{x},\, t) - G(\mathbf{x},\, t)||_F \leq k||\Delta\mathbf{x}||$$

   for all $\mathbf{x}$ and $\Delta\mathbf{x}$ and all t in the interval $[t_0,\, t_f]$ of interest.

2. The functions $\mathbf{f}(\mathbf{x}(t),\, t)$ and $G(\mathbf{x}(t),\, t)$ are continuous in their second (time) argument over the interval $[t_0,\, t_f]$ of interest.

3. The functions $\mathbf{f}(\mathbf{x}(t),\, t)$ and $G(\mathbf{x}(t),\, t)$ are uniformly bounded according to $||\mathbf{f}(\mathbf{x},\, t)|| \leq k(1 + ||\mathbf{x}||^2)$ and $||G(\mathbf{x},\, t)||_F \leq k(1 + ||\mathbf{x}||^2)$.

4. The vector $\mathbf{x}(t_0)$ is a random vector, with finite second moment, which is independent of the Brownian motion.

- In Itô's proof the solution for $\mathbf{x}(t)$ is given by assuming the existence of $\mathbf{x}_k(t)$ with $\mathbf{x}_k(t_0) = \mathbf{x}(t_0)$ and then forming $\mathbf{x}_{k+1}(t)$ through (9) with

$$\mathbf{x}_{k+1}(t) = \mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{f}(\mathbf{x}_k(\tau), \tau)\, d\tau + \int_{t_0}^{t} G(\mathbf{x}_k(\tau), \tau)\, \frac{d\boldsymbol{\beta}(\tau)}{d\tau} d\tau$$

- If the four sufficient conditions are met, then the sequence of $\mathbf{x}_k(t)$ converges in the mean sense and with probability one on any finite interval $[t_0,\, t_f]$ to the solution $\mathbf{x}(t)$

- The solution $\mathbf{x}(t)$ has the following useful properties

1. It is mean square continuous, i.e., $\underset{\tau \to t}{\text{l.i.m.}}\ \mathbf{x}(\tau) = \mathbf{x}(t)$.
2. The variables $\mathbf{x}(t) - \mathbf{x}(t_0)$ and $\mathbf{x}(t)$ are both independent of the future increments of $\boldsymbol{\beta}(t)$.

3. It is a Markov process. Consider $t \geq t'$, so that

$$\mathbf{x}(t) = \mathbf{x}(t') + \int_{t'}^{t} \mathbf{f}(\mathbf{x}(\tau), \tau)\, d\tau + \int_{t'}^{t} G(\mathbf{x}(\tau), \tau)\, \frac{d\boldsymbol{\beta}(\tau)}{d\tau} d\tau \qquad (10)$$

   This clearly shows that $\mathbf{x}(t)$ depends on $\mathbf{x}(t')$ and $d\boldsymbol{\beta}(\tau)$, $t' \leq \tau \leq t$, and the latter is independent of $\mathbf{x}(s)$ with $s \leq t'$. Thus, the conditional probability for $\mathbf{x}(t)$ given $\mathbf{x}(t')$ and $\mathbf{x}(s)$, $s \leq t'$, equals the distribution conditioned only on $\mathbf{x}(t')$. This proves that it is Markov.

4. The mean squared value of each component, i.e., $E\left\{x_i^2(t)\right\}$, is bounded by some finite value. Also, $\int_{t_0}^{t_f} E\left\{x_i^2(t)\right\}\; dt < \infty$.

5. The probability of a change in $\mathbf{x}(t)$ in a small interval Δt is of higher order than Δt:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \underset{||\boldsymbol{\xi}-\boldsymbol{\rho}|| \geq \delta}{\int_{-\infty}^{\infty}} p(\boldsymbol{\xi}(t+\Delta t)|\boldsymbol{\rho}(t))\; d\boldsymbol{\xi} = 0 \qquad (11)$$

   where the notation means that the integration over $\boldsymbol{\xi}$ is to be called outside the ball of radius δ about $\boldsymbol{\rho}$.

Itô Stochastic Differential Equations (v)

6. The *drift* of $\mathbf{x}(t)$ is $\mathbf{f}(\mathbf{x}(t),\, t)$. Using (8) we have

$$\begin{aligned}
\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (\boldsymbol{\xi} - \boldsymbol{\rho})\, p(\boldsymbol{\xi}(t+\Delta t) | \boldsymbol{\rho}(t))\ d\boldsymbol{\xi} \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} E\left\{ \mathbf{x}(t+\Delta t) - \mathbf{x}(t) | \mathbf{x}(t) = \boldsymbol{\rho} \right\} \\
&= \mathbf{f}(\boldsymbol{\rho},\, t)
\end{aligned} \tag{12}$$

This states that the mean rate of change in $\mathbf{x}(t)$ going from t to $t + \Delta t$ is $\mathbf{f}(\mathbf{x}(t),\, t)$ as $\Delta t \to \infty$.

7. The *diffusion* of $\mathbf{x}(t)$ is $G(\mathbf{x}(t),\, t)\, Q(t)\, G^T(\mathbf{x}(t),\, t)$:

$$\begin{aligned}
\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (\boldsymbol{\xi} - \boldsymbol{\rho})(\boldsymbol{\xi} - \boldsymbol{\rho})^T p(\boldsymbol{\xi}(t+\Delta t) | \boldsymbol{\rho}(t))\ d\boldsymbol{\xi} \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} E\left\{ [\mathbf{x}(t+\Delta t) - \mathbf{x}(t)][\mathbf{x}(t+\Delta t) - \mathbf{x}(t)]^T | \mathbf{x}(t) = \boldsymbol{\rho} \right\} \\
&= G(\boldsymbol{\rho},\, t)\, Q(t)\, G^T(\boldsymbol{\rho},\, t)
\end{aligned} \tag{13}$$

8. The higher-order infinitesimals in the progression of (11)-(13) are all zero:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (\xi_i - \rho_i)^k p(\boldsymbol{\xi}(t + \Delta t)|\boldsymbol{\rho}(t))\ d\boldsymbol{\xi} = 0$$

for $k \geq 2$. A similar relation exists for general products greater than second degree as well. This implies that the process does not diffuse "too fast."

Itô’s Formula (i)

- Suppose we wish to apply Itô’s calculus on a scalar function $\psi(\mathbf{x}(t), t)$ that has continuous first and second partial derivatives with respect to $\mathbf{x}(t)$ and is continuously differentiable with respect to time
  - As stated previously, formal rules of integration and differentiation are not valid in Itô’s calculus
  - To develop a stochastic differential equation for $\psi(\mathbf{x}(t), t)$ the *Ito formula* must be used
  - First calculate

$$d\psi(\mathbf{x}(t),\, t) = \psi(\mathbf{x}(t) + d\mathbf{x}(t),\, t + dt) - \psi(\mathbf{x}(t),\, t)$$

  - Now consider expanding the right-hand side using a Taylor series expansion, so that

$$d\psi(\mathbf{x}(t),\, t) = \frac{\partial \psi}{\partial t}\, dt + \frac{\partial \psi}{\partial \mathbf{x}^T}\, d\mathbf{x}(t) + \frac{1}{2}\frac{\partial^2 \psi}{\partial t^2}\, dt^2 + \frac{1}{2}\, d\mathbf{x}^T(t)\, \frac{\partial^2 \psi}{\partial \mathbf{x}\, \partial \mathbf{x}^T}\, d\mathbf{x}(t) + \cdots$$

- Substituting (8) and retaining only terms up to first order in dt and second order in $d\beta$ yields

$$d\psi(\mathbf{x}(t),\,t) = \frac{\partial\psi}{\partial t}\,dt + \frac{\partial\psi}{\partial\mathbf{x}^T}\,d\mathbf{x}(t) + \frac{1}{2}\mathrm{Tr}\left[G(\mathbf{x}(t),\,t)\,d\boldsymbol{\beta}(t)\,d\boldsymbol{\beta}^T(t)\,G^T(\mathbf{x}(t),\,t)\frac{\partial^2\psi}{\partial\mathbf{x}\,\partial\mathbf{x}^T}\right]$$

where the identity $\mathrm{Tr}(\mathbf{x}\,\mathbf{y}^T) = \mathbf{x}^T\mathbf{y}$ has been used

- Now, using the Levy property $[d\boldsymbol{\beta}(t)\,d\boldsymbol{\beta}^T(t)] = Q(t)\,dt$ gives

$$d\psi(\mathbf{x}(t),\,t) = \frac{\partial\psi}{\partial t}\,dt + \frac{\partial\psi}{\partial\mathbf{x}^T}\,d\mathbf{x}(t) + \frac{1}{2}\mathrm{Tr}\left[G(\mathbf{x}(t),\,t)\,Q(t)\,G^T(\mathbf{x}(t),\,t)\frac{\partial^2\psi}{\partial\mathbf{x}\,\partial\mathbf{x}^T}\right]dt \qquad (14)$$

Itô's Formula (iii)

- Equation (8) is often combined with (14) and written in the form

$$\boxed{d\psi(\mathbf{x}(t),\, t) = \frac{\partial \psi}{\partial t}\, dt + \mathcal{L}[\psi(\mathbf{x}(t),\, t)]\, dt + \frac{\partial \psi}{\partial \mathbf{x}^T}\, G(\mathbf{x}(t),\, t)\, d\boldsymbol{\beta}(t)} \tag{15}$$

where

$$\boxed{\mathcal{L}[\psi(\mathbf{x}(t),\, t)] \equiv \frac{\partial \psi}{\partial \mathbf{x}^T}\, \mathbf{f}(\mathbf{x}(t),\, t) + \frac{1}{2} \mathrm{Tr}\left[G(\mathbf{x}(t),\, t)\, Q(t)\, G^T(\mathbf{x}(t),\, t)\, \frac{\partial^2 \psi}{\partial \mathbf{x}\, \partial \mathbf{x}^T} \right]}$$

- The term $\mathcal{L}[\psi(\mathbf{x}(t), t)]$ is the differential generator of the process

Itô’s Formula – Example (i)

- Consider the following system

$$dx(t) = d\beta(t)$$

with $q(t)$ being a constant denoted by q

  - This equation states that $x(t)$ is itself Brownian motion, which is heuristically written as $\dot{x}(t) = w(t)$
  - Now consider the following nonlinear function

$$\psi(x(t),\, t) = e^{x(t)} = e^{\beta(t)}$$

  - From (15) this satisfies the following stochastic differential equation

$$d\psi(x(t),\, t) = e^{x(t)}\, dx(t) + \frac{1}{2} q\, e^{x(t)}\, dt$$

or

$$d\left[e^{\beta(t)}\right] = e^{\beta(t)}\, d\beta(t) + \frac{1}{2} q\, e^{\beta(t)}\, dt$$

  - Because of the last term, this does not satisfy formal rules for differentials

Itô's Formula – Example (ii)

- This can be overcome by defining $\gamma(t) \equiv e^{\beta(t)}$, which yields

$$d\gamma(t) = \frac{1}{2} q\,\gamma(t)\,dt + \gamma(t)\,d\beta(t)$$

with $\gamma(t_0) = 1$ with probability one

- This is now the appropriate stochastic differential equation in the form of (8) to yield a solution in the form of $e^{\beta(t)}$, since $\beta(t_0) = 0$ with probability one
- This example clearly shows that stochastic differential equations do not obey formal rules of integration either
- The differential equation that would have to be proposed by formal rules is

$$dz(t) = z(t)\,d\beta(t)$$

with $z(t_0) = 1$ with probability one

- The solution for this equation with $t_0 = 0$ using (15) is given by

$$z(t) = e^{\beta(t) - qt/2}$$

Fokker-Planck Equation (i)

- All the material to this point has been leading to answering the question "how does the pdf propagate in time?" for Gaussian inputs into a nonlinear system
  - The answer lies in the *Fokker-Planck equation*, also known as the *forward Kolmogorov equation*
  - To derive this equation, begin with the scalar version of (8)

$$dx(t) = f(x(t),\, t)\, dt + g(x(t),\, t)\, d\beta(t)$$

  - From the Itô formula in (15), dropping the explicit notation for time and state dependence for now, we have

$$d\psi = \left[\frac{\partial \psi}{\partial t} + f\,\frac{\partial \psi}{\partial x} + \frac{1}{2} g^2 q\,\frac{\partial^2 \psi}{\partial x^2}\right] dt + g\,\frac{\partial \psi}{\partial x}\, d\beta$$

  - Taking the expectation of both sides yields

$$\frac{d}{dt} E\left\{\psi\right\} = E\left\{f\,\frac{\partial \psi}{\partial x} + \frac{1}{2} g^2 q\,\frac{\partial^2 \psi}{\partial x^2}\right\}$$

Fokker-Planck Equation (ii)

- Using the definition of expectation gives

$$\frac{d}{dt}\int_{-\infty}^{\infty}\psi\, p(x(t)|x(t'))\;dx = \int_{-\infty}^{\infty}\left[f\,\frac{\partial\psi}{\partial x}+\frac{1}{2}g^2 q\,\frac{\partial^2\psi}{\partial x^2}\right]p(x(t)|x(t'))\;dx$$

where $p(x(t)|x(t'))$ denotes the conditional probability of $x(t)$ given $x(t')$ with $t > t'$

- Integrating by parts, and using $p \equiv p(x(t)|x(t'))$, yields

$$\int_{-\infty}^{\infty}\psi\frac{\partial p}{\partial t}\;dx = \int_{-\infty}^{\infty}\left[-\frac{\partial f p}{\partial x}+\frac{1}{2}\frac{\partial g^2 p}{\partial x^2}\right]\psi\;dx$$

when $p(x(t)|x(t'))$ and $\partial p(x(t)|x(t'))/\partial x$ vanish as $x \to \pm\infty$

- Since ψ is arbitrary, and simplifying the notation for $p(x(t)|x(t'))$ to be just $p(x(t),\,t)$, we now have

$$\frac{\partial p(x(t),\,t)}{\partial t} = -\frac{\partial}{\partial x}[f(x(t),\,t)\,p(x(t),\,t)]+\frac{1}{2}\frac{\partial^2}{\partial x^2}[g^2(x(t),\,t)\,p(x(t),\,t)]$$

- This is the Fokker-Planck equation for scalar systems

Fokker-Planck Equation (iii)

- The general Fokker-Planck equation for multidimensional systems is given by

$$\boxed{\begin{aligned}\frac{\partial}{\partial t}p(\mathbf{x}(t),\,t) &= -\sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left[f_i(\mathbf{x}(t),\,t)\,p(\mathbf{x}(t),\,t)\right]\\ &\quad+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2}{\partial x_i\,\partial x_j}\left[\{G(x(t),\,t)\,Q(t)\,G^T(x(t),\,t)\}_{ij}\;p(\mathbf{x}(t),\,t)\right]\end{aligned}}$$

where $f_i(\mathbf{x}(t),\ t)$ is the i^{th} element of $\mathbf{f}(\mathbf{x}(t),\ t)$, and $\{G\,Q\,G^T\}_{ij}$ is the ij^{th} element element of $G\,Q\,G^T$

- This equation only has a closed-form solution for a small number of cases
- Note that this form uses the Itô interpretation, which is commonly used among mathematicians

Fokker-Planck Equation (iv)

- The Stratonovich interpretation uses the following replacement

$$f_i(\mathbf{x}(t),\, t) \leftarrow f_i(\mathbf{x}(t),\, t) - \frac{1}{2}\,\mathbf{g}_i^T(\mathbf{x}(t),\, t)\, Q(t)\, \frac{\partial \mathbf{g}_i(\mathbf{x}(t),\, t)}{\partial x_i}$$

  where $\leftarrow$ denotes replacement, and where $\mathbf{g}_i^T(\mathbf{x}(t),\, t)$ is the i^{th} row of $G(\mathbf{x}(t),\, t)$
- The Stratonovich interpretation is more popular among engineers and physicists because standard rules of integration can be applied
  - For most systems of interest to engineers, the matrix G is usually not a function of $\mathbf{x}(t)$ so the interpretation issue is usually not a concern

Fokker-Planck Equation – Example (i)

- Consider the following first order system

$$dx(t) = f\,x(t)\,dt + d\beta(t)$$

or written heuristically as $\dot{x}(t) = f\,x(t) + w(t)$, with $E\{w(t)\} = 0$ and $E\{w(t)\,w(\tau)\} = q(t)\,\delta(t-\tau)$

  - The Fokker-Planck equation is given with $f(x(t),\, t) = f\,x(t)$ and $g(x(t),\, t) = 1$
    - Note that the Itô and Stratonovich interpretations are equivalent in this case
  - Assume a Gaussian solution with

$$p(\mathbf{x}(t),\, t) = \frac{1}{\sqrt{2\pi\, p(t)}} \exp\left\{ -\frac{[x(t) - \mu(t)]^2}{2\,p(t)} \right\}$$

where $\mu(t)$ is the mean and $p(t)$ is the variance

Fokker-Planck Equation – Example (ii)

- The Fokker-Planck equation now becomes

$$\frac{1}{2}p^{-3/2}(t)\left\{p^{-1}(t)\left[x(t)-\mu(t)\right]^2-1\right\}\dot{p}(t)+p^{-3/2}(t)[x(t)-\mu(t)]\,\dot{\mu}(t)$$

$$=-f\,p^{-1/2}(t)+f\,x(t)[x(t)-\mu(t)]\,p^{-3/2}(t)+\frac{1}{2}q(t)\,p^{-3/2}(t)\left\{p^{-1}(t)[x(t)-\mu(t)]^2-1\right\}$$

- Equating terms independent of $[x(t)-\mu(t)]$ yields

$$\dot{p}(t)=2f\,p(t)+q(t)$$

which is exactly the Kalman covariance propagation equation

- Equating terms that depend on $[x(t)-\mu(t)]$ yields

$$\dot{\mu}(t)+\frac{1}{2}\frac{x(t)-\mu(t)}{p(t)}\dot{p}(t)=f\,x(t)+q(t)\,\frac{1}{2}\frac{x(t)-\mu(t)}{p(t)}$$

- Substituting $\dot{p}(t)=2f\,p(t)+q(t)$ yields

$$\dot{\mu}(t)=f\,\mu(t)$$

which is exactly the Kalman propagation equation

- This clearly shows that the scalar Kalman propagation equations are consistent with the Fokker-Planck equation

Kushner Equation (i)

- The equation by Kushner modifies the Fokker-Planck equation with continuous measurements
  - Note that another form was introduced by Zakai
  - However, Kushner's equation is a nonlinear stochastic partial differential equation and satisfies the normalization requirement for a pdf, while Zakai's equation is a linear stochastic partial differential equation for the un-normalized pdf
  - The continuous-time measurements are modeled using the process described by

$$d\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t),\, t)\, dt + d\mathbf{b}(t)$$

where $\mathbf{b}(t)$ is Brownian motion, independent of $\beta(t)$ from (8), with diffusion given by

$$E\left\{d\mathbf{b}(t)\, d\mathbf{b}^T(t)\right\} = R(t)\, dt$$

  - This corresponds heuristically to

$$\dot{\mathbf{z}}(t) \equiv \mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t),\, t) + \mathbf{v}(t)$$

with $E\{\mathbf{v}(t)\,\mathbf{v}^T(\tau)\} = R(t)\,\delta(t-\tau)$

Kushner Equation (ii)

- We wish to establish the equation for the time history of the conditional density of the state $\mathbf{x}(t)$, but now conditioned on the entire history of measurements observed up to time t
- The conditional density now becomes

$$p(\mathbf{x}(t),\, t \,|\, \mathbf{z}(\tau), t_0 \leq \tau \leq t)$$

  or in simple shorthand notation $p\left(\mathbf{x}|\mathbf{z}\right)$

- The conditional density satisfies the Kushner equation, which is given by

$$\boxed{\frac{\partial p(\mathbf{x}|\mathbf{z})}{\partial t} = \mathcal{L}(\mathbf{x}|\mathbf{z}) + \left[\mathbf{h}(\mathbf{x}(t),\, t) - \mathbf{m}(\mathbf{x}(t),\, t)\right]^T R^{-1}(t) \left[\mathbf{y}(t) - \mathbf{m}(\mathbf{x}(t),\, t)\right] p(\mathbf{x}|\mathbf{z})}$$

  where

$$\boxed{\mathbf{m}(\mathbf{x}(t),\, t) \equiv \int_{-\infty}^{\infty} \mathbf{h}(\mathbf{x}(t),\, t) p(\mathbf{x}|\mathbf{z})\; d\mathbf{x}}$$

Kushner Equation (iii)

and

$$\mathcal{L}(\mathbf{x}|\mathbf{z}) \equiv -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left[f_i(\mathbf{x}(t),\, t)\, p(\mathbf{x}|\mathbf{z})\right] + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2}{\partial x_i\, \partial x_j}\left[\left\{G(x(t),\, t)\, Q(t)\, G^T(x(t),\, t)\right\}_{ij}\, p(\mathbf{x}|\mathbf{z})\right]$$

- Note that this equation corresponds to the Fokker-Planck equation
- If no measurement information exists, i.e. $R^{-1}(t) = 0$, then the Kushner equation reduces directly to the Fokker-Planck equation

History of Stochastic Filtering Theory

Author(s) (year)	Method	Solution	Comments
Kolmogorov (1941)	innovations	exact	L, S
Wiener (1942)	spectral factorization	exact	L, S, IM
Levinson (1947)	lattice filter	approximate	L, S, FM
Bode & Shannon (1950)	innovations, whitening	exact	L, S
Zadeh & Ragazzini (1950)	innovations, whitening	exact	L, NS
Kalman (1960)	orthogonal projection	exact	LQG, NS, D
Kalman & Bucy (1961)	recursive Riccati	exact	LQG, NS, C
Stratonovich (1960)	conditional Markov	exact	NL, NS
Kushner (1967)	PDE	exact	NL, NS
Zakai (1969)	PDE	exact	NL, NS
Handschin & Mayne (1969)	Monte Carlo	approximate	NL, NS, NG
Bucy & Senne (1971)	point-mass, Bayes	approximate	NL, NS, NG
Kailath (1971)	innovations	exact	L, NS, NG
Beneš (1981)	Beneš	ES of Zakai Equation	NL, FD
Daum (1986)	virtual measurement	ES of FPK Equation	NL, FD
Gordon, Salmond, & Smith (1993)	bootstrap	approximate	NL, NS, NG
Julier & Uhlmann (1997)	unscented transform	approximate	NL, NG

LQG denotes “linear quadratic-Gaussian,” PDE denotes “partial differential equation,” FPK denotes “Fokker-Planck-Kolmogorov,” ES denotes “exact solution,” L denotes “linear,” S denotes “stationary,” NS denotes “non-stationary,” C denotes “continuous,” D denotes “discrete,” IM denotes “infinite memory,” FM denotes “finite memory,” NG denotes “non-Gaussian,” NL denotes “nonlinear,” and FD denotes “finite-dimensional”