Lecture Note Sketches

Introduction to Stochastic Processes and Stochastic Differential Equations

Hermann Riecke

Engineering Sciences and Applied Mathematics

h-riecke@northwestern.edu

January 6, 2010

©Hermann Riecke

Contents

1 Introduction **7**

2 Probabilities and Probability Distributions **8**

2.1 Examples of Probability Distributions 12

2.1.1 Binomial Distribution 12

2.1.2 Poisson Distribution 13

2.1.3 Gaussian Distribution 15

2.2 Bayes' Formula 18

2.2.1 Data Assimilation 19

3 Stochastic Processes **21**

3.1 Wiener-Khinchin Theorem 25

3.2 Markov Processes. Chapman-Kolmogorov Equation 27

3.3 Master Equation 33

3.4 Differential Chapman-Kolmogorov Equation 35

3.4.1 Drift Term 37

3.4.2 Diffusion Term 39

4 Fokker-Planck Equation **40**

4.1 The Rayleigh Particle 40

4.2 The Macroscopic Equation 44

4.3 Brownian Motion in an External Potential 45

4.3.1 Markov Approximation for $X(t)$ and Weak Noise 46

4.3.2 Fluctuations in a Steady State 50

4.3.3 Bistable Systems: Escape from a Metastable State 51

4.3.4 First Passage Times: Second Approach 56

4.4 The Backward Fokker-Planck Equation 59

5 Langevin Equation **60**

5.1 Relation between Langevin Equation and Fokker-Planck Equation 63

5.1.1 Linear Langevin Equation 63

5.1.2 Nonlinear Langevin Equation 64

5.2 Mathematical Considerations: Ito vs Stratonovich 68

5.3 Ito Stochastic Integrals 71

5.4 Ito Stochastic Differential Equation 75
5.4.1 Ito's Formula 75
5.4.2 Solvability by Variable Transform 78
5.4.3 Fokker-Planck Equation from the Ito Stochastic Differential Equation 80
5.5 Stratonovich's Stochastic Differential Equation 81
5.6 Colored Noise: Ornstein-Uhlenbeck and its White-Noise Limit 83

6 Stochastic Resonance **86**
6.1 Examples 90
6.1.1 Ring Laser [22] 91
6.1.2 Mechanoreceptors in Crayfish[12] 93
6.1.3 Tactile Sensation[9] 94

7 Sketch of Numerical Methods for Stochastic Differential Equations **96**
7.1 Strong Approximation 96
7.1.1 Euler-Maruyama Scheme 96
7.1.2 Milstein Scheme 98
7.1.3 Implicit Schemes 103
7.2 Weak Approximation 104

8 Projects **106**

References

[1] L. Arnold. *Stochastic differential equations: theory and applications*. Wiley, New York, 1974. 519.2 A756sX.

[2] R. D. Astumian. Thermodynamics and kinetics of a brownian motor. *Science*, 276(5314):917–922, May 1997.

[3] R. D. Astumian. Paradoxical games and a minimal model for a brownian motor. *Am. J. Phys. paper for 442*, 73(2):178–183, February 2005.

[4] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani. Stochastic resonance in climatic-change. *Tellus*, 34(1):10–16, September 1982.

[5] L. Bocquet. High friction limit of the kramers equation: the multiple time-scale approach. *Am. J. Phys. paper for 442*, 65(2):140–144, February 1997.

[6] N. Brunel and V. Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.*, 11(7):1621–1671, October 1999.

[7] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, 15(1):1, 1943.

[8] W. T. Coffey, Y. P. Kalmykov, and J. T. Waldron. *The Langevin equation*. World Scientific, 1996.

[9] J. J. Collins, T. T. Imhoff, and P. Grigg. Noise-enhanced tactile sensation. *Nature*, 383(6603):770–770, October 1996.

[10] S. Denisov, P. Hanggi, and J. L. Mateos. Ac-driven brownian motors: a fokker-planck treatment. *Am. J. Phys. paper for 442*, 77(7):602–606, July 2009.

[11] D. Dimitropoulos and B. Jalali. Stochastic differential equation approach for waves in a random medium. *Phys. Rev. E*, 79(3):036606, March 2009.

[12] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss. Noise enhancement of information-transfer in crayfish mechanoreceptors by stochastic resonance. *Nature*, 365(6444):337–340, September 1993.

[13] M. O. Ernst and M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, Jan 2002.

[14] L. Gammaitoni, P. Hanggi, P. Jung, and F. Marchesoni. Stochastic resonance. *Rev. Mod. Phys.*, 70(1):223–287, January 1998.

[15] C. W. Gardiner. *Handbook of stochastic methods*. Springer, 1990.

[16] J. C. Hull. *Options, futures and other derivatives*. Prentice-Hall, 2006.

[17] P.E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer, Berlin, 1992.

[18] R. Kuske. Probability densities for noisy delay bifurcations. *J. Stat. Phys.*, 96(3-4):797–816, August 1999.

[19] R. Kuske, L. F. Gordillo, and P. Greenwood. Sustained oscillations via coherence resonance in sir. *J. Theor. Biol.*, 245(3):459–469, April 2007.

[20] B. J. Mason. Towards understanding and prediction of climatic variations. *Q. J. Roy. Met. Soc.*, 102(433):473–498, 1976.

[21] B. McNamara and K. Wiesenfeld. Theory of stochastic resonance. *Phys. Rev. A*, 39:4854, 1989.

[22] B. Mcnamara, K. Wiesenfeld, and R. Roy. Observation of stochastic resonance in a ring laser. *Phys. Rev. Lett.*, 60(25):2626–2629, June 1988.

[23] A. S. Pikovsky and J. Kurths. Coherence resonance in a noise-driven excitable system. *Phys. Rev. Lett.*, 78(5):775–778, February 1997.

[24] S. Reinker, Y. X. Li, and R. Kuske. Noise-induced coherence and network oscillations in a reduced bursting model. *Bull. Math. Biol.*, 68(6):1401–1427, August 2006.

[25] Z. Schuss. *Theory and applications of stochastic differential equations*. Wiley, 1980.

[26] E. Simonotto, M. Riani, C. Seife, M. Roberts, J. Twitty, and F. Moss. Visual perception of stochastic resonance. *Phys. Rev. Lett.*, 78(6):1186–1189, February 1997.

[27] S. H. Strogatz and R. E. Mirollo. Stability of incoherence in a population of coupled oscillators. *J. Stat. Phys.*, 63(3-4):613–635, May 1991.

[28] N. G. van Kampen. *Stochastic processes in physics and chemistry*. North-Holland, 1992.

[29] N.G. van Kampen. Fluctuation in unstable systems. In E. Tirapegui and Villarroel D., editors, *Instabilities and Nonequilibrium Structures*. Reidel, 1987.

[30] Y. Weiss, E. P. Simoncelli, and E. H. Adelson. Motion illusions as optimal percepts. *Nature Neurosc.*, 5(6):598–604, June 2002.

[31] P. Wilmott, W. Howison, and J. Dewynne. *The mathematics of financial derivatives*. Cambridge, 1995.

Index

absorbing boundary, 57
adiabatic approximation, 47
Arrhenius factor, 56

Backward Euler, 104
backward Fokker-Planck equation, 56, 59
Backward Milstein, 104
Black-Scholes equation, 106

climate, 86
Coherence resonance, 106
Colored Noise, 83
Crayfish, 93
cumulants, 11, 12, 60

Doob's theorem, 30
Dynkin equation, 54, 58

escape rate, 56
Euler-Maruyama Schem, 96

fluctuation-dissipation theorem, 42
Fokker-Planck equation, 40
Forward Euler, 104

identically distributed, 17
Implicit Schemes, 103

Kramers' equation, 46
Kuramoto model, 106

Lindeberg condition, 34

Markov process, 23
mean first passage time, 52
mean-squared limit, 70
metastable, 52
Milankovich cycles, 86
Milstein Scheme, 98
Moment Generating Function, 10
multiplicative noise, 64

network oscillations, 106
non-anticipating function, 72

order parameter, 52
Order-2 Weak Taylor Scheme, 105
Ornstein-Uhlenbeck process, 29, 42, 83

phase transition, 52
pitchfork bifurcation, 65
Poisson process, 29

Rare Events, 15
RC-circuit, 62
reaction coordinate, 52
reaction rate, 56
realization, 97
reflecting boundary, 57
Ring Laser, 91

strong approximation, 96

transition probability, 28

Weak Approximation, 104
weak approximation, 96
Wiener process, 28, 43

1 Introduction

Many systems have unpredictable components:

- weather
- stock market
- spreading of diseases
- motion of gas molecules
- molecular conformations: opening and closing of ion channels
- emission and absortion of photons: photo multipliers, photo receptors
- ...

Causes of unpredictability:

- many interacting agents (particles)
- chaotic dynamics, even in systems with few degrees of freedom
- quantum mechanics: only probabilities for transitions can be given

Probabilistical description of such systems:
as example, consider Brownian motion of a large particle in a thermal bath of smaller particles. We are interested in

- the evolution of the mean, variance, or higher moments of the position or velocity of the large particle
- possibly even the evolution of full probability distributions for the position or the velocity
- sample trajectories of the large particle

Approaches:

- (evolution) equations for the probability distributions:
  Chapman-Kolmogorov equation, master equation, Fokker-Planck equation
- differential equations with stochastic quantities: Langevin equation
  we will need to make sense of the stochastic differential equation (Ito vs. Stratonovich)

Recommended books:

- N. G. van Kampen. Stochastic processes in physics and chemistry. North-Holland, 1992 (on reserve in the library)
- C. W. Gardiner. Handbook of stochastic methods. Springer, 1990 (on reserve in the library)
- D. Henderson and P. Plaschko. Stochastic differential equations in science and engineering.
- N.G. van Kampen. Fluctuation in unstable systems. In E. Tirapegui and Villarroel D., editors, Instabilities and Nonequilibrium Structures. Reidel, 1987 (a scanned version of this paper is on the class web site
`https://courses.northwestern.edu/webapps/login`
under 2009FALL ES APPM 442-1 *Stochastic Differential Equations*
- Other useful references: [25, 1, 17, 8]

2 Probabilities and Probability Distributions

We consider stochastic events:

- tossing a coin: event ω is head or tail
- throwing a die
- position of molecule in a fluid
- emission of an electron in a (old-fashioned) TV-tube
- absorption of a photon
- ...

Consider sets $A_i \subseteq \Omega$ of events, Ω being all possible events.

Define a probability $P(A_i)$

1. $P(A_i) \geq 0$ for all A_i
2. $P(\Omega) = 1$
3. The probability of mutually exclusive events is additive:
if the A_i, $i = 1, 2, 3, \ldots$, form a countable (but possibly infinite) collection of sets with $A_i \cap A_j = \emptyset$ for $i \neq j$ then

$$P(A_i \cup A_i \cup \ldots \cup A_k) = P(A_i) + P(A_j) + \ldots + P(A_k)$$

e.g. for a die $A_1 = \{1\}$ and $A_2 = \{2, 3\}$ are mutually exclusive and the probability to get 1 or 2 or 3 is $P(A_1) + P(A_2)$. However, for $A_i = \{\text{even numbers }\}$ and $A_2 = \{2\}$ one has $P(A_1 \cup A_2) = P(A_1) \leq P(A_1) + P(A_2)$.

Note:

- from the definitions one obtains directly

$$P(\Omega\backslash A) = 1 - P(A) \qquad P(\emptyset) = 0.$$

- to make the connection with experiments one needs to measure probabilities: use the relative frequency of events as an approximation for the probability

For continuous variables x one has to define events a bit more carefully;

- if the event is defined that x has a specific value $x = x_0$, then the probability of this event is always 0
- need to define the event has x being in some interval:
  $P(x \in (x_0, x_0 + \Delta x))$ is a meaningful quantity and can be non-zero
  typically $P(x \in (x_0, x_0 + \Delta x)) = p(x_0)\Delta x + \mathcal{O}(\Delta x^2)$

Notation:

- A random variable is a function $X(\omega)$ of the event ω
- Often we need not explicitly denote the event and can simply write for the random variable X.
  In a given realization X has the value x.

Define

- Joint probability

$$P(A \cap B) = P(\omega \in A \text{ and } \omega \in B)$$

- Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

  or

$$P(A \cap B) = P(A|B)\, P(B)$$

- Statistical independence:
  Two events or sets of events are independent of each other if $P(A|B)$ does not depend on B: $P(A|B) = P(A)$ and analogously $P(B|A) = P(B)$
  this implies

$$P(A \cap B) = P(A)\, P(B)$$

  For multiple events A_i statistical independence requires that *for all possible combinations* $A_j \cap A_k \cap ... \cap A_m$ the joint probability factorizes

$$P(A_j \cap A_k \cap ... \cap A_m) = P(A_j)P(A_k) \ldots P(A_m)$$

- Mean of a random variable X

- X takes on only the discrete values X_i, $i = 1..N$

$$\mu(X) \equiv \langle X \rangle \equiv \sum_{i=1}^{N} x_i P(x_i)$$

- $X \in \mathbb{R}$

$$\mu(X) \equiv \langle X \rangle = \int x\, P(x)\, dx$$

- Higher moments

$$\langle X^n \rangle = \int x^n\ P(x) dx$$

- Variance of X is the mean of the square of the deviation from the mean

$$\sigma(X)^2 = \langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2$$

  $\sigma(X)$ is the standard deviation of X.

- Covariance between different random variables X_i and X_j

$$C_{ij} = \langle (X_i - \langle X_i \rangle)\,(X_j - \langle X_j \rangle) \rangle = \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle$$

  measures how correlated the two variables are (in terms of their deviations from their respective means:

  - For $C_{ij} > 0$ the random variables X_i and X_j are mostly on the same side of their means: correlated
  - For $C_{ij} < 0$ the random variables X_i and X_j are mostly on opposite sides of their means: anti-correlated
  - For $C_{ij} \sim 0$ the random variables X_i and X_j are just as often on opposite sides of their means as on the same side: uncorrelated

Moment Generating Function[1]

To get mean and variance make use of a *generating function*.

In general the moment generating function or characteristic function is defined as

$$\phi(\mathbf{s}, t) = \langle e^{i\mathbf{s}\cdot\mathbf{X}} \rangle = \int dx_1 \dots dx_n p(\mathbf{x}, t) e^{i\mathbf{s}\cdot\mathbf{x}}$$

The generating function amounts to a Fourier transform of the probability distribution.

The moments can then be expressed elegantly as

$$\langle \prod_{i=1}^{n} X_i^{m_i} \rangle = \left[\prod_{i=1}^{n} \left(-i \frac{\partial}{\partial s_i} \right)^{m_i} \phi(\mathbf{s}, t) \right] \Bigg|_{\mathbf{s}=0}$$

[1] [15, Chapter 2.6 and 2.7]

Conversely, using the Taylor expansion of the exponential $e^{i\mathbf{s}\cdot\mathbf{X}} = \prod_{j=1}^{n} e^{is_j X_j}$ the generating function can be expressed in terms of the moments

$$\phi(\mathbf{s},t) = \sum_{m_1=0}^{\infty} \cdots \sum_{m_n=0}^{\infty} \frac{(is_1)^{m_1}}{m_1!} \cdots \frac{(is_n)^{m_n}}{m_n!} \langle X_1^{m_1} \ldots X_n^{m_n} \rangle$$

The probability distribution can be obtained by an inverse Fourier transformation

$$P(\mathbf{x},t) = \frac{1}{(2\pi)^n} \int ds_1 \ldots ds_n e^{-i\mathbf{s}\cdot\mathbf{x}} \phi(\mathbf{s},t)$$

For independent distributions $P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i)$ one gets

$$\phi(s_1, s_2, \ldots, s_n) = \prod_{i=1}^{n} \phi(s_i)$$

It is sometimes also useful to introduce the *cumulants* $\langle\langle X_1^{m_1} \ldots X_n^{m_n} \rangle\rangle$, which are defined via

$$\ln\left(\phi(\mathbf{s},t)\right) = \sum_{m_1=0}^{\infty} \cdots \sum_{m_n=0}^{\infty} \frac{(is_1)^{m_1}}{m_1!} \cdots \frac{(is_n)^{m_n}}{m_n!} \langle\langle X_1^{m_1} \ldots X_n^{m_n} \rangle\rangle \tag{1}$$

Example:

$n = 2:$

$$\phi(s_1, s_2, t) = 1 + is_1 \langle X_1 \rangle + is_2 \langle X_2 \rangle - \frac{1}{2} s_1^2 \langle X_1^2 \rangle - \frac{1}{2} s_2^2 \langle X_2^2 \rangle - s_1 s_2 \langle X_1 X_2 \rangle + \mathcal{O}(s^3)$$

$$\begin{aligned}
\ln\left(\phi(s_1, s_2, t)\right) &= is_1 \langle X_1 \rangle + is_2 \langle X_2 \rangle - \frac{1}{2} s_1^2 \langle X_1^2 \rangle - \frac{1}{2} s_2^2 \langle X_2^2 \rangle - s_1 s_2 \langle X_1 X_2 \rangle + \mathcal{O}(s^3) + \\
&\quad + \left(-\frac{1}{2}\right) \left(is_1 \langle X_1 \rangle + is_2 \langle X_2 \rangle - \frac{1}{2} s_1^2 \langle X_1^2 \rangle - \frac{1}{2} s_2^2 \langle X_2^2 \rangle - s_1 s_2 \langle X_1 X_2 \rangle + \mathcal{O}(s^3) \right)^2 + \ldots \\
&= is_1 \langle X_1 \rangle + is_2 \langle X_2 \rangle \\
&\quad -\frac{1}{2} \left(s_1^2 \langle X_1^2 \rangle + s_2^2 \langle X_2^2 \rangle + 2 s_1 s_2 \langle X_1 X_2 \rangle - s_1^2 \langle X_1 \rangle^2 - s_2^2 \langle X_2 \rangle^2 - 2 s_1 s_2 \langle X_1 \rangle \langle X_2 \rangle \right) \\
&= is_1 \langle X_1 \rangle + is_2 \langle X_2 \rangle \\
&\quad -\frac{1}{2} \left(s_1^2 \left(\langle X_1^2 \rangle - \langle X_1 \rangle^2 \right) + s_2^2 \left(\langle X_2^2 \rangle - \langle X_2 \rangle^2 + 2 s_1 s_2 \left(\langle X_1 X_2 \rangle - \langle X_1 \rangle \langle X_2 \rangle \right) \right) \right)
\end{aligned}$$

Thus

$$\begin{aligned}
\langle\langle X_i \rangle\rangle &= \langle X_i \rangle \\
\langle\langle X_i^2 \rangle\rangle &= \langle X_i^2 \rangle - \langle X_i \rangle^2 \\
\langle\langle X_1 X_2 \rangle\rangle &= \langle X_1 X_2 \rangle - \langle X_1 \rangle \langle X_2 \rangle
\end{aligned}$$

Notes:

- Considering two random variables ($n = 2$): statistical independence is equivalent with
  - $\phi(s_1, s_2, t) = \phi_1(s_1, t)\phi_2(s_2, t)$
  - all moments factorize:
    $$\langle X_1^{m_1} X_2^{m_2} \rangle = \langle X_1^{m_1} \rangle \langle X_2^{m_2} \rangle$$
  - all cumulants $\langle\langle X_1^{m_1} X_2^{m_2} \rangle\rangle$ vanish when both m_1 and m_2 are nonzero: the expansion of $\ln(\phi(s_1, s_2, t)) = \ln(\phi_1(s_1, t)) + \ln(\phi_2(s_2, t))$ has no terms in which s_1 and s_2 appear at the same time.

- Expressions for general cumulants is cumbersome to obtain[15, Chapter 2.7]. For example

$$\langle\langle X_1 X_2 X_3 \rangle\rangle = \langle X_1 X_2 X_3 \rangle - \langle X_1 X_2 \rangle \langle X_3 \rangle - \langle X_2 X_3 \rangle \langle X_1 \rangle - \langle X_1 X_3 \rangle \langle X_2 \rangle + 2 \langle X_1 \rangle \langle X_2 \rangle \langle X_3 \rangle$$

2.1 Examples of Probability Distributions

2.1.1 Binomial Distribution

Consider an urn with N balls, m of which are red and $N - m$ are white. What is the probability $P(k; n)$ to retrieve exactly k red balls in n draws if each ball is placed back into the urn before the next is drawn? Assume that the probability is the same for all balls.

- probability to draw *any* read ball is given by the fraction of red balls among the balls:
$$\frac{m}{N}$$

- probability to draw k red balls in k specific draws
$$\left(\frac{m}{N}\right)^k$$
since successive drawings are independent of each other

- analogously: probability to draw $n - k$ white balls
$$\left(\frac{N-m}{N}\right)^{n-k}$$

- it does not matter in which draws the red or white balls are actually drawn: in how many ways can the k red draws be distributed over the total number of n draws? Think of picking k integers from $1 \ldots n$ without replacing them,
$$n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$$
possibilities. It does not matter in which order the integers are picked: do not distinguish those $k!$ possibilities
$$\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{(n-k)!\,k!}$$

- thus
$$P(k;n) = \left(\begin{array}{c} n \\ k \end{array} \right) \left(\frac{m}{N}\right)^k \left(\frac{N-m}{N}\right)^{n-k} = \left(\begin{array}{c} n \\ k \end{array} \right) p^k q^{n-k}$$
with $p = \frac{m}{N}$ probability for red ball and $q = 1-p$ probability for white ball.

Normalization
$$\sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) p^k q^{n-k} \underbrace{=}_{\text{binomial theorem}} (p+q)^n = 1$$

Mean value
$$\langle k \rangle = \sum_{k=0}^{n} k \left(\begin{array}{c} n \\ k \end{array} \right) p^k q^{n-k} = p\frac{\partial}{\partial p} \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) p^k q^{n-k} = p\frac{\partial}{\partial p} (p+q)^n = pn$$

Variance

$$\begin{aligned}
\sigma^2 &= \langle k^2 \rangle - \langle k \rangle^2 = p\frac{\partial}{\partial p}\left(p\frac{\partial}{\partial p}\sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) p^k q^{n-k}\right) - (pn)^2 = \\
&= p\frac{\partial}{\partial p}\left(p\frac{\partial}{\partial p}(p+q)^n\right) - (pn)^2 = \\
&= p\frac{\partial}{\partial p}\left(np(p+q)^{n-1}\right) - (pn)^2 = \\
&= pn(p+q)^{n-1} + np^2(n-1)(p+q)^{n-2} - (pn)^2 = \\
&= pn - p^2 n = npq
\end{aligned}$$

2.1.2 Poisson Distribution

Consider

- current in a cathode tube (old TV screen): consists of individual electrons that are emitted from the cathode at arbitrary, uncorrelated times
- customers approaching a bank teller, their arrival times are presumably also uncorrelated

What is the distribution $P(n,t)$ of the number n of electrons/customers that have arrived up to a time t?

The probability that the number increases from n to $n+1$ during a time interval Δt is proportional to Δt
$$P(n \to n+1 \text{ during } \Delta t) = \lambda \Delta t$$

Then we have

$$P(n, t+\Delta t) = \underbrace{\lambda\Delta t}_{\text{probability for } n \text{ to increase}} P(n-1,t) + \underbrace{(1-\lambda\Delta t)}_{\text{probbility for } n \text{ not to increase}} P(n,t)$$

For small Δt

$$\frac{d}{dt}P(n,t) = \lambda\left(P(n-1,t) - P(n,t)\right) \tag{2}$$

For this differential-difference equation we need initial conditions in t and in n (first order in time and in n). Assume that initially no electrons have arrived:

$$P(-1,t) = 0 \qquad P(n,0) = \delta_{n,0}$$

How to solve this equation?

Introduce the generating function for $P(n,t)$

$$G(u,t) = \sum_{n=0}^{\infty} u^n P(n,t)$$

using (2) one gets

$$\begin{aligned}
\frac{d}{dt}G(u,t) &= \lambda \sum_{n=0}^{\infty}\left(u^n P(n-1,t) - u^n P(n,t)\right) = \\
&= \lambda\left\{u\sum_{n=1}^{\infty} u^{n-1}P(n-1,t) - \sum_{n=0}^{\infty} u^n P(n,t)\right\} \\
&\underbrace{=}_{\text{relabel } m=n-1} \lambda\left(u-1\right)\sum_{n=0}^{\infty} u^n P(n,t) = \lambda(u-1)G(u,t)
\end{aligned}$$

Thus

$$G(u,t) = G(u,0)e^{\lambda(u-1)t}$$

Using $P(n,0) = 0$ for $n \geq 1$ we get k

$$G(u,0) = P(0,0) = 1$$

and

$$G(u,t) = e^{\lambda ut}e^{-\lambda t} = e^{-\lambda t}\sum_{n=0}^{\infty}\frac{1}{n!}\left(\lambda t\right)^n u^n$$

Therefore by comparing powers in u we get the Poisson distribution

$$P(n,t) = \frac{\left(\lambda t\right)^n}{n!}e^{-\lambda t}$$

For the Poisson distribution we had the generating function in terms of $u \equiv e^{is}$

$$\phi(s,t) = G(e^{is},t) = e^{\lambda ut}e^{-\lambda t} = e^{\lambda te^{is}}e^{-\lambda t}$$

- mean value

$$\langle n\rangle = \left[-i\frac{\partial}{\partial s}\phi(s,t)\right]\Big|_{s=0} = -ie^{-\lambda t}e^{\lambda te^{is}}i\lambda te^{is}\Big|_{s=0} = \lambda t$$

  as expected, λ is the mean rate at which electrons/customers arrive.
  The mean can also easily be calculated directly from $P(n,t)$

$$\sum_{n=0}^{\infty} nP(n,t) = e^{-\lambda t}\sum_{n=1}^{\infty}\frac{\left(\lambda t\right)^n}{(n-1)!} = \left(\lambda t\right)e^{-\lambda t}\sum_{n=0}^{\infty}\frac{\left(\lambda t\right)^n}{n!} = \lambda t$$

- variance

$$\begin{aligned}\langle n^2\rangle &= \left[\left(-i\frac{\partial}{\partial s}\right)^2\phi(s,t)\right]\Bigg|_{\mathbf{s=0}} = -i\frac{\partial}{\partial s}\left(-ie^{-\lambda t}e^{\lambda te^{is}}i\lambda te^{is}\right)\Bigg|_{s=0} = \\ &= -i\lambda te^{-\lambda t}\, e^{\lambda te^{is}}e^{is}\left(i\lambda te^{is}+i\right)\Big|_{s=0} = \lambda t\left(\lambda t+1\right)\end{aligned}$$

  thus

$$\sigma^2 = \lambda t$$

  Determining the variance by evaluating the sum is more difficult.

- The variability of a distribution can be characterized by the Fano factor

$$F = \frac{\sigma^2}{\mu}$$

  The Poisson distribution has $F = 1$, i.e. the variance is equal to the mean, corresponding to quite high variability.

Rare Events: Poisson Distribution from Binomial Distribution

The Poisson distribution arises, e.g., from the binomial distribution for events with low probability.

Consider the binomial distribution for $p \ll 1$ and many draws n, $np = \mu = \mathcal{O}(1)$,

$$\begin{aligned}P(k;n) &= \binom{n}{k} p^k(1-p)^{n-k} = \\ &= \frac{n!}{k!(n-k)!}\left(\frac{\mu}{n}\right)^k\left(1-\frac{\mu}{n}\right)^{n-k} = \\ &= \frac{1}{k!}\underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_{\to 1}\mu^k\underbrace{\left(1-\frac{\mu}{n}\right)^n}_{\to e^{-\mu}}\underbrace{\left(1-\frac{\mu}{n}\right)^{-k}}_{\to 1}\end{aligned}$$

Letting $n \to \infty$ with μ and k fixed one gets

$$P(k;n) \to \frac{1}{k!}\mu^k e^{-\mu}$$

2.1.3 Gaussian Distribution

The Gaussian or normal distribution is given by

$$p(\mathbf{x}) = \frac{1}{\sqrt{2\pi det(\mathbf{\Sigma})}}e^{-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^t\mathbf{\Sigma}^{-1}(\mathbf{x}-\bar{\mathbf{x}})}$$

where the mean is given by

$$\langle \mathbf{X}\rangle = \bar{\mathbf{x}}$$

and the correlation matrix by

$$\langle (X_i - \bar{x}_i)(X_j - \bar{x}_j) \rangle = \mathbf{\Sigma}_{ij}$$

The Gaussian distribution appears very widely because of the *Central Limit Theorem*: the sum of many independent, (almost) arbitrarily distributed random variables is Gaussian distributed.

More specifically[2]:

Consider many independent random variables X_i, $i = 1 \dots n$, with distributions $p_i(x)$. For simplicity assume vanishing mean. Their sum

$$S_N = \sum_{i=1}^{N} X_i$$

is also a random variable.

The distributions of X_i can be almost arbitrary as long as their variances are finite

$$var\{X_i\} = \langle (X_i - \langle X_i \rangle)^2 \rangle = \sigma_i^2$$

and they satisfy the Lindeberg condition

$$\lim_{N\to\infty} \left[\frac{1}{\sigma_N^2} \sum_{i=1}^{N} \int_{|x|>\tau\sigma_n} x^2 p_i(x) dx \right] \to 0$$

for any $\tau > 0$, where

$$\sigma_N^2 = \sum_{i=1}^{N} \sigma_i^2$$

Then in the limit $N \to \infty$

$$P(\frac{S_N}{\sigma_N}) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{S_N^2}{\sigma_N^2}}$$

i.e. is Gaussian (normal) distributed and its variance is σ_n^2.

Note:

- The Lindeberg condition requires that large values $|x|$ are sufficiently rare: the distributions $p_i(x)$ need to decay sufficiently fast for $|x| \to \infty$.
  If the σ_i are larger than some constant for all i (e.g. if all distributions $p_i(x)$ are equal) one has $\sigma_N \to \infty$ for $N \to \infty$.
  For $\tau = 0$ the fraction in the Lindeberg condition would always be 1.
  For $\tau > 0$ increasingy larger portions of the integral are omitted as $N \to \infty$, since σ_N increases with N. For the fraction to go to 0 for arbitrarily small $\tau > 0$ requires that the contributions to σ_i^2 are not dominated by large values of $|x|$.

[2]one could formulate this statement also in terms of the rescaled sum as in the simple proof below.

- The Cauchy or Lorentzian distribution

$$P(x) = \frac{a}{\pi}\frac{1}{a^2+x^2}$$

has diverging variance and it does not satisfy the Lindeberg condition.

For *identically distributed* variables the proof for the central limit theorem is relatively easy:

For simplicity assume all Y_i have zero mean and variance 1. For the generating function $\phi_Y(s)$ we have

$$\begin{aligned}
\langle Y\rangle &= -i\,\frac{d}{ds}\phi_Y(s)\Big|_{s=0} = 0 \\
\langle Y^2\rangle &= -\,\frac{d^2}{ds^2}\phi_Y(s)\Big|_{s=0} = 1
\end{aligned}$$

This gives the Taylor expansion for small s

$$\phi_Y(s) = 1 - \frac{1}{2}s^2 + \mathcal{O}(s^3)$$

Consider the rescaled sum of the Y_i

$$S_N = \sum_{i=1}^{N}\frac{1}{\sqrt{N}}Y_i$$

The generating function ϕ_{S_N} is then

$$\begin{aligned}
\phi_{S_N}(s) &= \langle e^{isS_N}\rangle = \langle e^{is\sum_{i=1}^{N}\frac{1}{\sqrt{N}}Y_i}\rangle = \\
&= \langle\prod_{i=1}^{N} e^{i\frac{s}{\sqrt{N}}Y_i}\rangle = \prod_{i=1}^{N}\phi_Y(\frac{s}{\sqrt{N}}) = \\
&= \left(1-\frac{1}{2}\frac{s^2}{N}+\mathcal{O}(\frac{s^3}{N^{\frac{3}{2}}})\right)^N \to e^{-\frac{1}{2}s^2} \quad \text{for } N\to\infty
\end{aligned}$$

which is the generating function of the Gaussian (normal) distribution with vanishing mean and unit variance, $N(0,1)$.

Gaussian Distribution as the Limit of Poisson Distribution

For large mean values the Poisson distribution approaches a Gaussian distribution.

The Poisson distribution (omit the t from our previous result)

$$P(k) = \frac{\lambda^k}{k!}e^{-\lambda}$$

has mean $\mu = \lambda$ and variance $\sigma^2 = \lambda$.

Expand k around the mean and scale the deviation with the standard deviation

$$k = \mu + \sigma x = \lambda + \sqrt{\lambda}x = \lambda\left(1+\frac{x}{\sqrt{\lambda}}\right) \qquad \text{with } x = \mathcal{O}(1)$$

Use Stirling's formula for $k!$ for large k, keeping more terms than usually,

$$\ln k! = k\ln k - k + \frac{1}{2}\ln k + \frac{1}{2}\ln 2\pi + \mathcal{O}(\frac{1}{k})$$

insert in Poisson distribution

$$\begin{aligned}
\ln P(k) &= k\ln\lambda - \lambda - \left(k\ln k - k + \frac{1}{2}\ln k + \frac{1}{2}\ln 2\pi + \mathcal{O}(\frac{1}{k})\right) = \\
&= -\lambda\left(1+\frac{x}{\sqrt{\lambda}}\right)\ln\left(1+\frac{x}{\sqrt{\lambda}}\right) + \lambda\frac{x}{\sqrt{\lambda}} - \frac{1}{2}\ln\lambda - \frac{1}{2}\ln\left(1+\frac{x}{\sqrt{\lambda}}\right) - \frac{1}{2}\ln 2\pi + \mathcal{O}(\frac{1}{\lambda})
\end{aligned}$$

Use

$$\ln\left(1+\frac{x}{\sqrt{\lambda}}\right) = \frac{x}{\sqrt{\lambda}} - \frac{1}{2}\frac{x^2}{\lambda} + \mathcal{O}(\frac{1}{\lambda^{\frac{3}{2}}})$$

get

$$\begin{aligned}
\ln P(k) &= -\lambda\left(\frac{x}{\sqrt{\lambda}} - \frac{1}{2}\frac{x^2}{\lambda} + \frac{x^2}{\lambda} - \frac{1}{2}\frac{x^3}{\lambda^{\frac{3}{2}}} + \mathcal{O}(\frac{1}{\lambda^{\frac{3}{2}}})\right) + \lambda\frac{x}{\sqrt{\lambda}} - \frac{1}{2}\ln\lambda - \mathcal{O}(\frac{1}{\lambda^{\frac{1}{2}}}) - \frac{1}{2}\ln 2\pi + \mathcal{O}(\frac{1}{\lambda}) = \\
&= -\frac{1}{2}x^2 - \frac{1}{2}\ln\lambda - \frac{1}{2}\ln 2\pi
\end{aligned}$$

$$P(k) \to \frac{1}{\sqrt{2\pi\lambda}}e^{-\frac{1}{2}x^2} = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\frac{(k-\mu)^2}{\sigma^2}} \qquad \text{with } \mu = \lambda \qquad \sigma^2 = \lambda$$

Similarly: close to the mean the binomial distribution is well approximated by the Gaussian distribution.
Using

$$k = np + \sqrt{n}x = \mu + \sqrt{n}x$$

for large n with p and q fixed one gets

$$P(k;n) = \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k} \to \frac{1}{\sqrt{2\pi npq}}e^{-\frac{1}{2}\frac{(k-np)^2}{(npq)^2}} \qquad \text{for } n \to \infty$$

2.2 Bayes' Formula

Write the joint probability in two different ways

$$P(A\cap B) = P(A|B)\,P(B) = P(B|A)\,P(A)$$

Then one gets

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)} \propto P(A|B)\,P(B)$$

This simple formula is useful in many applications.

2.2.1 Data Assimilation

Consider weather forecasting. Need to combine various uncertain pieces of information

- data from measurements
- the computational model itself is not certain since the physical model for the 'weather' is quite incomplete
- parameters in the model used for the prediction are uncertain
- initial data for the model: not all initial values are actually known, e.g., the algorithm needs the temperature at many more locations than can be measured

Three steps are iteratively performed

1. predict data at a later time t_{n+1}, e.g. temperature at a given location, using the model that is based on data at an earlier time t_n.
2. measure the values of the data at time t_{n+1}.
3. combine the two pieces of information to obtain better estimates for the model parameters and the initial conditions.
4. repeat

An essential question is: *how to combine the various pieces of information?*

In this simple discussion we lump the uncertainties in the model parameters together with the uncertainty of the initial conditions.

We would like to know the true temperature x given a set of measured values y_i making use of a predictive model. We won't be able to get a single value, instead we will aim for the distribution $P(x|\mathbf{y})$. Due to the measurement errors we obtain from the measurements at best the distribution $P(y_i|x)$.

To get the distribution $P(x|\mathbf{y})$ of interest we can use Bayes formula

$$\underbrace{P(x|\mathbf{y})}_{\text{posterior distribution}} = \frac{1}{P(\mathbf{y})} \underbrace{P(x|\mathbf{y})}_{\text{likelihood}} \underbrace{P(x)}_{\text{prior distribution}}$$

Note:

- The prior distribution is the distribution for the true value that we assume (know) before the measurement is done ('a priori')
- The likelihood is the probability for a given true value given the measurements.
- The posterior distribution is the distribution we obtain after we incorporate the measurement into our expectations ('a posteriori')

Specifically, assume that the measurements y_i all have normally distributed measurement errors, which may have different degrees of precision,

$$P(y_i|x) = N(x, \sigma_i) \equiv \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y_i - x)^2}{2\sigma_i^2}}$$

Assuming all n measurements are independent of each other we get

$$P(\mathbf{y}|x) = \prod_{i=1}^{n} P(y_i|x)$$

For simplicity assume all measurements have the same error, $\sigma_i = \sigma_y$.

$$P(\mathbf{y}|x) = \frac{1}{\sqrt{2\pi\sigma_y^2}^n} e^{-\frac{1}{2\sigma_y^2}\sum(y_i - x)^2}$$

For the prior distribution we take the distribution $P(x)$ we obtain from the model - based on previous data - before the new measurements are incorporated. $P(x)$ expresses the uncertainties in the previous data and model parameters. Assume for simplicity also a normal distribution

$$P(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x - \mu_{prior})^2}{2\sigma_x^2}}$$

Thus

$$\begin{aligned} P(x|\mathbf{y}) &= \frac{1}{P(\mathbf{y})} \frac{1}{\sqrt{2\pi\sigma_y^2}^n} e^{-\frac{1}{2\sigma_y^2}\sum_i(y_i - x)^2} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x - \mu_{prior})^2}{2\sigma_x^2}} = \\ &= \frac{1}{P(\mathbf{y})} \frac{1}{\sqrt{2\pi\sigma_y^2}^n} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{1}{2}\left[\left(\frac{n}{\sigma_y^2} + \frac{1}{\sigma_x^2}\right)x^2 - 2\left(\sum_i \frac{y_i}{\sigma_y^2} + \frac{\mu_{prior}}{\sigma_x^2}\right)x + \left[\sum_i \frac{y_i^2}{\sigma_y^2} + \frac{\mu_{prior}^2}{\sigma_x^2}\right]\right]} \end{aligned}$$

which is a normal distribution

$$N(\mu_{post}, \sigma_{post}) = \frac{1}{\sqrt{2\pi\sigma_{post}^2}} e^{-\frac{1}{2}\frac{(x - \mu_{post})^2}{\sigma_{post}^2}}$$

with variance

$$\sigma_{post}^2 = \frac{1}{\frac{n}{\sigma_y^2} + \frac{1}{\sigma_x^2}} \equiv (1 - K)\,\sigma_x^2$$

and mean value

$$\mu_{post} = \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 + n\sigma_y^2} \left(n\frac{\mu_y}{\sigma_y^2} + \frac{\mu_{prior}}{\sigma_x^2}\right) \equiv \mu_{prior} + K\,(\mu_y - \mu_{prior})$$

where $\mu_y = \frac{1}{n}\sum_i y_i$ and the gain K is given by

$$K = \frac{n\sigma_x^2}{\sigma_y^2 + n\sigma_x^2}$$

Notes:

- the measurements shift the mean of the posterior towards the mean of the measurements
  - the shift is proportional to the gain K, which increases with
    * the number n of measurements
    * the precision of the measurements relative to the uncertainty in the prior distribution (model uncertainty)
- including the measurements always reduces the uncertainty of the posterior distribution relative to that of the prior distribution
- depending on the amount (n) and the quality (σ_y^2) of the data the measurements or the model result dominate the result from this assimilation
- in human sensory processing, a similar approach is often also used to understand the perception that results from integrating different types of sensory information, like touch+vision [13], auditory+vision (e.g. localization of a sound source), or vision+vision [30, see homework], weighing each information with its reliability/precision.

3 Stochastic Processes

Given a set of random variables $\{X_n|n = 1\dots N\}$ with probability distribution $P(X_1, X_2, \dots, X_n\}$ we can define a stochastic process

$$Y(t) = f(X_1, \dots, X_n, t)$$

For each fixed realization of the random variables, $X_i = x_i$, we get a function

$$y(t) = f(x_1, \dots, x_n, t)$$

called sample function or realization of the process.

Examples:

1. Consider

   $$Y(t) = X_1 \cos t + X_2 \sin t$$

   with $\langle X_i \rangle = 0$ and $\langle X_i X_j \rangle = \delta_{ij}\sigma^2$. Then

   $$\begin{aligned} \langle Y(t) \rangle &= \langle X_1 \cos t + X_2 \sin t \rangle = 0 \\ \langle Y(t_1) Y(t_2) \rangle &= \langle (X_1 \cos t_1 + X_2 \sin t_1)(X_1 \cos t_2 + X_2 \sin t_2) \rangle = \\ &= \langle X_1^2 \rangle \cos t_1 \cos t_2 + \langle X_2^2 \rangle \sin t_1 \sin t_2 + \langle X_1 X_2 \rangle (\cos t_1 \sin t_2 + \sin t_1 \cos t_2) = \\ &= \sigma^2 \cos (t_2 - t_1) \end{aligned}$$

2. A discontinuous stochastic process (cf. HW 1)

   $$Y(t) = X_n \qquad \text{for } \xi + n < t < \xi + n + 1$$

   where $\{X_n|n = 1, 2, 3\dots\}$ is an infinite set of identically distributed independent stochastic variables and ξ is another independent stochastic variable that is uniformly distributed in $0 < \xi < 1$.

At each fixed time $Y(t)$ is a random variable with a certain probability distribution.

We can characterize a random process using the probability distributions

$$P_1(y_1, t_1) \qquad \text{probability for } Y \text{ to have value } y_1 \text{ at } t_1$$

$$P_1(y_1, t_1; y_2, t_2) \qquad \text{probability for } Y \text{ to have value } y_1 \text{ at } t_1 \text{ and } y_2 \text{ at } t_2$$

$$\ldots$$

$$P_n(y_1, t_1, ; \ldots ; y, t_n)$$

Note:

- the process can be discrete or continuous in time
- if y are continuous variables the probabilities are given by $P_n(y_1, t_1; \ldots ; y_n t_n) \prod_{i=1}^{n} dy_i$.

The probabilities need to satisfy the conditions

1. $P_n \geq 0$
2. $P_n(y_1, t_1; \ldots ; y_n, t_n)$ is symmetric under interchange of any indices (they are simply joint probabilities), e.g. $P(y_1, t_1; y_2, t_2) = P(y_2, t_2; y_1, t_1)$
3. $$\int P_n(y_1, t_1; \ldots ; y_{n-1}, t_{n-1}; y_n, t_n) dy_n = P_{n-1}(y_1, t_1; \ldots ; y_{n-1}, t_{n-1})$$
   since y_n has to take on *some* value. If that value does not matter then it also does not matter when one would measure that value $\Rightarrow$ no dependence on t_n.
4. $\int P_1(y_1, t_1) dy_1 = 1$

Notes:

- any set of functions that satisfy these conditions define a stochastic process (proof by Kolmogorov, see [28] p.62)
- One need not specify all P_n: P_n contains all the information about all P_m with $m < n$ $\Rightarrow$ any finite number of P_n need not be specified
- for a stationary process P_n depends only on the time differences $t_n - t_m$, but not on the times t_i themselves.

Means, correlations, etc. are defined as for random variables

- mean
$$\langle Y(t) \rangle = \int y \, P(y, t) dy$$

- correlation
$$C_{ij}(t_1,t_2) = \langle (Y_i(t_1) - \langle Y_i(t_1)\rangle)\,(Y_j(t_2) - \langle Y_j(t_2)\rangle)\rangle$$
Of particular importance is typically how fast (if at all) the correlation decays for $|t_1 - t_2| \to \infty$.
The diagonal elements $C_{ii}(t_1,t_2)$ give the respective autocorrelations.
The off-diagonal elements of $C_{ij}(t_1,t_2)$ give cross-correlations between Y_i and Y_j

Examples

1. Sequence of independent coin tosses: $y = \pm 1$

$$\begin{array}{rcll} P_1(y,t) & = & P(y) & \\ P_2(y_1,t_1;y_2,t_2) & \underbrace{=}_{\text{independence}} & P(y_1)\,P(y_2) & \text{for } t_1 \neq t_2 \end{array}$$

all joint probabilities factorize in this way.

$$\begin{aligned}
\langle Y\rangle &= -P(-1) + P(1) \\
C(t_1,t_2) &= \sum_{y_1=\pm 1}\sum_{y_2=\pm 1} y_1 y_2 \underbrace{P_2(y_1,t_1;y_2,t_2)}_{P(y_1)P(y_2)} - \left(\sum_{y_1} y_1 P_1(y_1)\right)\left(\sum_{y_2} y_2 P_1(y_2)\right) = \\
&= 0 \qquad \text{for } t_1 \neq t_2 \\
C(t,t) &= \sum_{y_1=\pm 1}\sum_{y_2=\pm 1} y_1 y_2 \underbrace{P_2(y_1,t;y_2,t)}_{P(y_1)\delta_{y_1 y_2}} - \left(\sum_{y_1} y_1 P_1(y_1)\right)\left(\sum_{y_2} y_2 P_1(y_2)\right) = \\
&= \langle Y^2\rangle - \langle Y\rangle^2 = P(+1) + P(-1) - (P(+1) - P(-1))^2
\end{aligned}$$

since the coin tosses are independent only at different times.

2. Markov process
different events are not independent, but the probability of an event depends only on the immediately previous event
Introduce conditional probability to obtain y_n at t_n given that $y = y_i$ at all previous times:
$$P_{1|n-1}(y_n t_n | y_{n-1}, t_{n-1}; \ldots; y_1, t_1)$$
For a Markov process one has
$$P_{1|n-1}(y_n t_n | y_{n-1}, t_{n-1}; \ldots; y_1, t_1) = P_{1|n-1}(y_n t_n | y_{n-1}, t_{n-1})$$
independent of earlier events.
To characterize a Markov process completely we therefore need only $P_1(y,t)$ and $P_2(y_1,t_1;y_2,t_2)$ or $P_{1|1}(y_2,t_2|y_1,t_1)$.
For example:
$$\begin{array}{rcl} P_3(y_1,t_1;y_2,t_2;y_3,t_3) & = & P_{1|2}(y_3,t_3|y_2,t_2;y_1,t_1)\,P_2(y_1,t_1;y_2,t_2) = \\ & \underbrace{=}_{\text{Markov process}} & P_{1|1}(y_3,t_3|y,t_2)\,P_{1|1}(y_2,t_2|y_1,t_1)P_1(y_1,t_1) \end{array}$$

since - as for random variables - we have the connection between joint and conditional probabilities

$$P_2(y_1,t_1;y_2,t_2) = P_{1|1}(y_2,t_2|y_1,t_1)\,P_1(y_1,t_1)$$

Note:

- Markov processes are somewhat similar to 1^{st}-order differential equations (one can actually write a $1^{st}-$order ODE as a Markov process).

3. Gambler tossing coins: capital Y
   gambler wins with probability $p_+ \equiv p$: $y \to y+1$
   gambler looses with probability $p_- = 1-p$: $y \to y-1$
   The capital at step n depends only on the capital at step $n-1$: Markov process for the capital
   Consider one such step

$$P_1(y,n) = P_1(y-1,n-1)\,p + P_1(y+1,n-1)\;(1-p)$$

   Thus

$$P_{1|1}(y_n,n|y_{n-1},n-1) = p\delta_{y_n,y_{n-1}+1} + (1-p)\delta_{y_n,y_{n-1}-1}$$

   and

$$P_2(y_n,n;y_{n-1},n-1) = P_1(y_{n-1},n-1)\left[p\delta_{y_n,y_{n-1}+1} + (1-p)\,\delta_{y_n,y_{n-1}-1}\right]$$

   Note:

   - y_n could also be considered the position of a one-dimensional random walker that takes only discrete steps of a fixed width.

4. Brownian motion on a short time scale
   The Brownian particle is hit by small particles with a random force at random time:

   - because of inertia v_{n+1} depends on v_n
     Moreover, the impact of the collision depends on the velocity of the Brownian particle:
     the probability that it is hit by small particles is somewhat higher ahead of it than behind it $\Rightarrow$ the particle slows down. No random walk in velocity space.
     But the impact depends only the current velocity, not previous velocities:
     $V(t)$ is a Markov process
   - the process for the position of the particle is not Markovian:
     x_{n+1} depends on x_n and on $v_n \approx \frac{1}{\Delta t}(x_n - x_{n-1})$

$$P(x_3,t_3|x_2,t_2;x_1,t_1) \neq P(x_3,t_3|x_2,t_2).$$

     Intuitively, the particle is more likely to continue in its direction of motion before the collision than to change the direction significantly.
   - **Notes:**
     - in the deterministic case the position satisfies 2^{nd}-order ODE

- over larger time scales and many collisions the particle 'forgets' its velocity and with it the information about the position before the current one.

Moment Generating Functional

Analogous to the generating functional for a random variable one has

$$\phi\{s(t)\} = \langle e^{i\int_{-\infty}^{+\infty} s(t)Y(t)dt}\rangle$$

which can be thought of as the extension of a vector-valued random variable Y_i to infinitely many components,

$$\phi(s_i) = \langle e^{i\sum_i s_i Y_i}\rangle$$

Again, it generates all moments of $Y(t)$ using Taylor expansion

$$\begin{aligned}\phi\{s(t)\} &= \langle 1 + i\int s(t)Y(t)dt + \frac{1}{2}\left(\int s(t)Y(t)dt\right)^2 + \ldots\rangle = \\ &= \sum_{i=0}^{\infty}\frac{i^n}{n!}\int s(t_1)\ldots s(t_n)\langle Y(t_1)\ldots Y(t_n)\rangle dt_1\ldots dt_n\end{aligned}$$

thus

$$\langle Y(t_1)\ldots Y(t_m)\rangle = (-i)^m \left.\frac{\delta^m \phi\{s(t)\}}{\delta s(t_1)\ldots\delta s(t_m)}\right|_{s(t)=0}$$

analogously to

$$\langle Y_1\ldots Y_m\rangle = (-i)^m \left.\frac{\partial^m \phi(s_1,\ldots,s_n)}{\delta s_1\ldots\delta s_m}\right|_{s_i=0}$$

3.1 Wiener-Khinchin Theorem[3]

For stationary stochastic processes there is a connection between the Fourier spectrum of the process and its autocorrelation. Often it is easier to determine the Fourier spectrum than to measure correlations directly. Then the Wiener-Khinchin theorem is useful.

Consider a stationary stochastic process $Y(t)$. Without loss of generality assume $Y(t)$ has vanishing mean.

For each realization of the underlying random variables the function $y(t) = f(X;t)$ can be Fourier transformed

$$y(t) = \sum_{n=-\infty}^{\infty} A_n e^{i\frac{2\pi}{T}nt}$$

with

$$A_n = \frac{1}{T}\int_0^T e^{-\frac{2\pi}{T}nt}y(t)dt$$

Replacing $y(t)$ by the stochastic process $Y(t)$ the coefficients A_n become random variables.

[3]cf. chapter III.3 in [28] ([28]III.3)

Using the Parseval identity

$$\sum_{n=-\infty}^{\infty} |A_n|^2 = \frac{1}{T}\int_0^T (Y(t))^2\, dt$$

one gets for the averages

$$\sum_{n=-\infty}^{\infty} \langle |A_n|^2\rangle = \frac{1}{T}\int_0^T \langle (Y(t))^2\rangle dt \qquad \underbrace{=}_{\text{stationary process}} \qquad \langle Y^2\rangle$$

One can think of $\langle Y^2\rangle$ as the total 'energy' in the random process.
The coefficient $\langle |A_n|^2\rangle$ gives the 'energy' for the corresponding frequencies $\omega = \frac{2\pi}{T}n$.
For very long averaging interval T the frequencies lie densely, $\Delta\omega = 2\pi/T$, and one can introduce a spectral density

$$S(\omega)\Delta\omega = \underbrace{2}_{Y\in\mathbb{R}:\ A_{-n}=A_n^*} \sum_{n=\frac{T}{2\pi}(\omega-\frac{1}{2}\Delta\omega)}^{\frac{T}{2\pi}(\omega+\frac{1}{2}\Delta\omega)} |A_n|^2$$

Evaluate $\langle |A_n|^2\rangle \equiv \langle A_n A_n^*\rangle$

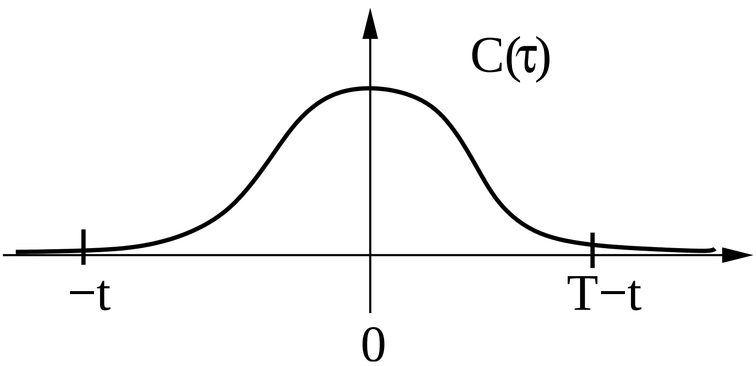
$$\langle |A_n|^2\rangle = \frac{1}{T^2}\int_0^T dt \int_0^T dt'\, e^{-i\frac{2\pi}{T}nt'+\frac{2\pi}{T}t}\langle Y(t')Y(t)\rangle$$

Since $\langle Y(t)\rangle = 0$ we have

$$\langle Y(t')Y(t)\rangle = C(t',t) \quad \underbrace{=}_{\text{stationary}} \quad C(t'-t)$$

$$\langle |A_n|^2\rangle \underbrace{=}_{\tau=t'-t} \frac{1}{T^2}\int_0^T dt \int_{-t}^{T-t} d\tau\, e^{-i\frac{2\pi}{T}n\tau}C(\tau)$$

Assume the autocorrelation decays on a time scale τ_c.



The τ-integration interval $(-t, T-t)$ is shifted across $C(\tau)$ by the t-integration. Except for $-t \gtrsim -\tau_c$ and $T-t \lesssim \tau_c$ the whole support of $C(\tau)$ is covered by the τ-integral. For $T \gg \tau_c$ those contributions become negligible and the τ-integral can be extended to $(-\infty,\infty)$

$$\langle |A_n|^2\rangle = \frac{1}{T^2}\int_0^T dt \int_{-\infty}^{\infty} d\tau\, e^{-i\frac{2\pi}{T}n\tau}C(\tau) = \frac{1}{T}\int_{-\infty}^{\infty} d\tau\, e^{-i\frac{2\pi}{T}n\tau}C(\tau)$$

If $C(\tau)$ is smooth (no characteristic delay time at which something happens) then $\langle |A_n|^2\rangle$ depends smoothly on n and n can be replaced by the frequency ω

$$\langle |A_n|^2\rangle = \frac{1}{T}\int_{-\infty}^{\infty} d\tau\, e^{-i\omega\tau} C(\tau) \qquad \text{with } \omega = \frac{2\pi}{T} n$$

To get to the spectral density need to count the number of modes in the frequency interval $\Delta\omega$

$$\Delta n = \frac{\Delta\omega}{\frac{2\pi}{T}}$$

Therefore

$$2 \sum_{n=\frac{T}{2\pi}(\omega-\frac{1}{2}\Delta\omega)}^{\frac{T}{2\pi}(\omega+\frac{1}{2}\Delta\omega)} \langle |A_n|^2\rangle = 2\frac{\Delta\omega}{\frac{2\pi}{T}}\frac{1}{T}\int_{-\infty}^{\infty} d\tau\, e^{-i\omega\tau} C(\tau)$$

and we have the Wiener-Khinchin theorem

$$S(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} d\tau\, e^{-i\omega\tau} C(\tau) = 2C(\omega)$$

The Fourier transform of the autocorrelation function is essentially given by the spectral density.

3.2 Markov Processes. Chapman-Kolmogorov Equation[4]

A Markov process is completely determined by $P_1(y,t)$ and $P_{1|1}(y_2,t_2|y_1,t_1)$. Are there any conditions on these probabilities?

For any probability distributions one has to require

$$\begin{aligned} P_1(y_2,t_2) &= \int dy_1 P_2(y_1,t_1;y_2,t_2) = \\ &= \int dy_1 P_{1|1}(y_2,t_2|y_1,t_1) P_1(y_1,t_1) \end{aligned} \tag{3}$$

and

$$\begin{aligned} P_{1|1}(y_3,t_2|y_1,t_1) &= \int dy_2 P_{2|1}(y_3,t_3;y_2,t_2|y_1,t_1) = \\ &= \int dy_2 P_{1|2}(y_3,t_3|y_2,t_2;y_1,t_1) P_{1|1}(y_2,t_2|y_1,t_1) \end{aligned}$$

For Markov processes the second condition becomes for $t_1 < t_2 < t_3$

$$P_{1|1}(y_3,t_2|y_1,t_1) = \int dy_2 P_{1|1}(y_3,t_3|y_2,t_2) P_{1|1}(y_2,t_2|y_1,t_1) \tag{4}$$

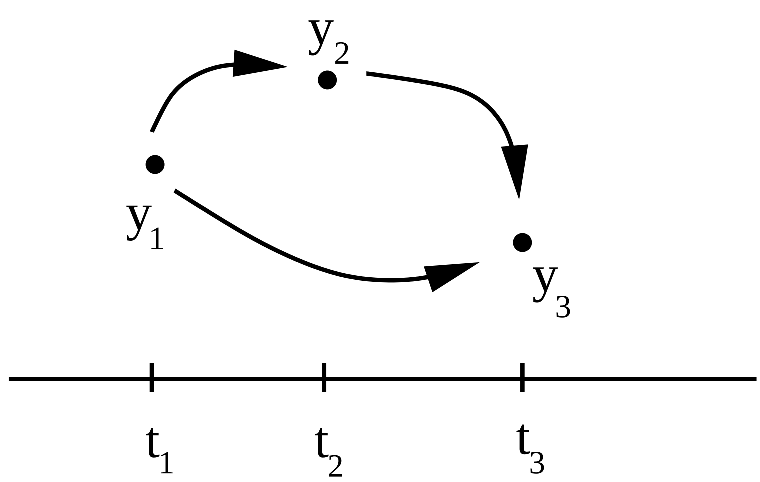
For stationary Markov processes only time differences matter

$$P_{1|1}(y_2,t_2|y_1,t_1) = T_{t_2-t_1}(y_{2,}|y_1)$$

[4] [28]IV.2

with $T_\tau(y_2|y_1)$ giving the *transition probability* of the process. It satisfies

$$T_{\tau+\tau'}(y_3|y_1) = \int T_\tau(y_3|y_2)T_{\tau'}(y_2|y_1)dy_2 \tag{5}$$



Notes:

- Equation (3) determines in a straightforward manner the unconditional probability P_1 at the later time t_2 from P_1 at the earlier time t_1. Thus, we need to specify only $P_1(y, t_0)$ for some t_0.
- Equation (4) is the Chapman-Kolmogorov equation.
- The Chapman-Kolmogovor equation is a nonlinear integral equation for the conditional probability $P_{1|1}$. It is difficult to solve. It is easier to deal with in the form of a differential equation (see later (14) below).
- A Markov process is completely determined by solutions to (3,4) and any such solution defines a Markov process.

Examples

1. **Wiener process**
   is defined by

$$\begin{aligned} P_{1|1}(y_2, t_2|y_1, t_1) &= \frac{1}{\sqrt{2\pi\,(t_2 - t_1)}}e^{-\frac{1}{2}\frac{(y_2-y_1)^2}{t_2-t_1}} \\ P_1(y, 0) &= \delta(y) \end{aligned} \tag{6}$$

   which yields with (3)

$$P_1(y, t) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{1}{2}\frac{y^2}{t}} \tag{7}$$

   - The Wiener process is non-stationary (P_1 depends explicitly on t).
   - It satisfies the Chapman-Kolmogorov equation (check!).

- It was introduced to model the position of a particle undergoing Brownian motion or a random walker for long times (when the particle 'forgets' its previous velocity):
  $P_1(y,t,)$ exhibits the diffusive spread expected from a random walker:
  distance that the walker can move during a given time interval:

$$y_2 - y_1 \propto \sqrt{t_2 - t_1}$$

  - path is continuous
  - for short times distance can be large, no maximal velocity
    path need not be differentiable

$$\frac{y_2 - y_1}{\Delta t} = \mathcal{O}(\Delta t^{-\frac{1}{2}}) \to \infty \qquad \text{for} \qquad \Delta t \to 0$$

  - for large times it grows sublinearly: no ballistic motion

2. **Poisson process**
   Consider $Y(t)$ taking on only integer values $n = 0, 1, 2, \ldots$.
   Define a Markov process via

$$\begin{aligned} P_1(n,0) &= \delta_{n,0} \\ P_{1|1}(n_2,t_2|n_1,t_1) &= \frac{(t_2-t_1)^{n_2-n_1}}{(n_2-n_1)!} e^{-(t_2-t_1)} \qquad \text{for } n_2 > n_1 \end{aligned}$$

   - Each sample function $y(t)$ consists of unit steps which occur at random times. The number of steps (increments $n_2 - n_1$) between times t_2 and t_1 are Poisson-distributed.
   - The process is not stationary.

3. **Ornstein-Uhlenbeck process**
   It is a stationary Markov process given by

$$\begin{aligned} P_1(y_1) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_1^2} \\ T_\tau(y_2|y_1) &= \frac{1}{\sqrt{2\pi\left(1-e^{-2\tau}\right)}} e^{-\frac{1}{2}\frac{\left(y_2-y_1e^{-\tau}\right)^2}{\left(1-e^{-2\tau}\right)}} \end{aligned}$$

   Notes:

   - The Ornstein-Uhlenbeck process was introduced to model the velocity of a Brownian particle ($v = y$)
   - Gaussian distribution
   - Vanishing mean
   - In contrast to the position of the Brownian particle, the velocity does not spread to large values, damping pushes it back to 0. That makes the process stationary.

- Correlation function decays exponentially

$$\begin{aligned}
C(\tau) &= \int y_3 y_1 P_2(y_3, y_1)\, dy_3 dy_1 = \\
&= \int y_3 y_1 T_\tau(y_3|y_1) P_1(y_1)\, dy_3 dy_1 \\
&= \ldots \text{ complete the square etc. } \ldots \\
&= e^{-\tau}
\end{aligned}$$

Over time the particle is hit so often that it 'forgets' its previous velocity: the correlation vanishes for $\tau \gg 1$

4. **Doob's theorem**:

i) *The Ornstein-Uhlenbeck process is the only stationary, Gaussian Markov process.*

Consider first a general stationary Gaussian process $Y(t)$ (possibly non-Markovian) after shifting the mean and rescaling Y the distribution $P_1(y)$ can be written as

$$P_1(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

The process is stationary

$$P_{1|1}(y_2, t_2|y_1, t_1) \equiv T_\tau(y_2|y_1) = d e^{-\frac{1}{2}\left(a y_1^2 + 2b y_1 y_2 + c y_2^2\right)}$$

$T_\tau(y_2|y_1)$ must be normalized (the particle has to end up *somewhere*)

$$\int T_\tau(y_2|y_1) dy_2 = 1$$

$$\int d\, e^{-\frac{1}{2}\left(a y_1^2 + 2b y_1 y_2 + c y_2^2\right)} dy_2 = d\, e^{-\frac{1}{2}a y_1^2} \int e^{-\frac{1}{2}c\left(y_2 + \frac{b}{c}y_1\right)^2} dy_2\ e^{\frac{1}{2}\frac{b^2}{c}y_1^2}$$

requiring

$$a = \frac{b^2}{c} \qquad d = \sqrt{\frac{c}{2\pi}}$$

Also, the consistency condition (3) requires

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_2^2} &= \int T_\tau(y_2|y_1) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_1^2} dy_1 = \\
1 &= d\, e^{\frac{1}{2}c y_2^2} \int e^{-\frac{1}{2}\left(a y_1^2 + 2b y_1 y_2 - y_1^2\right)} e^{-\frac{1}{2}y_1^2} dy_1
\end{aligned}$$

which leads to

$$b^2 = c(c-1)$$

Express remaining unknown c in terms of the correlation function $C(\tau)$

$$\begin{aligned}
C(\tau) &= \int y_2 y_1 T_\tau(y_2|y_1) P_1(y_1)\, dy_2 dy_1 = \\
&= \frac{d}{\sqrt{2\pi}} \int \quad y_2 y_1\, e^{-\frac{1}{2}\left(a y_1^2 + 2b y_1 y_2 - y_1^2\right)} e^{-\frac{1}{2}c y_1^2} dy_1 dy_2
\end{aligned}$$

yielding

$$c = \frac{1}{1 - C(\tau)^2}$$

This results in

$$T_\tau(y_2|y_1) = \frac{1}{\sqrt{2\pi\,(1 - C(\tau)^2)}} e^{-\frac{1}{2}\frac{(y_2 - C(\tau)y_1)^2}{1 - C(\tau)^2}} \tag{8}$$

Note:

- Any stationary Gaussian process can be written in the form (8). Different stationary Gaussian processes differ only in their correlation functions and their means.

Now consider a Markovian stationary Gaussian process.
To make use of the Markov property introduce $t_3 > t_2$

$$C(t_3 - t_1) = \int y_3 y_1 T_{t_3 - t_1}(y_3|y_1) P_1(y_1)\, dy_3 dy_1$$

and use

$$T_{t_3 - t_1}(y_3|y_1) = \int T_{t_3 - t_2}(y_3|y_2) T_{t_2 - t_1}(y_2|y_1) dy_2$$

$$C(t_3 - t_1) = \int y_3 dy_3 \int T_{\tau'}(y_3|y_2) T_\tau(y_2|y_1) dy_2 \int y_1 P_1(y_1) dy_1$$

Evaluating the triple integral yields (done in maple or mathematica) gives

$$C(t_3 - t_1) = C(t_3 - t_2) \cdot C(t_2 - t_1)$$

To solve this functional relation take derivative with respect to t_2 which is an arbitrary intermediate time

$$0 = -C'(t_3 - t_2) \cdot C(t_2 - t_1) + C(t_3 - t_2) \cdot C'(t_2 - t_1)$$

implying

$$\frac{C'(t_3 - t_2)}{C(t_3 - t_2)} = \frac{C'(t_2 - t_1)}{C(t_2 - t_1)}$$

For fixed t_2 the left-hand side depends on t_3 while the right-hand side depends on t_1
$\Rightarrow$ both sides have to be constant,

$$\frac{C'(\tau)}{C(\tau)} = const.$$

$$C(\tau) = e^{-\gamma\tau}$$

with γ the correlation time. Thus the process is the Ornstein-Uhlenbeck process.

ii) *Any stationary Gaussian process with exponential correlation function is Markovian.*

Show first that a Gaussian process is completely determined by its mean $\langle Y(t)\rangle$ and its correlation $C(t_1, t_2)$.
Consider the generating function of the process, which completely determines the process,

$$\phi\{s(t)\} = \langle e^{i\int_{-\infty}^{+\infty} s(t)Y(t)dt}\rangle.$$

Allowing t to be a continuous variable the expectation value is given by a path integral,

$$\langle e^{i\int_{-\infty}^{+\infty} s(t)Y(t)dt} \rangle = \int P\{Y(t)\} e^{i\int_{-\infty}^{+\infty} s(t)Y(t)dt} dY(t).$$

Consider first discrete times t_j, $j = 1..N$ and consider then the limit $N \to \infty$.
The process is Gaussian

$$P_N(y_N, t_N; \ldots; y_1, t_1) = \frac{1}{\mathcal{N}} e^{-\frac{1}{2}\sum_{ij=1}^{N}(y_i-\bar{y}_i)A_{ij}(y_j-\bar{y}_j)}$$

where ȳ is the mean, A is the inverse of the correlation matrix, and $\mathcal{N}$ is a normalization.
Generating function

$$\phi(s_1, \ldots, s_N) = \frac{1}{\mathcal{N}} \int dy_1 \ldots dy_n e^{i\sum_j s_j y_j} e^{-\frac{1}{2}\sum_{ij=1}^{N}(y_i-\bar{y}_i)A_{ij}(y_j-\bar{y}_j)}$$

The multiple Gaussian integrals will give again Gaussians with the mean appearing in the imaginary term: as illustration, consider the scalar case $N = 1$

$$\int e^{isy-\frac{1}{2}A(y-\bar{y})^2} dy = \int e^{-\frac{1}{2}A\left(y-\bar{y}-\frac{1}{A}is\right)^2} dy \;\cdot\; e^{+\frac{1}{2}A\bar{y}\frac{1}{A}is+\frac{1}{2}A\left(\frac{is}{A}\right)^2}$$

Thus

$$\phi(s_1, \ldots, s_N) \propto e^{i\sum_j s_j\alpha_j - \frac{1}{2}\sum_{ij} s_i\beta_{ij}s_j}$$

In the limit $N \to \infty$ the sums turn into integrals

$$\phi\{s(t)\} = \frac{1}{\mathcal{N}'} e^{i\int s(t)\alpha(t)dt - \frac{1}{2}\int s(t)\beta(t,t')s(t')dtdt'}$$

where $\beta(t, t')$ can be assumed to be symmetric, $\beta(t, t') = \beta(t', t)$, since the anti-symmetric part does not contribute to the integral.
Also, since $\phi\{s(t) = 0\} = 1$ we have $\mathcal{N}'$=1.
Thus, the Gaussian process is determined completely by the functions $\alpha(t)$ and $\beta(t, t')$.

We know

$$\begin{aligned} \left.\frac{\delta\phi\{s(t)\}}{\delta s(t_a)}\right|_{s(t)=0} &= -i\langle Y(t_a)\rangle \\ \left.\frac{\delta^2\phi\{s(t)\}}{\delta s(t_a)\delta s(t_b)}\right|_{s(t)=0} &= -\langle Y(t_a)Y(t_b)\rangle = -C(t_a, t_b) - \langle Y(t_a)\rangle\langle Y(t_b)\rangle \end{aligned}$$

Evaluating the derivatives we get

$$\left.\left(i\alpha(t_a) - \frac{1}{2}\left(\int \beta(t_a, t')s(t')dt' + \int s(t)\beta(t, t_a)dt\right)\right)\phi\{s(t)\}\right|_{s(t)=0} = -i\langle Y(t)\rangle$$

$$i\alpha(t) = -i\langle Y(t)\rangle \tag{9}$$

and

$$\left.\left(-\alpha(t_a)\alpha(t_b) - \frac{1}{2}\left[\beta(t_a, t_b) + \beta(t_b, t_a)\right]\right)\phi\{s(t)\}\right|_{s(t)=0} = -\alpha(t_a)\alpha(t_b) - \beta(t_a, t_b)$$

leading to

$$\beta(t_a,t_b) = -\langle Y(t_a)\rangle\langle Y(t_b)\rangle - (-C(t_a,t_b) - \langle Y(t_a)\rangle\langle Y(t_b)\rangle) = C(t_a,t_b) \tag{10}$$

Thus, for a Gaussian process the mean and the correlation function are sufficient to determine $\phi\{s(t)\}$ completely. Since $\phi\{s(t)\}$ generates all probability distributions, such a process is completely determined by its mean and correlation function.

Now: if the process is stationary and the correlation function of the process is exponential, $C(t_a - t_b) = e^{-(t_a - t_b)}$, it has the same correlations as the Ornstein-Uhlenbeck process and therefore must be identical to the Ornstein-Uhlenbeck process, which is Markovian.

Consequence:

- For a Gaussian stationary process exponential correlations imply that the process is Markovian.

Note:

- Inserting the exponentially decaying correlation function into (8) is not sufficient to show that the process is the Ornstein-Uhlenbeck process; we need to show that the process does not differ from the Ornstein-Uhlenbeck process in higher probability distribution functions like $P_{1|2}(y_3,t_3|y_2,t_2;y_1,t_1)$. We did this by showing that a Gaussian process is completely determined by its mean and its correlation.

3.3 Master Equation[5]

The Chapman-Kolmogorov equation is a *nonlinear* integral equation for the transition probability and difficult to solve even in the stationary case

$$T_{\tau+\tau'}(y_3|y_1) = \int T_\tau(y_3|y_2)T_{\tau'}(y_2|y_1)dy_2$$

One can simplify it if the transition probability has a simple form for small durations τ

$$T_\tau(y_2|y_1) = \tau \underbrace{W(y_2|y_1)}_{\text{transition rate}} + \underbrace{(1-a_0(y_1)\,\tau)}_{\text{probability that } y \text{ stays at } y_1} \delta(y_2 - y_1) + o(\tau) \tag{11}$$

($o(\tau)$ are terms that are smaller than $\mathcal{O}(\tau)$, they could in principle be bigger than $\mathcal{O}(\tau^2)$, e.g. $\mathcal{O}(\tau \ln \tau)$)

Note:

- $W(y_2|y_1)$ is the probability to jump from y_1 to y_2 during time τ: it is a transition rate

[5] [28]V.1

Total probability is conserved

$$1 = \int T_\tau(y_2|y_1)dy_2 = 1 - a_0(y_1)\tau + \tau \int W(y_2|y_1)dy_2$$

$$a_0(y_1) = \int W(y_2|y_1)dy_2 \tag{12}$$

Note:

- the higher the probability to jump away to any location y_2 the lower the probability to stay at y_1
- the ansatz (11) implies discontinuous solutions $y(t)$:
  even for arbitrarily small time intervals τ the rate $W(y_2|y_1)$ with which the particle jumps a finite distance $y_2 - y_1$ is finite (and independent of the duration)
- to get a continuous Markov process one needs in general (another Lindeberg condition)

  $$\lim_{\tau\to 0} \frac{1}{\tau} \int_{|y_2-y_1|>\epsilon} T_\tau(y_2|y_1)dy_2 = 0 \qquad \text{for any } \epsilon > 0$$

  here one gets

  $$\int_{|y_2-y_1|>\epsilon} W(y_2|y_1)dy_2 \quad = \quad 0 \qquad \text{for any } \epsilon > 0$$

  implying

  $$W(y_2|y_1) = 0$$

Now use this expansion in the Chapman-Kolmogorov equation for small τ'

$$\begin{aligned} T_{\tau+\tau'}(y_3|y_1) &= \int \delta(y_3 - y_2)\left(1 - a_0(y_2)\tau'\right) T_\tau(y_2|y_1) + \tau' W(y_3|y_2)T_\tau(y_2|y_1)dy_2 = \\ &= \left(1 - a_0(y_3)\tau'\right) T_\tau(y_3|y_1) + \tau' \int W(y_3|y_2)T_\tau(y_2|y_1)dy_2 \end{aligned}$$

for $\tau' \to 0$

$$\frac{d}{d\tau}T_\tau(y_3|y_1) = -a_0(y_3)T_\tau(y_3|y_1) + \int W(y_3|y_2)T_\tau(y_2|y_1)dy_2$$

replacing $a_0(y_3)$ by using the conservation of probability (12) we get the master equation

$$\frac{d}{d\tau}T_\tau(y_3|y_1) = \int \underbrace{W(y_3|y_2)T_\tau(y_2|y_1)}_{\text{rate of jumps into } y_3} - \underbrace{W(y_2|y_3)T_\tau(y_3|y_1)}_{\text{rate of jumps out of } y_3} \quad dy_2 \tag{13}$$

Notes:

- the master equation describes the increase and decrease of the probability of a state due to ‘fluxes’ into and out of that state (somewhat similar to a continuity equation)
- a similar master equation can also be derived for the non-stationary case

- for many systems the transition rates $W(y_2|y_1)$ can be determined (measured, modeled) more easily than the full transition probabilities $T_\tau(y_2|y_1)$. One can then use the master equation to determine $T_\tau(y_2|y_1)$.

For systems with discrete states one gets the analogous master equation

$$\frac{d}{d\tau}T_\tau(n_3|n_1) = \sum_{n_2} \underbrace{W(n_3|n_2)T_\tau(n_2|n_1)}_{\text{rate of jumps into } y_3} - \underbrace{W(n_2|n_3)T_\tau(n_3|n_1)}_{\text{rate of jumps out of } y_3}$$

3.4 Differential Chapman-Kolmogorov Equation[6]

Consider different limit to include also continuous processes. Focus again on stationary case.

Assume for all $\epsilon > 0$:

1. jumps: for $|y_2 - y_1| \geq \epsilon$

$$\lim_{\tau\to 0} \frac{1}{\tau}T_\tau(y_2|y_1) = \underbrace{W(y_2|y_1)}_{\text{jump rate}}$$

2. continous component: mean distance moved during small time intervals

$$\lim_{\tau\to 0} \frac{1}{\tau}\int_{|y_2-y_1|<\epsilon} (y_2 - y_1)\, T_\tau(y_2|y_1)dy_2 = A(y_1) + \mathcal{O}(\epsilon)$$

3. continous component: mean spread during small time intervals

$$\lim_{\tau\to 0} \frac{1}{\tau}\int_{|y_2-y_1|<\epsilon} (y_2 - y_1)^2\, T_\tau(y_2|y_1)dy_2 = B(y_1) + \mathcal{O}(\epsilon)$$

Calculate the evolution of $\langle f(y)\rangle$ for a smooth $f(y)$. Since $f(y)$ is arbitrary that expecation value will generate equation for $T_\tau(y_2|y_1)$

$$\langle f(y)\rangle = \int f(y)T_t(y|y_1)dy$$

This average depends on t through $T_t(y|y_1)$ and it also depends on the initial value y_1

$$\frac{d}{dt}\langle f(y\rangle = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int f(y_3)\left[T_{t+\Delta t}(y_3|y_1) - T_t(y_3|y_1)\right]dy_3.$$

Using the Chapman-Kolmogorov equation we get

$$\frac{d}{dt}\langle f(y\rangle = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int f(y_3)\left[\int T_{\Delta t}(y_3|y_2)T_t(y_2|y_1)dy_2 - T_t(y_3|y_1)\right]dy_3.$$

Make now use of the assumptions for small Δt:

[6][15]3.4

- there is a contribution from jumps via $W(y_3|y_1)$
- smooth part of $T_{\Delta t}(y_3|y_2)$ is peaked near $|y_3 - y_2| \ll 1$

Expand $f(y_3)$

$$f(y_3) = f(y_2) + (y_3 - y_2)\, f'(y_2) + \frac{1}{2}\,(y_3 - y_2)\, f''(y_2) + \dots$$

Separate integral to use the expansion only in the domain $|y_3 - y_2| < \epsilon$,

$$\begin{aligned}
\int dy_3 \int dy_2 \dots \quad = \quad & \int\int_{|y_3-y_2|<\epsilon} dy_3 dy_2 f(y_2) T_{\Delta t}(y_3|y_2) T_t(y_2|y_1) + \\
& + \int\int_{|y_3-y_2|<\epsilon} dy_3 dy_2 \left((y_3 - y_2)\, f'(y_2) + \frac{1}{2}\,(y_3 - y_2)\, f''(y_2) \right) T_{\Delta t}(y_3|y_2) T_t(y_2|y_1) + \\
& + \int\int_{|y_3-y_2|\geq\epsilon} dy_3 dy_2 f(y_3) T_{\Delta t}(y_3|y_2) T_t(y_2|y_1) - \\
& - \int dy_3 f(y_3) T_t(y_3|y_1) \cdot \underbrace{\int dy_2 T_{\Delta t}(y_2|y_3)}_{=1 \text{ inserted}}
\end{aligned}$$

In 2^{nd} integral perform integration over y_3 using the definitions for A and B.
In 3^{rd} and 4^{th} integral exchange the dummy integration variables y_2 and y_3.
The 1^{st} integral removes the range $|y_3 - y_2| < \epsilon$ from the 4^{th} integral

$$\begin{aligned}
\frac{d}{dt}\langle f(y)\rangle \quad = \quad & \int \left[f'(y_2) A(y_2) + \frac{1}{2} f''(y_2) B(y_2) \right] T_t(y_2|y_1) dy_2 + \\
& + \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[\int\int_{|y_3-y_2|>\epsilon} f(y_2) \left(T_{\Delta t}(y_2|y_3) T_t(y_3|y_1) - T_{\Delta t}(y_3|y_2) T_t(y_2|y_1) \right) dy_2 dy_3 \right]
\end{aligned}$$

Integrate 1^{st} integral by parts

$$\begin{aligned}
\frac{d}{dt} \int f(y_2) T_t(y_2|y_1) dy_2 \quad = \quad & \int f(y_2) \left[-\frac{\partial}{\partial y_2} \left(A(y_2) T_t(y_2|y_1) \right) + \frac{1}{2} \frac{\partial^2}{\partial y_2^2} \left(B(y_2) T_t(y_2|y_1) \right) \right] dy_2 + \\
& + \int dy_2 f(y_2) \int_{|y_3-y_2|>\epsilon} dy_3 \left(W(y_2|y_3) T_t(y_3|y_1) - W(y_3|y_2) T_t(y_2|y_1) \right)
\end{aligned}$$

Since $f(y)$ is arbitrary the integrands have to be equal and one gets the *differential Chapman-Kolmogorov equation*

$$\begin{aligned}
\frac{\partial}{\partial t} T_t(y_2|y_1) \quad = \quad & -\frac{\partial}{\partial y_2} \left(A(y_2) T_t(y_2|y_1) \right) + \frac{1}{2} \frac{\partial^2}{\partial y_2^2} \left(B(y_2) T_t(y_2|y_1) \right) + \\
& + \int_{|y_3-y_2|>\epsilon} dy_3 \left(W(y_2|y_3) T_t(y_3|y_1) - W(y_3|y_2) T_t(y_2|y_1) \right)
\end{aligned} \tag{14}$$

Note:

- Strictly speaking, the integral is actually the Cauchy principal value, i.e. in the limit $\epsilon \to 0$. The Cauchy principal value need not exist if $W(y_2|y_3)$ diverges for $y_2 \to y_3$. We will assume in the following that there is no such problem and will write simply the usual integral sign instead.
- For boundary terms to vanish restrict $f(y)$ such that it vanishes outside the domain of interest.

The differential Chapman-Kolmogorov equation can also be derived in the non-stationary case,

$$\begin{aligned}\frac{\partial}{\partial t}P_{1|1}(y_2,t_2|y_1,t_1) &= -\frac{\partial}{\partial y_2}\left(A(y_2,t_2)P_{1|1}(y_2,t_2|y_1,t_1)\right) + \\ &+\frac{1}{2}\frac{\partial^2}{\partial y_2^2}\left(B(y_2,t_2)P_{1|1}(y_2,t_2|y_1,t_1)\right) + \\ &+\int_{|y_3-y_2|>\epsilon} dy_3\left(W(y_2|y_3,t_3)P_{1|1}(y_3,t_3|y_1,t_1) - W(y_3|y_2,t_2)P_{1|1}(y_2,t_2|y_1,t_1)\right)\end{aligned} \tag{15}$$

The integral term corresponds to the master equation (13) and describes jumps. We will discuss the meaning of the other two terms below.

From this equation for the transition probability one can also obtain an equation for the probability $P(y,t) \equiv P_1(y,t) = \int P_{1|1}(y,t|y_1,0)P(y_1,0)\,dy_1$ by multiplying (14) or (15) by $P(y_1,0)$ and integrating over y_1

$$\begin{aligned}\frac{\partial}{\partial t}P(y,t) &= -\frac{\partial}{\partial y}\left(A(y,t)P(y,t)\right) + +\frac{1}{2}\frac{\partial^2}{\partial y^2}\left(B(y,t)P(y,t)\right) + \\ &+\int_{|y-y_3|>\epsilon} W(y|y_3,t_3)P(y_3,t_3) - W(y_3|y,t)P(y,t)\,dy_3\end{aligned} \tag{16}$$

3.4.1 Drift Term

Consider

$$\frac{\partial}{\partial t}T_t(y|y_1) = -\frac{\partial}{\partial y}\left(A(y)T_t(y|y_1)\right) \tag{17}$$

Introduce $y_0(t)$ satisfying

$$\frac{d}{dt}y_0 = A(y_0,t) \qquad \text{with } y_0(0) = y_1 \tag{18}$$

Show that

$$T_t(y|y_1) = \delta(y - y_0(t))$$

satisfies the drift equation (17).

Left-hand side

$$\frac{\partial}{\partial t}T_t(y|y_1) = -\frac{d}{dy}\delta(y-y_0(t))\cdot\frac{d}{dt}y_0 = -\frac{d}{dy}\delta(y-y_0(t))\cdot A(y_0,t)$$

To evaluate the right-hand-side of (17) use the definition of distributions like the δ-function or its derivative via integrals over smooth functions[7],

$$\begin{aligned}
\int f(y) \left[\frac{\partial}{\partial y}\left(A(y)\delta(y-y_0(t))\right)\right] dy &= -\int f'(y)A(y)\delta(y-y_0(t))dy = \\
&= -f'(y_0(t))A(y_0) = \\
&= -\int A(y_0)f'(y)\delta(y-y_0(t))dy = \\
&= \int f(y) \left[A(y_0)\frac{d}{dy}\delta(y-y_0(t))\right] dy
\end{aligned}$$

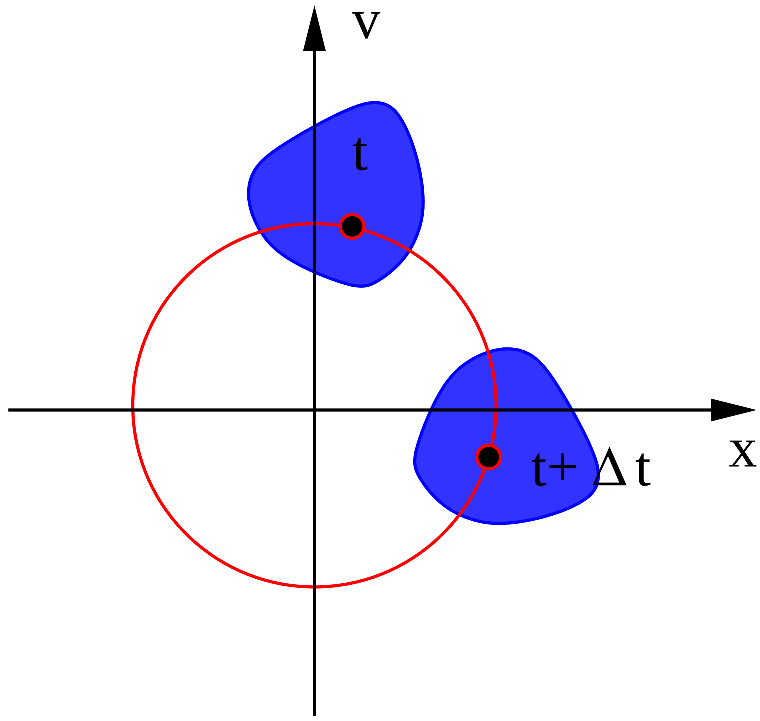
Thus

$$\begin{aligned}
\frac{\partial}{\partial y}\left(A(y)T_t(y|y_1)\right) &= \frac{d}{dy}\left(A(y_0(t))\delta(y-y_0(t))\right) \\
&= A(y_0(t))\frac{d}{dy}\delta(y-y_0(t))
\end{aligned}$$

Notes:

- the drift equation describes the deterministic motion of a particle with velocity $A(y_0)$: if it initially has the well-defined position y_1 it stays on the trajectory given by (18)
- using
  $$P_1(y,t) = T_t(y|y_1)P_1(y_1,0)$$
  (17) yields
  $$\frac{\partial}{\partial t}P_1(y,t) = -\frac{\partial}{\partial y}\left(A(y)P_1(y,t)\right)$$
  which is the Liouville equation of statistical mechanics for the evolution of the probability distribution of an ensemble of particles evolving deterministically under (18).
  Example: for the spring-mass system (harmonic oscillator) the motion in phase space (x,v) is given by an ellipse. The motion of the individual harmonic oscillator (marked by circles) is described by $P_1(y,t) = \delta(y-y_0(t))$. The evolution of an ensemble of harmonic oscillators is given by the motion of the marked region.

[7] use integration by parts with vanishing boundary terms.



3.4.2 Diffusion Term

Consider

$$\frac{\partial}{\partial t}T_t(y|y_1) = \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(B(y)T_t(y|y_1)\right)$$

Determine evolution of a δ-peak for short times

$$T_0(y|y_1) = \delta(y - y_1)$$

Initially $T_t(y|y_1)$ is so sharply peaked that the derivative of $B(y)$ can be ignored

$$\frac{\partial}{\partial t}T_t(y|y_1) = \frac{1}{2}B(y)\frac{\partial^2}{\partial y^2}T_t(y|y_1)$$

Since $T_t(y|y_1)$ is sharply peaked at y_1 one can initially also ignore the y-dependence of $B(y)$ and gets then

$$T_{\Delta t}(y|y_1) = \frac{1}{\sqrt{2\pi B\Delta t}}\, e^{-\frac{1}{2}\frac{(y-y_1)^2}{B\Delta t}}$$

Again for $P_1(y, 0) = \delta(y - y_1)$ one gets

$$P_1(y, \Delta t) = \frac{1}{\sqrt{2\pi B\Delta t}}\, e^{-\frac{1}{2}\frac{(y-y_1)^2}{B\Delta t}}$$

Notes:

- This term describes diffusion-like spreading of the probability distribution
- the sample paths are continuous:
  only transitions with $y - y_1 = \mathcal{O}(\Delta t^{\frac{1}{2}})$ are likely, i.e. $y - y_1 \to 0$ for $\Delta t \to 0$

- the sample paths need not be differentiable:

$$\frac{y - y_1}{\Delta t} = \mathcal{O}(\Delta t^{-\frac{1}{2}}) \to \infty$$

- Together, the drift term and the diffusion term constitute the Fokker-Planck equation for $T_t(y_2|y_1)$

$$\frac{\partial}{\partial t} T_t(y_2|y_1) = -\frac{\partial}{\partial y_2} \left(A(y_2) T_t(y_2|y_1) \right) + \frac{1}{2} \frac{\partial^2}{\partial y_2^2} \left(B(y_2) T_t(y_2|y_1) \right) \tag{19}$$

  or for $P(y,t)$

$$\frac{\partial}{\partial t} P(y,t) = -\frac{\partial}{\partial y} \left(A(y,t) P(y,t) \right) + + \frac{1}{2} \frac{\partial^2}{\partial y^2} \left(B(y,t) P(y,t) \right) \tag{20}$$

4 Fokker-Planck Equation

4.1 The Rayleigh Particle[8]

Consider a Brownian particle on time scales that may be shorter than the correlation time for the velocity, but still much longer than the time between individual collisions:

- the velocity v should be Markovian
- the position x need not be Markovian: for short enough times the particle 'remembers' its previous position

Describe the velocity of the particle: forces of the collisions are finite therefore the velocity is continuous $\rightarrow$ only Fokker-Planck terms contribute.
How do we get the coefficients for the drift and the diffusion?

Macroscopically, averaging over all the collisions, the particle is exposed to drag

$$\frac{d}{dt} v = -\gamma v$$

From the Fokker-Planck equation we get

$$\begin{aligned}
\frac{d}{dt} \langle V \rangle \quad &= \quad \int dv\, v \frac{\partial}{\partial t} T_t(v|v_1) = \\
&= \quad \int dv \left\{ -v \frac{\partial}{\partial v} \left(A(v) T_t(v|v_1) \right) + \frac{1}{2} \frac{\partial^2}{\partial v^2} \left(B(v) T_t(v|v_1) \right) \right\} = \\
&\underbrace{=}_{\text{integrate by parts}} \quad \int dv\, A(v) T_t(v|v_1) = \\
&= \quad \langle A(V) \rangle
\end{aligned}$$

[8] [28, VIII.4]

Choose

$$A(v) = -\gamma v$$

then

$$\langle A(V) \rangle = - \langle \gamma V \rangle = -\gamma \langle V \rangle \tag{21}$$

and the mean satisfies the macroscopic equation.

Note:

- For nonlinear $A(v)$ one has
$$\langle A(v) \rangle \neq A(\langle v \rangle)$$
and the mean $\langle V \rangle$ would not necessarily satisfy the macroscopic equation.
This leads to corrections (see Sec.4.2).

The coefficient $B(v)$ describes the diffusive spread of the velocity due to the randomness of the collisions: 'noise'. In the steady state one gets a stationary probability distribution function for V.
In thermodynamic equilibrium the velocity is distributed according to the Maxwell distribution,

$$P_e(v) = \sqrt{\frac{M}{2\pi kT}} e^{-\frac{1}{2}\frac{Mv^2}{kT}}$$

where M is the mass of the particle, k is the Boltzmann constant, and T is the temperature.
Note the Boltzmann form of the distribution $P_e \propto e^{-E/kT}$.
P_e has to satisfy the Fokker-Planck equation (20)

$$\begin{aligned} 0 &= \gamma \frac{\partial}{\partial v}(vP_e) + \frac{1}{2}\frac{\partial^2}{\partial v^2}(B(v)P_e(v)) = \\ &= \gamma P_e + \gamma v \left(-\frac{Mv}{kT}\right) P_e + \frac{1}{2} \underbrace{B}_{\text{try constant } B} \left[-\frac{M}{kT} + \left(\frac{Mv}{kT}\right)^2 \right] P_e \end{aligned}$$

Comparing coefficients generates two conditions

$$\gamma = \frac{BM}{2kT} \qquad \frac{M\gamma}{kT} = \frac{1}{2} B \frac{M^2}{k^2T^2}$$

Since they are consistent with each other we have a solution and have determined the noise term in the Fokker-Planck equation

$$\frac{\partial}{\partial t} P(v,t) = \gamma \frac{\partial}{\partial v}(vP(v,t)) + \gamma \frac{kT}{M} \frac{\partial^2}{\partial v^2} P(v,t) \tag{22}$$

and in the analogous equation for $T_t(y_2|y_1)$.

Notes:

- $B \equiv \left\langle (\Delta v)^2 \right\rangle / \Delta t$ gives the strength of the fluctuations. It is related to the drag, which represents the dissipation:[9]

$$\underbrace{\frac{\left\langle (\Delta v)^2 \right\rangle}{2\Delta t}}_{\text{fluctuations}} = B = \underbrace{\gamma \frac{kT}{M}}_{\text{dissipation}} \qquad \text{both Einstein-like relations combined}$$

  This illustrates that fluctuations and dissipation are due to the same mechanism: collisions with other particles.
  In statistical mechanics the *fluctuation-dissipation theorem* gives a quite general relation between fluctuations (correlations) and dissipation (response functions) in the *linear* regime near equilibrium [28, p.89].

To determine the moment $\langle X(t)X(t')\rangle$ we need the transition probability $T_t(v|v_1)$ in addition to $P(v,t)$. The Fokker-Planck equation for the transition probability is solved by the transition probability of the Ornstein-Uhlenbeck process

$$T_t(v|v_1) = \sqrt{\frac{M}{2\pi kT\left(1 - C(t)^2\right)}} e^{-\frac{1}{2}\frac{M(v - v_1 C(t))^2}{kT(1-C(t)^2)}}$$

with

$$C(t) = e^{-\gamma t}$$

and

$$P_1(v) = \sqrt{\frac{M}{2\pi kT}} e^{-\frac{1}{2}\frac{Mv^2}{kT}}$$

Consider now the position $x(t)$ of the particle

$$X(t) = \int_0^t V(t')dt'$$

where we assume $X(0) = 0$. $V(t)$ is a Gaussian process. Since the sum of Gaussian processes is again a Gaussian process, $X(t)$ is also a Gaussian process.

We want $\langle X(t)\rangle$ and $\langle X(t_1)X(t_2)\rangle$ for a particle that starts at a fixed position $x = 0$. The average is over the different realizations of the random process V, i.e. over different trajectories that all start at the same location.

Since $\langle V(t)\rangle = 0$ we have also

$$\langle X(t)\rangle = \int_0^t \langle V(t')\rangle \, dt' = 0$$

and

$$\langle X(t_1)X(t_2)\rangle = \int_0^{t_1} dt' \int_0^{t_2} dt'' \, \langle V(t')V(t'')\rangle$$

with $\langle V(t)V(t+\tau)\rangle = \langle V(t)^2\rangle \, e^{-\gamma\tau} = \langle V^2\rangle \, e^{-\gamma t}$.

[9] Einstein derived these relations for the position.

Thus

$$\langle X(t_1)X(t_2)\rangle = \left\langle V^2\right\rangle \int_0^{t_1} dt' \int_0^{t_2} dt'' e^{-\gamma|t'-t''|}$$

Without loss of generality assume $t_2 \geq t_1$

$$\begin{aligned}
\langle X(t_1)X(t_2)\rangle &= \langle V^2\rangle \quad \int_0^{t_1} dt' \underbrace{\int_0^{t'} dt'' e^{-\gamma(t'-t'')}}_{\frac{1}{\gamma}\left(1-e^{-\gamma t'}\right)} + \int_0^{t_1} dt' \underbrace{\int_{t'}^{t_2} dt'' e^{-\gamma(t''-t')}}_{-\frac{1}{\gamma}\left(e^{-\gamma(t_2-t')}-1\right)} = \\
&= \quad \frac{\langle V^2\rangle}{\gamma}\left\{t_1 + \frac{1}{\gamma}\left(e^{-\gamma t_1} - 1\right) + t_1 - \frac{1}{\gamma}\left(e^{-\gamma(t_2-t_1)} - e^{-\gamma t_2}\right)\right\} = \\
&= \quad \frac{\langle V^2\rangle}{\gamma^2}\left\{2\gamma t_1 - 1 + e^{-\gamma t_1} + e^{-\gamma t_2} - e^{-\gamma(t_2-t_1)}\right\}
\end{aligned}$$

Thus

$$\left\langle X(t)^2\right\rangle = \frac{2\left\langle V^2\right\rangle}{\gamma^2}\left\{\gamma t + e^{-\gamma t} - 1\right\} \to \left\{ \begin{array}{ll} \left\langle V^2\right\rangle t^2 + \dots & \text{for } \gamma t \ll 1 \\ \frac{2}{\gamma}\left\langle V^2\right\rangle t + \dots & \text{for } \gamma t \gg 1 \end{array}\right.$$

Notes:

- the process is not stationary: the initial condition was fixed at $t = 0$ and $\langle X^2(t)\rangle$ gives the spread of the particle after that time.
- for $\gamma t \ll 1$ the velocity is still correlated $\langle V(0)V(t)\rangle = \mathcal{O}(1)$:
  - inertia is relevant and particle moves almost ballistically: distance covered is linear in the time
  - $\langle X(t)^2\rangle$ is independent of γ, it is only determined by the velocity scale $\langle V^2\rangle$
  - X is non-Markovian
- for $\gamma t \gg 1$ the velocity is uncorrelated $\langle V(0)V(t)\rangle \to 0$:
  - particle undergoes random walk, i.e. Wiener process
  - for $\gamma t_1, \gamma t_2, \gamma(t_2 - t_1) \gg 1$ one gets

    $$\langle X(t_1)X(t_2)\rangle = \frac{2}{\gamma}\left\langle V^2\right\rangle \min(t_1, t_2) \tag{23}$$

    (the $\min(t_1, t_2)$ arises because we assumed $t_1 < t_2$ in the derivation).
  - Einstein relation for the position

    $$\left\langle X(t)^2\right\rangle = \frac{2kT}{\gamma M} t$$

4.2 The Macroscopic Equation[10]

In our discussion of the Rayleigh particle we had set in (21)

$$\langle A(V)\rangle = A(\langle V\rangle)$$

in order to relate the macroscopic equation to the Fokker-Planck equation.
In general this is not valid, instead one has for solutions of the Fokker-Planck equation

$$\frac{d}{dt}\langle V\rangle = \langle A(V)\rangle$$

Consider the corrections for weak fluctuations → only velocities close to $\langle V\rangle$ are relevant.

Expand

$$A(V) = A(\langle V\rangle) + (V-\langle V\rangle)\,A'(\langle V\rangle) + \frac{1}{2}(V-\langle V\rangle)^2\,A''(\langle V\rangle) + \dots$$

then

$$\langle A(V)\rangle = A(\langle V\rangle) + \frac{1}{2}\left(\langle V^2\rangle - \langle V\rangle^2\right)^2 A''(\langle V\rangle) + \dots$$

$$\frac{d}{dt}\langle V\rangle = A(\langle V\rangle) + \frac{1}{2}\sigma^2 A''(\langle V\rangle) + \dots$$

This is not a closed equation for $\langle V\rangle$, it depends on σ^2 and higher moments of V
Need at least an evolution equation for σ^2

$$\begin{aligned}\frac{d}{dt}\langle V^2\rangle &= \int -v^2\frac{\partial}{\partial v}\left(A(v)T_t(v|v_1)\right) + \frac{1}{2}v^2\frac{\partial^2}{\partial v^2}\left(B(v)T_t(v|v_1)\right)dv = \\ &\underbrace{=}_{\text{i.b.p.}} 2\langle VA(V)\rangle + \langle B(V)\rangle\end{aligned}$$

Thus

$$\frac{d}{dt}\sigma^2 = \langle B(V)\rangle + 2\langle VA(V)\rangle - 2\langle V\rangle\langle A(V)\rangle = \langle B(V)\rangle + 2\langle (V-\langle V\rangle)\,A(V)\rangle$$

Expand again

$$\langle (V-\langle V\rangle)\;A(\langle V\rangle + V - \langle V\rangle)\rangle = A'(\langle V\rangle)\underbrace{\langle (V-\langle V\rangle)^2\rangle}_{\sigma^2} + \underbrace{\frac{1}{2}A''(\langle V\rangle)\langle (V-\langle V\rangle)^3\rangle + \dots}_{\text{higher moments}}$$

Small fluctuations →B is small, no need to expand it.

Thus, ignoring contributions from higher moments one gets

$$\begin{aligned}\frac{d}{dt}\sigma^2 &= B(\langle V\rangle) + 2\sigma^2 A'(\langle V\rangle)\\ \frac{d}{dt}\langle V\rangle &= A(\langle V\rangle) + \frac{1}{2}\sigma^2 A''(\langle V\rangle)\end{aligned}$$

Notes:

[10] [28, V.8]

- for the single macroscopic equation for $\langle V\rangle$ to be valid one needs

$$\left|\frac{\sigma^2 A''}{A}\right| \ll 1$$

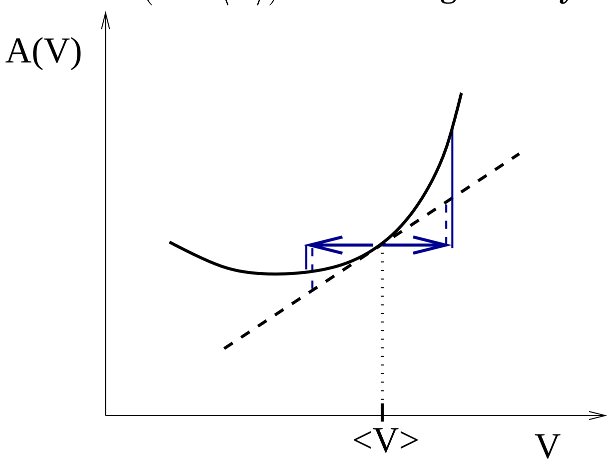
with $\sigma^2 = \mathcal{O}$(B/A') one gets

$$B \ll \left|\frac{AA'}{A''}\right|$$

i.e. the curvature of A needs to be small, only small deviation from linear dependence or equivalently, the noise strength should be small enough for v to remain in the linear regime

- if the deterministic force $A(V)$ is nonlinear, fluctuations do modify the mean velocity $\langle V\rangle$ through A'': the change in the force for positive and for negative fluctuations away from the mean do in general not compensate each other

- to this order the fluctuations are symmetric (V-dependence of B not important) and the term $(V - \langle V\rangle)\, A'$ is averaged away



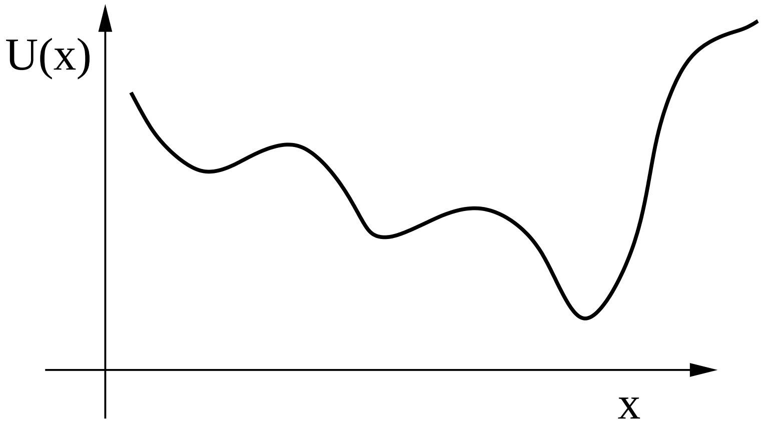
- for σ^2 to saturate one needs $A'(\langle V\rangle) < 0$: positive dissipation (drag)
  for $A'(\langle V\rangle) > 0$ the system is unstable for this mean value of V

- σ^2 and $\langle V\rangle$ relax on the same time scale $A'(\langle V\rangle)$ (linearize the $\langle V\rangle$-equation) $\rightarrow$ no separation of time scales and no decoupling of the two equations if the fluctuations are to be kept at all.

4.3 Brownian Motion in an External Potential[11]

Consider the motion of a Brownian particle in potential wells and across potential barriers. The particle need not be a true particle, it can be thought of as a variable characterizing other random processes like chemical reactions (barrier = activation energy) etc.

[11][29, can be downloaded from the class web site]



For a true Brownian particle one would have to consider a Rayleigh particle in an external potential
Fokker-Planck equation for the velocity V alone is not sufficient, since the force depends on the position X
the deterministic equations are given by

$$\begin{aligned}\frac{dx}{dt} &= v \\ \frac{dv}{dt} &= \frac{1}{M}F(x,t)\end{aligned}$$

where M is the mass of the particle.
$\rightarrow$ both X and V are random variables and one gets a *bivariate* Fokker-Planck equation for $T_t(X,V|X_1,V_1)$, which is a 3-dimensional PDE (Kramers' equation).

We would like to simplify the situation to $P(x,t)$: need $A(x)$ and therefore we need

$$\frac{dx}{dt} = f(x,t)$$

4.3.1 Markov Approximation for $X(t)$ and Weak Noise

To simplify matters: consider sufficiently long time scales such that the Markov approximation for X is sufficient.
For constant force F the situation would be just like that for the free Rayleigh particle:

- $\gamma t \lesssim 1$: inertia is relevant, X is not Markovian
- $\gamma t \gg 1$: inertia irrelevant, $v = F/\gamma$, and the deterministic equation for x contains only x

  $$\frac{dx}{dt} = v = \frac{F(x)}{\gamma}$$

  X is Markovian

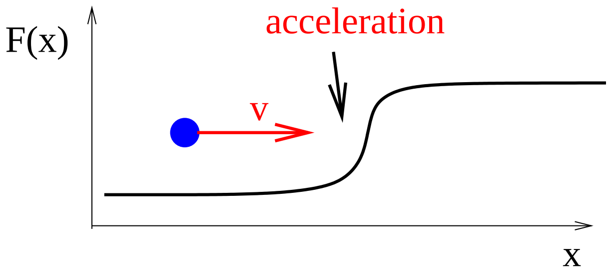
Non-constant force:

- Time-dependent $F(t)$:
for adiabatic approximation $v(t) = \frac{F(t)}{\gamma}$ to be valid need that the velocity relaxes faster than the force changes

$$\gamma \gg \frac{1}{F}\frac{dF}{dt}$$

- Space-dependent force $F(x)$:
as the particle moves it experiences a time-dependent force

$$\frac{dF}{dt} = \frac{dF}{dx}\langle V\rangle$$



thus need

$$\gamma \gg \langle V\rangle \frac{1}{F}\frac{dF}{dx}$$

Assume: $F(x)$ varies slowly in space. Then can use adiabatic approximation

$$v(x) = \frac{F(x)}{\gamma} = -\frac{1}{\gamma}\frac{dU(x)}{dx}$$

Macroscopic equation for the position

$$-A(x) = \frac{dx}{dt} = v = -\frac{1}{\gamma}\frac{dU}{dx} \tag{24}$$

Assuming $B(x) = b = const.$ we get then the Fokker-Planck equation for $P(x, t)$

$$\frac{\partial P}{\partial t} = \frac{1}{\gamma}\frac{\partial}{\partial x}\left(U'(x)P\right) + \frac{1}{2}b\frac{\partial^2}{\partial x^2}P$$

Notes:

- The macroscopic equation (24) is not restricted to describing a Brownian particle. It describes many dissipative (overdamped) systems: x could represent the magnetization of a magnet, the density of a liquid, the concentration of a mixture, a variable characterizing a chemical reaction.
- Fluctuations are in particular interesting near phase transitions (where a state becomes linearly unstable and fluctuations are amplified and in multi-stable systems where fluctuations allow transitions between the different linearly stable states (see Sec.4.3.3)

Consider again weak noise: $b \ll 1$

Expect: the particle follows mostly the macroscopic trajectory

$$\frac{dx}{dt} = -\frac{1}{\gamma}\frac{dU}{dx}$$

Go into a frame moving along with the macroscopic motion $x = \phi(t)$ and rewrite the probability

$$P(x,t) = P(\xi(x,t),t)$$

with

$$\xi(x,t) = \frac{1}{\sqrt{b}}\left(x - \phi(t)\right) \qquad \text{i.e.} \qquad x = \phi(t) + \sqrt{b}\,\xi$$

Note:

- ξ is a stretched variable, i.e. expect $x - \phi(t) = \mathcal{O}(\sqrt{b})$ and P is sharply peaked around $x = \phi(t)$. With this scaling $\xi = \mathcal{O}(1)$.

$$\begin{aligned} \frac{\partial}{\partial t}P &\rightarrow \frac{\partial}{\partial t}P - \frac{1}{\sqrt{b}}\frac{d\phi}{dt}\frac{\partial}{\partial \xi}P \\ \frac{\partial}{\partial x}P &\rightarrow \frac{1}{\sqrt{b}}\frac{\partial}{\partial \xi}P \end{aligned}$$

In this frame the potential becomes explicitly time-dependent:

$$\begin{aligned} U(x) &= U(\phi(t) + \sqrt{b}\xi) = \\ &= U(\phi(t)) + \sqrt{b}\xi U'(\phi(t)) + \frac{1}{2}b\xi^2 U''(\phi(t)) + \ldots \end{aligned}$$

Insert into Fokker-Planck equation

$$\frac{\partial}{\partial t}P - \frac{1}{\sqrt{b}}\frac{d\phi}{dt}\frac{\partial}{\partial \xi}P = +\frac{1}{\gamma}\frac{1}{b}\frac{\partial}{\partial \xi}\left(\left(\frac{\partial}{\partial \xi}U\right)\, P\right) + \frac{1}{2}\frac{\partial^2}{\partial \xi^2}P$$

Expect that an equation for the fluctuations arises at $\mathcal{O}(1)$ because the diffusion term is of that order.
Expand in the drift term

$$\begin{aligned} \frac{\partial}{\partial \xi}\left(\left(\frac{\partial}{\partial \xi}U\right)\, P\right) &= \frac{\partial}{\partial \xi}\left(\sqrt{b}U'(\phi)P + b\xi U''(\phi)P + \ldots\right) = \\ &= \sqrt{b}U'(\phi)\frac{\partial}{\partial \xi}P + bU''(\phi)\frac{\partial}{\partial \xi}(\xi P) + \ldots \end{aligned}$$

Collect orders in b

- $\mathcal{O}(b^{-1/2})$:

$$\frac{d\phi}{dt} = -\frac{1}{\gamma}U'(\phi)$$

  recovering the macroscopic equation of motion

- $\mathcal{O}(b^0)$:

$$\frac{\partial}{\partial t}P = \frac{1}{\gamma}U''(\phi)\frac{\partial}{\partial \xi}(\xi P) + \frac{1}{2}\frac{\partial^2}{\partial \xi^2}P$$

Determine mean and variance (using $P(\xi) \to 0$ for $\xi \to \pm\infty$)

$$\frac{d}{dt}\langle \xi \rangle = \frac{1}{\gamma}U''(\phi)\int \xi \frac{\partial}{\partial \xi}(\xi P)\, d\xi + \frac{1}{2}\int \xi \frac{\partial^2}{\partial \xi^2}P d\xi \underbrace{=}_{\text{i.b.p.}} -\frac{1}{\gamma}U''(\phi)\langle \xi \rangle$$

Analogously for $\langle \xi^2 \rangle$, yielding

$$\begin{aligned} \frac{d}{dt}\langle \xi \rangle &= -\frac{1}{\gamma}U''(\phi)\langle \xi \rangle \\ \frac{d}{dt}\left\langle \xi^2 \right\rangle &= -\frac{2}{\gamma}U''(\phi)\left\langle \xi^2 \right\rangle + 1 \end{aligned}$$

Using integrating factors one obtains the solutions as

$$\begin{aligned} \langle \xi \rangle_t &= \langle \xi \rangle_{t=0} e^{-\frac{1}{\gamma}\int_0^t U''(\phi(t'))dt'} \\ \left\langle \xi^2 \right\rangle_t &= \left\langle \xi^2 \right\rangle_{t=0} e^{-\frac{2}{\gamma}\int_0^t U''(\phi(t'))dt'} + \int_0^t e^{-\frac{2}{\gamma}\int_{t'}^t U''(\phi(t''))dt''} dt' \end{aligned}$$

Notes:

- all along the macroscopic trajectory $x = \phi(t)$ to leading order in b (b^0) the potential in the Fokker-Planck equation is approximated by a quadratic (i.e. a linear force)
- since the drift term $A(\xi)$ of the lowest order (b^0) Fokker-Planck equation is linear, i.e. $A(\xi) = \frac{1}{\gamma}U''(\phi)\xi$, P is a Gaussian centered around the macroscopic path

$$P(x,t|x_1,t_1) = \frac{1}{\sqrt{2\pi\left\langle \xi^2 \right\rangle_t}} e^{-\frac{1}{2}\frac{(x-\phi(t))^2}{\left\langle \xi^2 \right\rangle_t}}$$

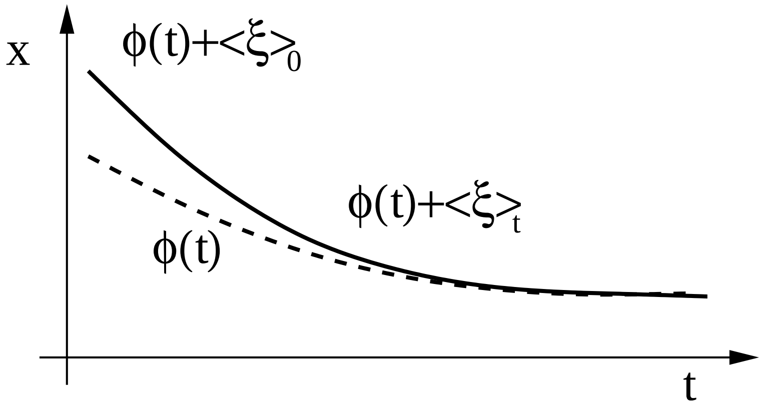
- for the full process P need not be Gaussian since in general the potential will not be quadratic
- for $U''(\phi) > 0$:
  - the deviations from the macroscopic trajectory go to zero in the mean

    $$\langle \xi \rangle_t \to 0 \qquad \text{for} \quad t \to \infty$$

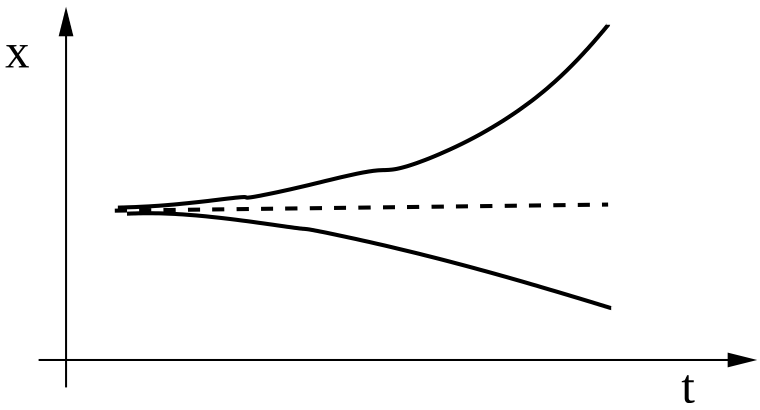
    thus, deviations from the macroscopic trajectory decrease with time.
    In particular, near a stable equilibrium the particle approaches the equilibrium in the mean.

– $\langle \xi^2 \rangle_t$ is bounded: the spreading by the fluctuations is balanced by the convergence of the trajectories. In particular, near a stable equilibrium

$$\langle \xi^2 \rangle_t \to \frac{\gamma}{2U''(\phi_{eq})}$$



- For $U''(\phi) < 0$: trajectory unstable, fluctuations grow without bound (until they reach a stable regime), trajectories of different realizations of the fluctuations diverge from each other.



4.3.2 Fluctuations in a Steady State

If the system approaches a stable steady state the distribution P can be given exactly without assuming small deviations from the steady state.

Fokker-Planck equation

$$\frac{\partial P}{\partial t} = \frac{1}{\gamma}\frac{\partial}{\partial x}\left(U'(x)P\right) + \frac{1}{2}b\frac{\partial^2}{\partial x^2}P$$

For steady state $\partial_t P = 0$ we can integrate once to get

$$\frac{1}{2}\gamma b\frac{\partial}{\partial x}P = -U'(x)P + C$$

Assume the particle is confined to a finite domain:

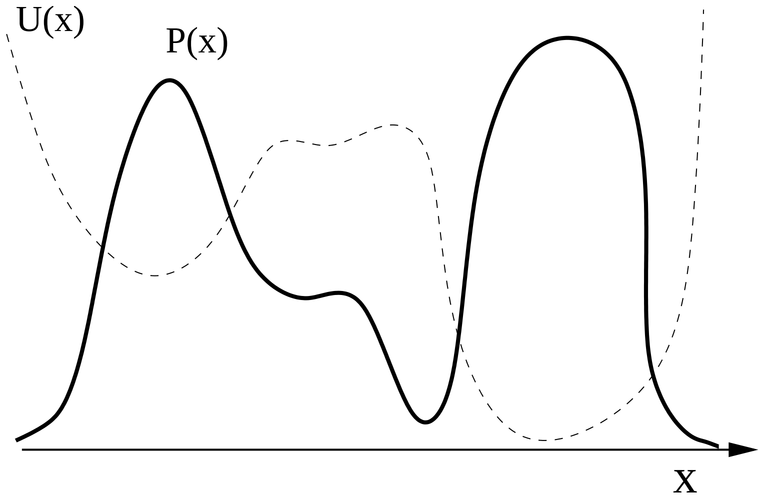
$$U(x) \to \infty \quad \text{and} \quad P(x) \to 0 \quad \text{for } |x| \to \infty$$

If $P(x)$ decreases fast enough so that $U'(x)P \to 0$ one gets $C = 0$.

$$P(x) = \frac{1}{N} e^{-\frac{2U(x)}{b\gamma}} \qquad \text{with} \quad N = \int P(x)dx$$

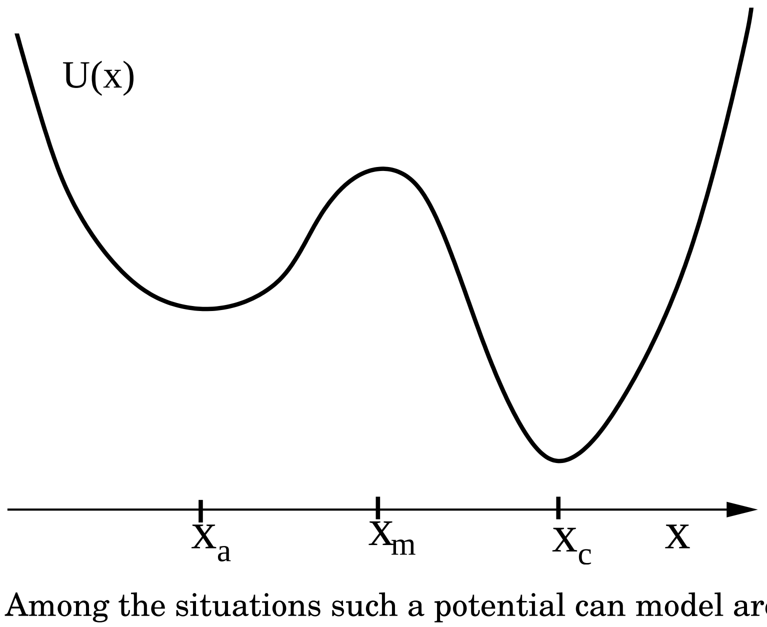
Notes:

- the speed of the decay of $P(x)$ for $|x| \to \infty$ is consistent with the assumption.
- In thermal equilibrium
$$P(x) = \frac{1}{N} e^{-\frac{U(x)}{kT}}$$
therefore
$$\frac{1}{2} b\gamma = kT \tag{25}$$



4.3.3 Bistable Systems: Escape from a Metastable State

Consider a particle in a potential with 2 minima



Among the situations such a potential can model are

- Chemical reactions, e.g.

$$A \quad + \quad B \quad \rightleftarrows C$$

  x is a reaction coordinate, it measures the progress of the reaction, it could be related to the distance between the reacting molecules

$$\begin{aligned} x &= x_a \quad \rightarrow \quad \text{A and B are separate} \\ x &= x_c \quad \rightarrow \quad \text{A and B are bound together forming C} \end{aligned}$$

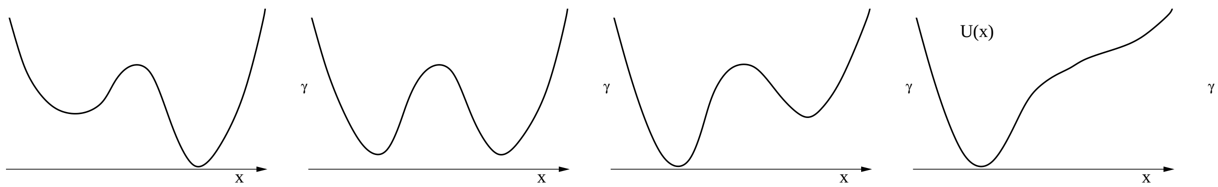
  In order to form the cound state $AB \equiv C$ typically an energy barrier has to be overcome
  This model is very primitive:

  - only single reaction coordinate (in addition to position, the molecule orientation, conformation or even simply bending of the molecule could be relevant, etc.)
  - molecule velocities are ignored
  - ...

- 1^{st}-order phase transition
  x is given by an *order parameter*. E.g., in liquid-gas transition x related to density of the fluid

$$\begin{aligned} x &= x_a \quad \rightarrow \quad \text{liquid state} \\ x &= x_c \quad \rightarrow \quad \text{gaseous state} \end{aligned}$$

  when the temperature changes the shape of the potential changes



Notes:

- x_a and x_c are (linearly) stable states:

  - in the absence of fluctuations they persist forever
    e.g., liquid state can persist even above the boiling point
  - with fluctuations the state with higher energy is only *metastable*: there is a finite probability that fluctuations will push the system across that energy barrier
    life-time of the metastable state depends on the strength of the fluctuations relative to the height of the barrier:
    nucleation seeds (dust, boiling stones) lower the energy barrier and can trigger the transition

Goal: determine the *mean first passage time* $\tau(x_c|x_a)$: average of the time t_{ac} the particle takes to leave the metastable well.

Note:

- t_{ac} depends also on the initial position x_1 inside the first well and the final position x_2 in the second well.
  $\rightarrow$ for the definition to make sense the average $\langle t_{ac}\rangle$ must be much longer than the time for the particle to traverse either well
  $\rightarrow$ the barrier must be sufficiently high

Consider an intermediate position

$$x_1 \underbrace{\rightarrow}_{\Delta t} x' \rightarrow x_2$$

- in time interval Δt system goes from x_1 (which does not have to be inside the well with minimum x_a to some x')
- from x' it goes on to x_2 somewhere in the well with minimum x_b: this takes on average $\tau(x_2|x')$
- average time given by average over all allowed positions x' weighted by the probability to get their from x_1

For $x_1 < x_2$ consider the mean first passage time from x_1 to x_2 (which will eventually be assumed to be x_a and x_c, respectively)

$$\tau(x_2|x_1) = \underbrace{\Delta t}_{\text{time to get to } x'} + \underbrace{\int_{-\infty}^{x_2}}_{\textit{first}\text{ passage time}} \tau(x_2|x')\, T_{\Delta t}(x'|x_1)dx'$$

Note:

- for *first* passage time the intermediate state x' is not allowed to be beyond the final state x_2, that defines the upper limit for the x'-integral

$T_{\Delta t}(x'|x_1)$ satisfies the Fokker-Planck equation

$$\frac{\partial}{\partial t}T_t(x'|x_1) = -\frac{\partial}{\partial x'}\left(A(x')T_t(x'|x_1)\right) + \frac{1}{2}\frac{\partial^2}{\partial x'^2}\left(B(x')T_t(x'|x_1)\right)$$

For small Δt the solution $T_{\Delta t}(x'|x_1)$ that starts at x_1, i.e. with initial condition $\delta(x'-x_1)$, is given by

$$T(x'|x_1) = \delta(x'-x_1) + \Delta t\left\{-\frac{\partial}{\partial x'}\left(A(x')\delta(x'-x_1)\right) + \frac{1}{2}\frac{\partial^2}{\partial x'^2}\left(B(x')\delta(x'-x_1)\right)\right\}$$

Insert that into the integral

$$\begin{aligned}\tau(x_2|x_1) &= \Delta t + \tau(x_2|x_1) + \\ &+\Delta t\int_{-\infty}^{x_2}\tau(x_2|x')\left[-\frac{\partial}{\partial x'}\left(A(x')\delta(x'-x_1)\right) + \frac{1}{2}\frac{\partial^2}{\partial x'^2}\left(B(x')\delta(x'-x_1)\right)\right]dx'\end{aligned}$$

Integrate by parts (boundary terms vanish since $\tau(x_2|x_2) = 0$ and δ-function localized at $x_1 > -\infty$).

$$\tau(x_2|x_1) = \Delta t + \tau(x_2|x_1) + \Delta t \left\{ A(x_1)\frac{\partial}{\partial x_1}\tau(x_2|x_1) + \frac{1}{2}B(x_1)\frac{\partial^2}{\partial x_1^2}\tau(x_2|x_1) \right\}$$

This yields the *Dynkin equation*

$$A(x_1)\frac{\partial}{\partial x_1}\tau(x_2|x_1) + \frac{1}{2}B(x_1)\frac{\partial^2}{\partial x_1^2}\tau(x_2|x_1) = -1 \tag{26}$$

with boundary condition

$$\tau(x_2|x_2) = 0$$

Notes:

- The Dynkin equation involves the adjoint of the Fokker-Planck operator (cf. integration by parts): the derivatives are acting on the second argument of τ
- The second boundary condition has to be chosen such that the boundary terms in the integration by parts vanish.

Solve using integrating factor for $A(x) = -U'(x)/\gamma$ and $B(x) = b$:
introduce $v(x_1) = \frac{\partial}{\partial x_1}\tau(x_2|x_1)$

$$v' - \frac{2}{b\gamma}U'v = -\frac{2}{b}$$

$$\frac{d}{dx_1}\left(e^{-\frac{2}{b\gamma}U}v\right) = -\frac{2}{b}e^{-\frac{2}{b\gamma}U}$$

Then

$$v(x_1) = e^{\frac{2}{b\gamma}U(x_1)}\left\{\int_{-\infty}^{x_1} -\frac{2}{b}e^{-\frac{2}{b\gamma}U(x')}dx' + C\right\}$$

Assume that for $x \to -\infty$ the potential diverges, $U(x \to -\infty) \to \infty$, then for $v(x_1)$ to remain finite for $x_1 \to -\infty$ one needs to have $C = 0$[12].

Thus

$$\tau(x_2|x_1) = \int_{\hat{x}}^{x_1} e^{\frac{2}{b\gamma}U(x')}\int_{-\infty}^{x'} -\frac{2}{b}e^{-\frac{2}{b\gamma}U(x'')}dx''$$

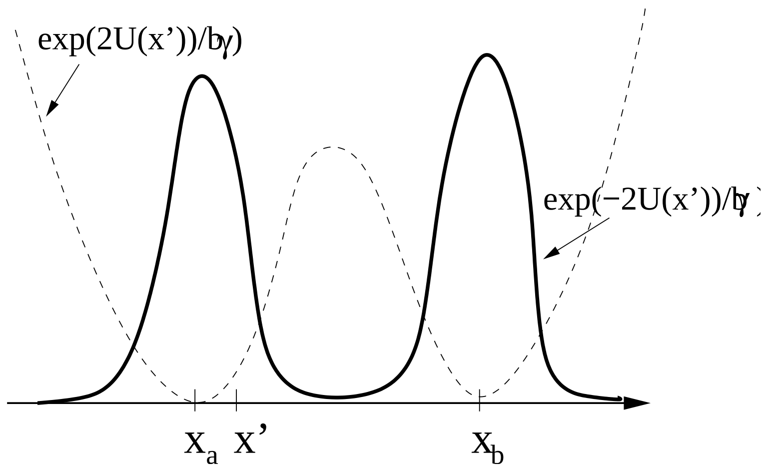
With the choice $\hat{x} = x_2$ the boundary condition $\tau(x_2|x_2) = 0$ is satisfied[13] and

$$\tau(x_2|x_1) = \frac{2}{b}\int_{x_1}^{x_2} e^{\frac{2}{b\gamma}U(x')}\left\{\int_{-\infty}^{x'} e^{-\frac{2}{b\gamma}U(x'')}dx''\right\}dx'$$

For weak noise, $b \ll 1$, the exponentials are sharply peaked:

[12] This amounts to a reflecting boundary condition for $x \to -\infty$

[13] This amounts to an absorbing boundary condition at $x = x_2$.



The x''-integral has sizeable contributions only for x'' near x_a and near x_c.
The x'-integrand is the product of $e^{\frac{2}{b\gamma}U(x')}$ and the area under the solid curve from $x=-\infty$ to x'.
$\Rightarrow$ it has sizeable contributions only if x' is close to the maximum x_m of the barrier.

Approximate $U(x)$ in the two exponentials

$$
\begin{aligned}
U(x') &= U(x_m) + \frac{1}{2}U''(x_m)\,(x'-x_m)^2 \qquad \text{for } x' \text{ near } x_m \\
U(x'') &= U(x_a) + \frac{1}{2}U''(x_a)\,(x''-x_a)^2 \qquad \text{for } x'' \text{ near } x_a
\end{aligned}
$$

Note:

- $U''(x_a) > 0$ while $U''(x_m) < 0$

For x_1 close to x_a one gets then

$$
\tau(x_2|x_1) = \frac{2}{b}e^{\frac{2}{b\gamma}(U(x_m)-U(x_a))}\int_{x_1}^{x_2} e^{\frac{1}{b\gamma}U''(x_m)(x'-x_m)^2}\left\{\int_{-\infty}^{x'} e^{-\frac{1}{b\gamma}U"(x_a)(x''-x_a)^2}dx''\right\}dx'
$$

Away from x_m the Gaussian involving x' decays very rapidly: can replace $x_{1,2}$ by $x_m \pm \delta$ with $\delta = \mathcal{O}(\sqrt{b})$.

Since now x' is restricted to $x' > x_m - \delta$ the x''-integral can be extended to $+\infty$ since $\int_{x_m-\delta}^{\infty} e^{-\frac{1}{b\gamma}U'(x_a)(x''-x_a)^2}dx''$ is very small.

$$
\tau(x_2|x_1) = \frac{2}{b}e^{\frac{2}{b\gamma}(U(x_m)-U(x_a))}\left(\int_{x_m-\delta}^{x_m+\delta} e^{\frac{1}{b\gamma}U''(x_m)(x'-x_m)^2}dx'\right)\left(\int_{-\infty}^{+\infty} e^{-\frac{1}{b\gamma}U"(x_a)(x''-x_a)^2}dx''\right)
$$

Can extend now the limits of the x'-integral to $\pm\infty$ and evaluate the integrals

$$
\tau(x_2|x_1) = \frac{2\pi b\gamma}{b}\frac{1}{\sqrt{-U''(x_m)U''(x_a)}}e^{\frac{2}{b\gamma}(U(x_m)-U(x_a))} \tag{27}
$$

One often introduces

$$
\omega_a^2 = U''(x_a) \qquad \omega_m^2 = -U''(x_m) \qquad W = U(x_m) - U(x_a)
$$

with the noise in thermal equilibrium given by $\frac{1}{2}b\gamma = kT$ (cf. (25))

$$\frac{1}{\tau(x_2|x_1)} = \frac{1}{2\pi\gamma}\omega_1\omega_m e^{-\frac{W}{kT}} \tag{28}$$

Notes:

- $\frac{1}{\tau(x_2|x_1)}$ is an escape rate or a reaction rate
- $e^{-\frac{W}{kT}}$ is the Arrhenius factor for activated reactions
- ω_a characterizes the frequency of oscillations in the initial well (if there were no damping). It set the frequency with which the barrier crossing is 'attempted': ω_a small (wide potential well) $\Rightarrow$ few attempts at crossing and small escape rate
- ω_m characterizes the width of the barrier: ω_m small $\Rightarrow$flat top $\Rightarrow$ hard to get across barrier $\Rightarrow$ small escape rate

4.3.4 First Passage Times: Second Approach[14]

To determine higher moments of the first-passage-time another approach is useful.

Consider the probability $G(x,t)$ of a particle to be within the interval $[a,b]$ at time t if it was released at $x \in [a,b]$ at time $t = 0$

$$G(x,t) = \int_a^b P_{1|1}(x',t|x,0)\,dx'$$

Assume the particle leaves the interval at a time T,

$$\text{Prob}(T \geq t) = G(x,t)$$

We seek to derive a differential equation for $G(x,t)$: want to have t and x both on the conditional side of $P_{1|1}$. If the system is translation invariant in time we have

$$P_{1|1}(x',t|x,0) = P_{1|1}(x',0|x,-t)$$

and

$$\frac{\partial}{\partial t}P_{1|1}(x',t|x,0) = \frac{\partial}{\partial t}P_{1|1}(x',0|x,-t) \tag{29}$$

In terms of the initial conditions $P_{1|1}(x',0|x,t)$ satisfies the *backward* Fokker-Planck equation

$$\frac{\partial}{\partial t}P_{1|1}(x',0|x,t) = -A(x)\frac{\partial}{\partial x}P_{1|1}(x',0|x,t) - \frac{1}{2}B(x)\frac{\partial^2}{\partial x^2}P_{1|1}(x',0|x,t)$$

Thus, because of the minus-sign in (29) one gets

$$\frac{\partial}{\partial t}P_{1|1}(x',t|x,0) = A(x)\frac{\partial}{\partial x}P_{1|1}(x',0|x,t) + \frac{1}{2}B(x)\frac{\partial^2}{\partial x^2}P_{1|1}(x',0|x,t)$$

[14] [15, 5.2.7]

Integrating over the first argument of $P_{1|1}(x',t|x,0)$ we get an equation for the survival probability

$$\frac{\partial}{\partial t}G(x,t) = A(x)\frac{\partial}{\partial x}G(x,t) + \frac{1}{2}B(x)\frac{\partial^2}{\partial x^2}G(x,t) \tag{30}$$

Initially the particle is released at x

$$P_{1|1}(x',0|x,0) = \delta(x'-x)$$

yielding the initial condition for $G(x,t)$

$$G(x,0) = \begin{cases} 1 & x \in [a,b] \\ 0 & x \notin [a,b] \end{cases}$$

Boundary conditions at $x = x_B$ with $x_B = a$ or $x_B = b$:

- absorbing boundary

$$\begin{aligned} \text{Prob}(T \geq t) = 0 \qquad & \text{for } x = x_B \\ G(x_B,t) \;=\; & 0 \end{aligned}$$

- reflecting boundary
  it does not matter whether the particle is released at $x_B - \delta x$ or at $x_B + \delta x$

$$\frac{\partial}{\partial x}G(x,t) = 0 \qquad \text{at } x = x_B$$

We would like to get moments of the mean first passage time. More generally

$$\langle f(T)\rangle = \int_0^\infty f(t)P_{esc}(t)dt$$

where $P_{esc}(t)dt$ is the probability that the particle leaves the interval during the interval $[t, t+dt]$.
$P_{esc}(t)dt$ is the amount by which the probability for the particle to be inside the interval decreases during dt

$$P(t)\,dt = -dG(x,t) = -\frac{\partial}{\partial t}G(x,t)\,dt$$

Thus,

$$\langle f(T)\rangle = -\int_0^\infty f(t)\frac{\partial}{\partial t}G(x,t)\,dt$$

and

$$\begin{aligned} \tau_n(x) \equiv \langle T^n\rangle &= -\int_0^\infty t^n\frac{\partial}{\partial t}G(x,t)\,dt = \\ &= n\int_0^\infty t^{n-1}G(x,t)\,dt \end{aligned}$$

The boundary terms vanish since the particle is guaranteed to leave the interval eventually (unless both boundaries are reflecting, of course).

Obtain a differential equation for τ_n by multiplying (30) by t^{n-1} and integrating it over time

$$\int_0^\infty t^{n-1}\frac{\partial}{\partial t}G(x,t)dt = A(x)\frac{\partial}{\partial x}\int_0^\infty t^{n-1}G(x,t)dt + \frac{1}{2}B(x)\frac{\partial^2}{\partial x^2}\int_0^\infty t^{n-1}G(x,t)dt$$

Using

$$\int_0^\infty t^{n-1}\frac{\partial}{\partial t}G(x,t)dt = -\tau_{n-1}(x)$$

we get

$$-n\tau_{n-1}(x) = A(x)\frac{\partial}{\partial x}\tau_n(x) + \frac{1}{2}B(x)\frac{\partial^2}{\partial x^2}\tau_n(x,) \tag{31}$$

For $n = 1$ this reduces to

$$-1 = A(x)\frac{\partial}{\partial x}\tau_1(x) + \frac{1}{2}B(x)\frac{\partial^2}{\partial x^2}\tau_1(x) \tag{32}$$

which is again the Dynkin equation (26).

Example:
$A(x) = -U'(x)/\gamma$ and $B(x) = b$ with an absorbing boundary at $x = 0$ and a reflecting boundary at $x = -L$

Introducing $v(x) = \frac{\partial}{\partial x}\tau_n(x)$ we get

$$v' - \frac{2}{b\gamma}U'v = -\frac{2}{b}n\tau_{n-1}$$

$$\frac{d}{dx}\left(e^{-\frac{2}{b\gamma}U}v\right) = -\frac{2}{b}n\tau_{n-1}e^{-\frac{2}{b\gamma}U}$$

Then

$$v(x) = e^{\frac{2}{b\gamma}U(x)}\left\{\int_{-L}^{x} -\frac{2}{b}n\tau_{n-1}e^{-\frac{2}{b\gamma}U(x')}dx' + C\right\}$$

The reflecting boundary condition requires that $v'(x) = 0$ for $x = -L \Rightarrow C = 0$.

Thus

$$\tau_n(x) = \int_{\hat{x}}^{x} e^{\frac{2}{b\gamma}U(x')}\int_{-L}^{x'} -\frac{2}{b}n\tau_{n-1}e^{-\frac{2}{b\gamma}U(x'')}dx''$$

To satisfy the absorbing boundary condition at $x = 0$ we need $\hat{x} = 0$

$$\tau_n(x) = \frac{2}{b}n\int_x^0 e^{\frac{2}{b\gamma}U(x')}\left\{\int_{-L}^{x'}\tau_{n-1}e^{-\frac{2}{b\gamma}U(x'')}dx''\right\}dx'$$

Note:

- for two absorbing boundaries the solution can also be given in terms of similar integrals, but it is much more complicated [15, 5.2.7]

4.4 The Backward Fokker-Planck Equation

The backward Fokker-Planck equation describes the dependence of $P(y_2, t_2|y_1, t_1)$ on the initial condition $(y_{1,}t_1)$.

To invoke again the Chapman-Kolmogorov equation consider three times $t_1 - \Delta t$, t_1, and t_2, this time the small increment is in the initial time,

$$P(t_2, y_2|t_1 - \Delta t, y_1) = \int P(t_2, y_2|t_1, z)\, P(t_1, z|t_1 - \Delta t, y_1)\, dz$$

For simplicity assume this process does not include jump components, i.e. $P(t_2, y_2|t_1, y_1)$ is smooth. Therefore

during the small interval $[t_1 - \Delta t, t_1]$ the particle does not get very far and z is near y_1: expand in $z - y_1$

$$\begin{aligned} P(t_2, y_2|t_1, z) &= P(t_2, y_2|t_1, y_1) + (z - y_1)\frac{\partial}{\partial y_1}P(t_2, y_2|t_1, y_1) + \\ &\quad + \frac{1}{2}(z - y_1)^2 \frac{\partial^2}{\partial y_1^2}P(t_2, y_2|t_1, y_1) + \mathcal{O}\left((z - y_1)^3\right) \end{aligned}$$

Insert this expansion for small Δt

$$\begin{aligned} P(t_2, y_2|t_1 - \Delta t, y_1) &= P(t_2, y_2|t_1, y_1) \underbrace{\int P(t_1, z|t_1 - \Delta t, y_1)\, dz}_{=1} + \\ &\quad + \frac{\partial}{\partial y_1}P(t_2, y_2|t_1, y_1) \underbrace{\int (z - y_1)\, P(t_1, z|t_1 - \Delta t, y_1)\, dz}_{\to \Delta t A(y_1, t_1 - \Delta t)} \\ &\quad + \frac{1}{2}\frac{\partial^2}{\partial y_1^2}P(t_2, y_2|t_1, y_1) \underbrace{\int (z - y_1)^2\, P(t_1, z|t_1 - \Delta t, y_1)\, dz}_{\to \Delta t B(y_1, t_1 - \Delta t)} + \\ &\quad + \mathcal{O}\left(\int (z - y_1)^3\, P(t_1, z|t_1 - \Delta t, y_1)\, dz\right) \end{aligned}$$

Thus, assuming P, A, and B are smooth in t we get

$$\frac{1}{\Delta t}\left(P(t_2, y_2|t_1, y_1) - P(t_2, y_2|t_1 - \Delta t, y_1)\right) \to \frac{\partial}{\partial t_1}P(t_2, y_2|t_1, y_1)$$

and

$$\frac{\partial}{\partial t_1}P(t_2, y_2|t_1, y_1) = -A(y_1, t_1)\frac{\partial}{\partial y_1}P(t_2, y_2|t_1, y_1) - \frac{1}{2}B(y_1, t_1)\frac{\partial^2}{\partial y_1^2}P(t_2, y_2|t_1, y_1)$$

Notes:

- this *backward* Fokker-Planck equation describes the dependence of the transition probability on the *initial* conditions

- while in the forward Fokker-Planck equation the drift and diffusion terms are inside the derivative, these terms are outside the derivative in the backward Fokker-Planck equation
- to be well posed, the backward Fokker-Planck equation needs a *final condition* rather than an initial condition.
- for processes that also exhibit jumps one can derive a backward derivative Chapman-Kolmogorov equation (cf. ch.3.6 in [15])

5 Langevin Equation[15]

So far we dealt with equations for the probability distributions or transition probabilities (Chapman-Kolmogorov, Fokker-Planck).

Consider now an equation directly for the stochastic variable itself

Approach:

1. start with the macroscopic equation of motion
2. add "suitable" noise term
3. adjust noise strength "suitably"

Consider the Langevin equation for Brownian motion

$$\frac{dV}{dt} = -\gamma V + L(t)$$

where $L(t)$ represents the effect of the many molecules hitting the Brownian particle:

- the average is meant to be contained in the macroscopic equation

$$\langle L(t) \rangle = 0$$

- the kicks by the molecules are very brief and they are uncorrelated for different times

$$\langle L(t)L(t') \rangle = \Gamma\delta(t - t')$$

  we expect that we can determine the noise strength Γ from a comparison with the distribution in thermondynamic equilibrium

- higher moments of $L(t)$:
  Assume the process is Gaussian, i.e. all higher cumulants (cf. (1)) vanish.
  For a Gaussian process we know (cf. (9,10))

$$\phi(\{s(t)\}) = \exp\left\{ i \int s(t) \langle L(t) \rangle \, dt - \frac{1}{2} \int \int s(t) \left(\langle L(t)L(t') \rangle - \langle L(t) \rangle \langle L(t') \rangle \right) s(t') dt \, dt^{'} \right\}$$

[15] [28], IX.1-3

thus

$$\ln\left(\phi(\{s(t\})\right) = i \int s(t) \left\langle L(t) \right\rangle dt - \frac{1}{2} \int \int s(t) \left\langle \left\langle L(t) L(t') \right\rangle \right\rangle s(t') dt\, dt'$$

which does not contain any contributions higher than quadratic in $s(t)$. Thus, all higher cumulants vanish.

To obtain $\langle V(t)\rangle$ and $\langle V(t)V(t')\rangle$ determine explicit solution of the Langevin equation

$$V(t) = V_0 e^{-\gamma t} + \int_0^t e^{-\gamma(t-t')} L(t') dt'$$

Here $V(t)$ is still a stochastic process. To obtain this equation one could consider a specific realization of the noise term $L(t)$ and determine the solution $v(t)$ for that realization that satisfies the initial condition v_0 . That amounts to a realization of the stochastic process $V(t)$.

Mean:

$$\langle V(t)\rangle = v_0 e^{-\gamma t} + \int_0^t e^{-\gamma(t-t')} \langle L(t')\rangle dt' = v_0 e^{-\gamma t} \tag{33}$$

Second moment (assuming $t_2 \geq t_1$):

$$\begin{aligned}
\langle V(t_1)V(t_2)\rangle &= v_0^2 e^{-\gamma(t_1+t_2)} + v_0 e^{-\gamma t_1} \int_0^{t_2} e^{-\gamma(t-t')} \langle L(t')\rangle dt' + v_0 e^{-\gamma t_2} \int_0^{t_1} e^{-\gamma(t-t')} \langle L(t')\rangle dt' \\
&\quad + \int_0^{t_1} \int_0^{t_2} e^{-\gamma(t_1-t')} e^{-\gamma(t_2-t")} \langle L(t')L(t")\rangle dt' dt" \\
&= v_0^2 e^{-\gamma(t_1+t_2)} + \Gamma e^{-\gamma(t_1+t_2)} \int_0^{t_1} dt' \left\{ \int_0^{t'+\epsilon} e^{\gamma(t'+t")} \delta(t'-t") dt" + \int_{t'+\epsilon}^{t_2} e^{\gamma(t'+t")} \delta(t'-t") dt" \right\} \\
&= v_0^2 e^{-\gamma(t_1+t_2)} + \Gamma e^{-\gamma(t_1+t_2)} \int_0^{t_1} dt' e^{2\gamma t'} \\
&= v_0^2 e^{-\gamma(t_1+t_2)} + \frac{\Gamma}{2\gamma} \left(e^{-\gamma(t_1+t_2)} \left(e^{2\gamma t_1} - 1 \right) \right) \\
\langle V(t_1)V(t_2)\rangle &= \left(v_0^2 - \frac{\Gamma}{2\gamma} \right) e^{-\gamma(t_1+t_2)} + \frac{\Gamma}{2\gamma} e^{-\gamma(t_2-t_1)}
\end{aligned} \tag{34}$$

To compare with the equilibrium solution: $t_{1,2} \to \infty$ with $t_2 - t_1 = \tau$

$$\langle V(t_1)V(t_2)\rangle \to \frac{\Gamma}{2\gamma} e^{-\gamma\tau} \tag{35}$$

In equilibrium one has

$$\frac{1}{2} M \langle V^2 \rangle = \frac{1}{2} kT$$

Using (35) one obtains

$$\langle V(t)^2 \rangle = \frac{\Gamma}{2\gamma} = \frac{kT}{M} \tag{36}$$

and

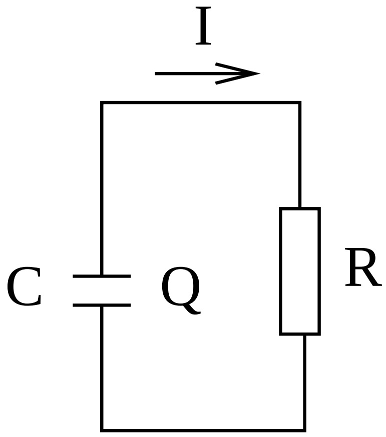
$$\langle V(t_1)V(t_2)\rangle = \left(v_0^2 - \frac{kT}{M}\right) e^{-\gamma(t_1+t_2)} + \frac{kT}{M} e^{-\gamma(t_2-t_1)} \tag{37}$$

Note:

- (36) is again a fluctuation-dissipation relation relating the fluctuations (Γ) with the dissipation γ.
- it would be very difficult to determine the noise strength directly from the molecular interactions $\Rightarrow$ the success of the Langevin approach relies partially on this comparison with equilibrium statistical mechanics.

Example:

Noise in RC-circuit:



macroscopic equation (Kirchhoff's laws):

$$\frac{dQ}{dt} = -I = -\frac{U}{R} \underbrace{=}_{C=\frac{Q}{U}} -\frac{Q}{RC}$$

the macroscopic equation contains dissipation $\Rightarrow$ there will also be fluctuations

$$\frac{dQ}{dt} = -\frac{1}{RC}Q + L(t)$$

Noise strength: in thermal equilibrium the energy stored in the capacitor is $\frac{1}{2}kT$

The work done by bringing the charge dQ onto the capacitor with voltage $U = E \cdot d$ is

$$dW = EdQ \cdot d = UdQ$$

The energy stored is

$$W = \int UdQ = \int \frac{Q}{C}dQ = \frac{1}{2}\frac{Q^2}{C}$$

$$\left\langle \frac{1}{2}\frac{Q^2}{C} \right\rangle = \frac{1}{2C}\left\langle Q^2 \right\rangle = \frac{1}{2}kT$$

From the Langevin equation we obtain (cf. (37))

$$\left\langle Q^2 \right\rangle = \frac{\Gamma}{2} RC$$

thus

$$\Gamma = \frac{2kT}{R}$$

Note:

- the noise depends only on the resistor $\Rightarrow$ the source of the noise is in the resistor (collision of conduction electrons with the atoms)
- the noise leads to a fluctuating current

  $$\left\langle \delta I(t)^2 \right\rangle = \left\langle L(t)^2 \right\rangle = \Gamma = \frac{2kT}{R}$$

  which decreases with increasing resistivity
  Alternatively, one can say the noise leads to a fluctuation voltage across the resistor

  $$\left\langle \delta U(t)^2 \right\rangle = R^2 \left\langle \delta I(t)^2 \right\rangle = 2kTR$$

  which increases with increasing resistivity.

5.1 Relation between Langevin Equation and Fokker-Planck Equation

Because in the Langevin equation the noise is δ-correlated in time it describes a Markov process $\Rightarrow$ expect that the same process also can be described by a Fokker-Planck equation

More precisely

$$y(t+\tau) = \lim_{\epsilon \to 0} \int_0^{t-\epsilon} -y + L(t')dt' + \int_t^{t+\tau} -y + L(t')dt'$$

i.e.

$$y(t+\tau) - y(t) = \int_t^{t+\tau} -y(t') + L(t')dt'$$

Since $L(t)$ and $L(t')$ are independent of each other for $t \neq t'$ we have that $y(t+\tau) - y(t)$ is independent of $y(t")$ for all $t" < t$, i.e. for all $\tau > 0$ the value $y(t+\tau)$ depends only on $y(t)$ and not on any previous values.

5.1.1 Linear Langevin Equation

Consider the linear Langevin equation

$$\frac{dV}{dt} = -\gamma V + L(t) \tag{38}$$

with Gaussian noise $L(t)$.

The velocity $V(t)$ is given by a sum over the noise term at different times

$$V(t) = V_0 e^{-\gamma t} + \int_0^t e^{-\gamma(t-t')} L(t') dt'$$

which are all Gaussian.

Thus

- $V(t)$ is also a Gaussian process
- for a complete characterization of the process $V(t)$ only $\langle V(t)\rangle$ and $\langle V(t_1)V(t_2)\rangle$ are needed (cf. (33,37)

The mean and second moments are the same as those for the Fokker-Planck equation

$$\frac{\partial P(v,t|v_0,t_0)}{\partial t} = \gamma \left\{ \frac{\partial}{\partial v}\left(vP(v,t|v_0,t_0)\right) + \frac{kT}{M}\frac{\partial^2}{\partial v^2} P(v,t|v_0,t_0) \right\} \tag{39}$$

(see homework). Therefore the Langevin equation (38) is equivalent to the Fokker-Planck equation (39).

5.1.2 Nonlinear Langevin Equation

Now consider

$$\frac{dy}{dt} = f(y) + g(y)L(t)$$

Through $g(y)$ the noise strength depends on the state of the system: multiplicative noise.

Notes:

- formally one can rewrite the Langevin equation with multiplicative noise as a Langevin equation for a new variable $\tilde{y}$ with additive noise

  $$\tilde{y} = F(y) \equiv \int^y \frac{1}{g(y')} dy'$$

  then

  $$\frac{d\tilde{y}}{dt} = \frac{1}{g(y)}\frac{dy}{dt} = \frac{f(y)}{g(y)} + L(t)$$

  i.e.

  $$\frac{d\tilde{y}}{dt} = \frac{f(F^{-1}(\tilde{y}))}{g(F^{-1}(\tilde{y}))} + L(t).$$

  For instance

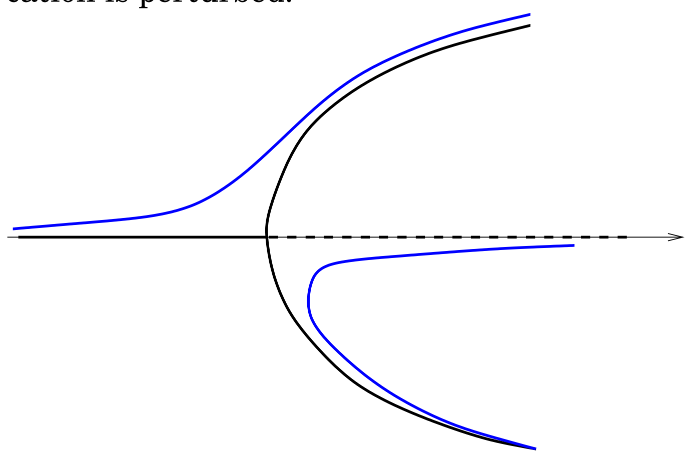
  $$\frac{dy}{dt} = -y + yL(t)$$

  becomes with $\tilde{y} = \ln y$ (assuming $y > 0$)

  $$\frac{d\tilde{y}}{dt} = -1 + L(t)$$

- There are important qualitative differences between additive noise and multiplicative noise, which can of course not be removed by such a transformation

  1. Consider
  $$\frac{dy}{dt} = ay - y^3 + L(t)$$
  Among many possibilities, this equation could describe the buckling of a beam under a longitudinal load with y denoting the amount of buckling. Without noise this system exhibits a pitchfork bifurcation
  $$y_\infty = \begin{cases} 0 & \text{for } a \le 0 \\ \pm\sqrt{a}, 0 & \text{for } a > 0 \end{cases}$$
  with $\pm\sqrt{a}$ representing a buckling to the left or right, respectively.
  Here the noise terms breaks the reflection symmetry $y \to -y$ (in this example it represents a fluctuating forcing transverse to the beam) and the pitchfork bifurcation is perturbed.

  2. Consider
  $$\frac{dy}{dt} = ay - y^3 + yL(t)$$
  Now the fluctuation force modifies a, i.e. the longitudinal load. The reflection symmetry is preserved, the pitchfork bifurcation is still perfect, i.e. $y = 0$ is a solution for all values of a, i.e. the unbuckled state exists for all a and noise strengths.

- If $f(y)$ is nonlinear or $g(y)$ not constant then y is not a Gaussian process even if $L(t)$ is Gaussian.

To obtain the Fokker-Planck equation we need $A(y)$ and $B(y)$

$$A(y) = \lim_{\tau\to 0} \frac{1}{\tau} \int_{|y-y_0|<\epsilon} (y - y_0) T_\tau(y|y_0)dy = \lim_{\tau\to 0} \frac{1}{\tau} \langle y(\tau) - y_0 \rangle$$
$$B(y) = \lim_{\tau\to 0} \frac{1}{\tau} \left\langle (y(\tau) - y_0)^2 \right\rangle$$

with y_0 being the specified (non-random) initial condition at $t = 0$.

To determine these expectation values integrate the Langevin equation[16] from $t = 0$ to

[16] For simplicity we assume that the process is stationary.

$t=\tau \ll 1$

$$y(\tau)-y_0=\int_0^\tau f(y(t))dt+\int_0^\tau g(y(t))L(t)dt$$

Expand

$$f(y)=f(y_0)+(y(\tau)-y_0)\,f'(y_0)+\ldots$$
$$g(y)=g(y_0)+(y(\tau)-y_0)\,g'(y_0)+\ldots$$

Then

$$\begin{aligned} y(\tau)-y_0 &= f(y_0)\tau+f'(y_0)\int_0^\tau (y(t)-y_0)\,dt+\ldots \\ &\quad +g(y_0)\int_0^\tau L(t)dt+g'(y_0)\int_0^\tau (y(t)-y_0)\,L(t)dt+\ldots \end{aligned}$$

Expand again in the integrand

$$\begin{aligned} y(\tau)-y_0 &= f(y_0)\tau+f'(y_0)\int_0^\tau \left\{f(y_0)t+\ldots+g(y_0)\int_0^t L(t')dt'+\ldots\right\}dt+\ldots \\ &\quad +g(y_0)\int_0^\tau L(t)dt+g'(y_0)\int_0^\tau \left\{f(y_0)t+\ldots+g(y_0)\int_0^t L(t')dt'+\ldots\right\}L(t)dt+\ldots \end{aligned}$$

We are interested in the average.

$$\langle L(t)\rangle=0 \qquad \langle L(t)L(t')\rangle=\Gamma\delta(t-t')$$

Thus

$$\begin{aligned} \langle y(\tau)-y_0\rangle &= f(y_0)\tau+f'(y_0)f(y_0)\underbrace{\int_0^\tau tdt}_{\mathcal{O}(\tau^2)}+\ldots+f'(y_0)g(y_0)\int_0^\tau\int_0^t \underbrace{\langle L(t')\rangle}_{=0}dt'+\ldots \\ &\quad +g(y_0)\cdot 0+g'(y_0)g(y_0)\int_0^\tau\int_0^t \underbrace{\langle L(t')L(t)\rangle}_{=\Gamma\delta(t-t')}dt'dt+\ldots \end{aligned}$$

$$\langle y(\tau)-y_0\rangle=f(y_0)\tau+g'(y_0)g(y_0)\Gamma\int_0^\tau\int_0^t \delta(t-t')dt'dt+\ldots$$

What is $\int_0^t \delta(t-t')dt'$? The question arises because the δ-function is located at end of the integration interval.

It seems reasonable to consider the δ- correlation as the limit of a a process with finite, but short correlation time $\Rightarrow \delta(t-t')$ is the limit of some symmetric, sharply peaked function, e.g.

$$\delta_\epsilon(t-t')=\begin{cases} \frac{1}{\epsilon} & -\frac{\epsilon}{2}\le t-t'\le\frac{\epsilon}{2} \\ 0 & \text{otherwise} \end{cases}$$

If we take the limit $\epsilon\to 0$ only in the very end, after all the integrals are taken one gets for any smooth function $h(t)$

$$\int_0^t h(t')\delta(t-t')dt'=\frac{1}{2}h(t)$$

Then

$$\langle y(\tau) - y_0 \rangle = f(y_0)\tau + g'(y_0)g(y_0)\Gamma\frac{1}{2}\tau + \dots$$

and

$$A(y) = f(y) + g'(y)g(y) \tag{40}$$

Analogously one gets

$$\left\langle (y(\tau) - y_0)^2 \right\rangle = [g(y_0)]^2 \int_0^\tau dt \int_0^\tau dt' \, \langle L(t)L(t') \rangle + \mathcal{O}(\tau^2)$$

implying

$$B(y) = \Gamma \, [g(y)]^2 \tag{41}$$

Thus we get the Fokker-Planck equation

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial y}\left[\left(f(y) + \frac{1}{2}\Gamma g(y)g'(y)\right) P\right] + \frac{1}{2}\Gamma\frac{\partial^2}{\partial y^2}\left[(g(y))^2 P\right] \tag{42}$$

Using

$$\frac{\partial}{\partial y}\left(g^2 P\right) = 2gg'P + g^2\frac{\partial}{\partial y}P = gg'P + g\frac{\partial}{\partial y}(gP)$$

one can rewrite the Fokker-Planck equation as

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial y}\left[f(y)P\right] + \frac{1}{2}\Gamma\frac{\partial}{\partial y}\left[g(y)\frac{\partial}{\partial y}\left[g(y)P\right]\right]$$

Note:

- If $g(y)$ is not constant the noise term contributes to the drift term $A(y)$: "noise-induced drift", i.e. even if $f(y) = 0$ the average $\langle y \rangle$ will be time-dependent and is driven purely by the noise term and the dependence of its effect on the state y.
- The noise-induced drift points to a problem that can arise when one wants to identify the correct Fokker-Planck equation:
  - Systems with *external noise*, i.e. the macroscopic dynamics is separate from the noise (one could imagine the noise can be turned off), e.g. a transmission line into which a noisy signal is fed, a bridge under the random force of cars driving on it:
    - macroscopic dynamics $f(y)$ is known in the absence of noise and the noise, which conceptually can be turned on or off, can modify the drift term
  - System with *internal noise*, e.g. Brownian motion, chemical reactions, viscous fluid flow. Here the macroscopic motion arises from the noisy microscopic motion, the noise cannot be turned off. Therefore the macroscopic dynamics $f(y)$ cannot be separated from the noise.
    - when the noise affects the system only additively the drift term $A(y)$ is not modified by the noise and the Langevin approach should be fine (viscosity in fluid flow acts only on the linear term, the nonlinear term is the advection term).

      In particular, if the dynamics of the system are also linear the mean satisfies the macroscopic equation.

* when the noise acts nonlinearly then it is not clear what to take for $f(y)$ because $f(y)$ already contains aspects of the noise. E.g., in chemical reactions the nonlinear terms represent reactions of molecules, which are the cause of the noise $\Rightarrow$ the nonlinear Langevin equation is then most likely not appropriate. One would have to start from the master equation and obtain suitable reductions [28, Chap. X].

5.2 Mathematical Considerations: Ito vs Stratonovich[17]

Mathematically we are having a problem: consider the simplest Langevin equation

$$\frac{dy}{dt} = L(t)$$

then

$$y(t) = \int_0^t L(t')dt'$$

Being a continuous Markov process $y(t)$ can be described by a Fokker-Planck equation. From (40,41) we have

$$\frac{\partial P}{\partial t} = \frac{1}{2}\Gamma\frac{\partial^2}{\partial y^2}P$$

i.e. $y(t)$ is the Wiener process $W(t)$ (cf. (6,7)).

The Wiener process is continuous *but nowhere differentiable,* i.e. $\frac{dy}{dt}$ does not exist!

To avoid using $L(t)$ itself use an integral formulation

$$W(\tau) = \int_t^{t+\tau} L(t')dt'$$

Write $dW(t)$ instead of $L(t)dt$

The Langevin equation becomes

$$dy = f(y)dt + g(y)dW(t)$$

or in integral form

$$y(t+\tau) = y(t) + \int_t^{t+\tau} f(y(t'))dt' + \underbrace{\int_t^{t+\tau} g(y(t'))dW(t')}_{\text{Riemann-Stieljes integral}}$$

The Riemann-Stieltjes integral is defined for general $u(t)$ and $v(t)$ as

$$\int_t^{t+\tau} u(t')dv(t') = \lim_{N\to\infty}\sum_{i=1}^{N} u(t_i^*)\left[v(t_{i+1}) - v(t_i)\right] \qquad \text{with } t_i^* \in [t_i, t_{i+1}]$$

Notes:

[17] [15, 4.]

- For $v(t) = t$ one recovers directly the Riemann integral $\lim_{N\to\infty} \sum_{i=1}^{N} u(t_i^*)\Delta t_i = \int_t^{t+\tau} u(t')dt'$.
- For 'nice' functions with bounded variation the value of the integral does not depend on the choice of $t_i^* \in [t_i, t_{i+1}]$.
- the Wiener process $W(t)$ has *unbounded variation*:

$$\sum_{i=1}^{N} |W(t_{i+1}) - W(t_i)| \to \infty \qquad \text{for } N \to \infty.$$

  Then the integral does depend on the choice of t_i^*, e.g.,

$$\langle S_N \rangle \equiv \left\langle \sum_{i=1}^{N} W(t_i^*) \left(W(t_{i+1}) - W(t_i)\right) \right\rangle = \sum_{i=1}^{N} \left\{ \langle W(t_i^*) W(t_{i+1}) \rangle - \langle W(t_i^*) W(t_i) \rangle \right\}$$

  Using (cf. (23))

$$\langle W(t_1) W(t_2) \rangle = \min(t_1, t_2)$$

  one gets

$$\langle S_N \rangle = \sum_{i=1}^{N} (t_i^* - t_i)$$

  choosing

$$t_i^* = \alpha t_{i+1} + (1-\alpha) t_i$$

  one gets

$$\langle S_N \rangle = \sum_{i=1}^{N} \alpha (t_{i+1} - t_i) = \alpha \tau$$

Two definitions for the stochastic integral

- Stratonovich ($\alpha = \frac{1}{2}$)

$$\int u(t)dv(t) = \lim_{N\to\infty} \sum_{i=1}^{N} \frac{1}{2} \left(u(t_i) + u(t_{i+1})\right) \left(v(t_{i+1}) - v(t_i)\right)$$

- Ito ($\alpha = 0$)

$$\int u(t)dv(t) = \lim_{N\to\infty} \sum_{i=1}^{N} u(t_i) \left(v(t_{i+1}) - v(t_i)\right)$$

Notes:

- for $u = g$ and $dv = dW$ this means
  - for Stratonovich integral the prefactor $g(y)$ is averaged across the kick
  - for Ito integral the prefactor $g(y)$ is determined before the kick

- the limit $\lim_{N\to\infty}$ is to be understood as a mean-squared limit $ms - \lim_{N\to\infty}$

$$ms - \lim_{N\to\infty} Y_n = Y \qquad \Leftrightarrow \qquad \lim_{N\to\infty} \left\langle (Y_n - Y)^2 \right\rangle = 0 \tag{43}$$

Reconsider the derivation of the terms $A(y)$ and $B(y)$ of the Fokker-Planck equation

$$\begin{aligned}
y(\tau) - y_0 &= \int_0^\tau f(y(t'))dt' + \int_0^\tau g(y(t')dW(t') \\
&= f(y_0)\tau + \mathcal{O}(\tau^2) + g(y_0) \underbrace{\int_0^\tau dW(t')}_{W(\tau)-W(0)=W(\tau)} + g'(y_0) \int_0^\tau (y(t') - y_0)\, dW(t') \\
&= f(y_0)\tau + g(y_0)W(\tau) + g'(y_0) \int_0^\tau \left[f(y_0)t' + g(y_0)W(t') \right] dW(t')
\end{aligned}$$

Since the ensemble average of $dW(t)$ vanishes one has also

$$\int_0^\tau t' \langle dW(t') \rangle = 0$$

and

$$\langle y(\tau) - y_0 \rangle = f(y_0)\tau + g'(y_0)g(y_0) \left\langle \int_0^\tau W(t')dW(t') \right\rangle$$

and

$$A(y) = f(y) + g(y)g'(y)\frac{1}{\tau} \left\langle \int_0^\tau W(t')dW(t') \right\rangle$$

Evaluate the stochastic integral:

1. Stratonovich:

$$\begin{aligned}
\left\langle S \int_0^\tau W(t')dW(t') \right\rangle &= \lim_{N\to\infty} \sum_{i=1}^{N} \frac{1}{2} \left\langle \left(W(t_{i+1}) + W(t_i)\right) \left(W(t_{i+1}) - W(t_i)\right) \right\rangle \\
&= \lim_{N\to\infty} \frac{1}{2} \sum_{i=1}^{N} \left\langle W(t_{i+1})^2 \right\rangle - \left\langle W(t_i)^2 \right\rangle \\
&= \frac{1}{2} \left(\left\langle W(\tau)^2 \right\rangle - \left\langle W(0)^2 \right\rangle \right)
\end{aligned}$$

We computed $\langle W(t)^2 \rangle$ for the Wiener process in the context of the Rayleigh particle (cf. 23)), $W(t)$ is the position of the Brownian particle at time t,

$$\langle W(\tau)W(\tau') \rangle = \min(\tau, \tau').$$

Therefore

$$\left\langle S \int_0^\tau W(t')dW(t') \right\rangle = \frac{1}{2}\tau = \left\langle \frac{1}{2} W(\tau)^2 \right\rangle \tag{44}$$

and

$$A_S(y) = f(y) + \frac{1}{2} g(y)g'(y)$$

2. Ito:

$$\begin{aligned}
\left\langle I \int_0^\tau W(t')dW(t') \right\rangle &= \lim_{N\to\infty} \sum_{i=1}^{N} \langle W(t_i)\left(W(t_{i+1}) - W(t_i)\right)\rangle \\
&= \lim_{N\to\infty} \sum_{i=1}^{N} \langle W(t_{i+1})W(t_i)\rangle - \langle W(t_i)^2\rangle \\
&= \lim_{N\to\infty} \sum_{i=1}^{N} \{t_i - t_i\} = 0
\end{aligned}$$

Therefore

$$A_I(y) = f(y)$$

Notes:

- In the Stratonovich interpretation of the stochastic integral the same noise-induced drift arises as we found when interpreting the δ-correlation as a symmetric *smooth* correlation function in the limit of vanishing correlation time.
- In the Ito interpretation no noise-induced drift term arises.
- By adjusting the drift term appropriately, both interpretations can be used for the same process.
- In the Stratonovich interpretation the usual integration rules hold (cf. (44))

$$\left\langle \int w dw \right\rangle = \left\langle \frac{1}{2} w^2 \right\rangle$$

- In the Ito interpretation we found a new integration rule

$$\left\langle \int w dw \right\rangle = 0$$

- If $g(y)$ is constant, i.e. for additive noise, there is no difference between the two approaches.

5.3 Ito Stochastic Integrals[18]

To use the Ito interpretation of the stochastic integrals we need to evaluate integrals like

$$\int G(t')dW(t')$$

where $G(t)$ is a non-anticipating function.

Define:

[18] [15, Chap.4.2]

- $G(t)$ is *non-anticipating* if it is statistically independent of $W(s) - W(t)$ for $s > t$.

The integrals are evaluated in the mean-square limit (cf. (43)),

$$\int G(t')dW(t') = ms - \lim_{N\to\infty} \sum_{i=1}^{N} G(t_{i-1}) \left(W(t_i) - W(t_{i-1})\right)$$

Note:

- the stochastic integral $\int G(t')dW(t')$ is a different kind of integral than $\int G(t')dt'$ and there is in general no connection between the two.
- the integral $\int G(t')dW(t')$ depends on the process W which is here assumed to be the Wiener process.

We need a number of properties of stochastic integrals.

a) $dW^2 = dt$ **and** $dW^{2+n} = 0$ **for** $n > 0$

The precise statements are

$$\int_0^t G(t') \left(dW(t')\right)^2 = ms - \lim_{N\to\infty} \sum_{i=1}^{N} G_{i-1} \left(W_i - W_{i-1}\right)^2 = \int_0^t G(t')dt' \tag{45}$$

and for $n > 0$

$$\int_0^t G(t') \left(dW(t')\right)^{2+n} = ms - \lim_{N\to\infty} \sum_{i=1}^{N} G_{i-1} \left(W_i - W_{i-1}\right)^{2+n} = 0 \tag{46}$$

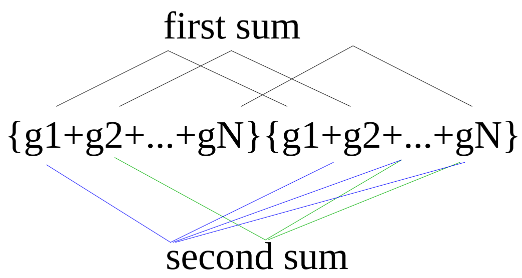
Here $G_i = G(t_i)$ and $W_i = W(t_i)$.

Thus:

- dW is a differential of order $\frac{1}{2}$
- $dW^2 = dt$

To proof this identity consider the appropriate limit

$$\lim_{N\to\infty} \left\langle \left\{ \sum_{i=1}^{N} G_{i-1} \left[\underbrace{\Delta W_i^2}_{(W_i - W_{i-1})^2} - \Delta t_i \right] \right\}^2 \right\rangle = \lim_{N\to\infty} \left\langle \sum_{i=1}^{N} G_{i-1}^2 \left(\Delta W_i^2 - \Delta t_i\right)^2 + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} G_{i-1} G_{j-1} \left(\Delta W_i^2 - \Delta t_i\right) \left(\Delta W_j^2 - \Delta t_j\right) \right\rangle$$



[19]

Term with single sum:

G is non-anticipating $\Rightarrow G_{i-1}$ and ΔW_i are statistically independent

$$\left\langle G_{i-1}^2 \left(\Delta W_i^2 - \Delta t_i\right)^2 \right\rangle = \left\langle G_{i-1}^2 \right\rangle \left\langle \left(\Delta W_i^2 - \Delta t_i\right)^2 \right\rangle$$

Since W is Gaussian distributed the fourth (and higher) cumulants vanishes

$$0 = \left\langle\left\langle \Delta W_i^4 \right\rangle\right\rangle = \left\langle \Delta W_i^4 \right\rangle - 3 \left\langle \Delta W_i^2 \right\rangle^2$$

using also that all terms with odd powers of ΔW_i vanish.

Thus

$$\begin{aligned} \left\langle \Delta W_i^4 \right\rangle &= 3 \left\langle \Delta W_i^2 \right\rangle^2 = 3 \left\langle (W_i - W_{i-1})^2 \right\rangle^2 = \\ &\underbrace{=}_{\text{Wiener process}} 3 \left\{ t_i - 2t_{i-1} + t_{i-1} \right\}^2 = 3 \Delta t_i^2 \end{aligned}$$

and

$$\left\langle G_{i-1}^2 \left(\Delta W_i^2 - \Delta t_i\right)^2 \right\rangle = \left\langle G_{i-1}^2 \right\rangle \left\{ 3\Delta t_i^2 - 2\Delta t_i^2 + \Delta t_i^2 \right\} = 2 \left\langle G_{i-1}^2 \right\rangle \Delta t_i^2$$

Therefore

$$\lim_{N\to\infty} \left\langle \sum_{i=1}^{N} G_{i-1}^2 \left(\Delta W_i^2 - \Delta t_i\right)^2 \right\rangle = 2 \lim_{N\to\infty} \left\{ \underbrace{\sum_{i=1}^{N} \left\langle G_{i-1}^2 \right\rangle \Delta t}_{=\mathcal{O}(1)} \cdot \Delta t \right\} = 0$$

Term with double sum:

$j > i \Rightarrow G_{i-1} G_{j-1} \left(\Delta W_i^2 - \Delta t_i\right)$ is statistically independent of $\Delta W_j^2 - \Delta t_j$

$$\left\langle G_{i-1} G_{j-1} \left(\Delta W_i^2 - \Delta t_i\right) \left(\Delta W_j^2 - \Delta t_j\right) \right\rangle = \left\langle G_{i-1} G_{j-1} \left(\Delta W_i^2 - \Delta t_i\right) \right\rangle \underbrace{\left\langle \Delta W_j^2 - \Delta t_j \right\rangle}_{=0}$$

Thus

$$\lim_{N\to\infty} \left\langle \left\{ \sum_{i=1}^{N} G_{i-1} \left[\underbrace{\Delta W_i^2}_{(W_i - W_{i-1})^2} - \Delta t_i \right] \right\}^2 \right\rangle = 0$$

[19] or simpler visualization via sum over all elements in a symmetric matrix: first sum consists of the diagonal terms, second (double) sum is over the upper triangle.

The proof of $dW^{2+n} = 0$ is analagous. It makes use of the higher cumulants and is more cumbersome.

b) Integration of polynomials

$$d\left(W\right)^{n+1} = \left(W + dW\right)^{n+1} - W^{n+1} = \sum_{m=1}^{n+1} \left(\begin{array}{c} n+1 \\ m \end{array}\right) W^{n-m+1} dW^m$$

with $dW^{2+k} = 0$ for $k > 0$ one gets

$$d\left(W\right)^{n+1} = (n+1)W^n dW + \frac{1}{2}(n+1)n\, W^{n-1} \underbrace{\left(dW\right)^2}_{=dt}$$

Integrating the previous equation we get

$$\int_0^t W^n(t')dW(t') = \frac{1}{n+1}\left(W(t)^{n+1} - W(0)^{n+1}\right) - \frac{1}{2}n\int_0^t W(t')^{n-1}dt' \qquad \textbf{(47)}$$

Example: $n = 1$

$$\int_0^t W(t')dW(t') = \frac{1}{2}\left(W(t)^2 - W(0)^2\right) - \frac{1}{2}t$$

therefore

$$\left\langle \int_0^t W(t')dW('t) \right\rangle = \frac{1}{2}t - \frac{1}{2}t = 0$$

as before.

c) Differentiation

When taking derivatives (i.e. when expanding) one has to keep in mind that $(dW)^2$ is of the same order as dt

$$\begin{aligned} df(W(t),t) &= f(W(t) + dW, t + dt) - f(W(t),t) = \\ &= \frac{\partial}{\partial W} f\, dW + \frac{1}{2}\frac{\partial^2}{\partial W^2} f\, dW^2 + \frac{\partial}{\partial t} f \end{aligned}$$

i.e.

$$df(W(t),t) = \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial W^2}\right) dt + \frac{\partial f}{\partial W} dW$$

d) Mean-Value Formula

For non-anticipating functions one has for the Ito stochastic integral

$$\left\langle \int_0^t G(t')dW(t') \right\rangle = 0$$

since

$$\left\langle \sum_{i=1}^N G_{i-1}\left(W_i - W_{i-1}\right) \right\rangle = \sum_{i=1}^N \langle G_{i-1}\rangle \underbrace{\langle W_i - W_{i-1}\rangle}_{=0} = 0$$

Note:

- For the Stratonovich interpretation of the integral this average need not be 0, since G_i can be correlated with $W_i - W_{i-1}$ (G_i is after the kick). Aspects like this make the Stratonovich formulation very difficult/cumbersome to use for proofs.

e) Correlation Formula

One can show (see homework)

$$\left\langle \int_0^t G(t')dW(t') \int_0^t H(t'')dW(t'') \right\rangle = \int_0^t \langle G(t')H(t') \rangle \, dt'$$

5.4 Ito Stochastic Differential Equation[20]

5.4.1 Ito's Formula

Consider the stochastic differential equation

$$dy = f(y)dt + g(y)dW$$

One can transform the variables so that the new variable satisfies a stochastic differential equation with additive rather than multiplicative noise.

While for multiplicative noise there is a significant difference between the differential equation in the Ito or in the Stratonovich sense, there is no difference for additive noise
$\Rightarrow$ the transformation of the variable must be different in the two cases, since in the original version the stochastic differential equation is interpreted differently for Ito and for Stratonovich.

What stochastic differential equation does $v(y,t)$ satisfy if y satisfies the stochastic differential equation above?

$$\begin{aligned} dv &= \frac{\partial v}{\partial y}dy + \frac{1}{2}\frac{\partial^2 v}{\partial y^2}(dy)^2 + \frac{\partial v}{\partial t}dt + h.o.t. \\ &= \frac{\partial v}{\partial y}\{f(y)dt + g(y)dW\} + \frac{1}{2}\frac{\partial^2 v}{\partial y^2}\{f(y)dt + g(y)dW\}^2 + \frac{\partial v}{\partial t}dt + h.o.t. \end{aligned}$$

Thus for Ito stochastic differential equations the change of variables is given by Ito's formula

$$dv = \left(\frac{\partial v}{\partial t} + \frac{\partial v}{\partial y}f + \frac{1}{2}\frac{\partial^2 v}{\partial y^2}g^2 \right) dt + \frac{\partial v}{\partial y} g \, dW.$$

Example: The Kubo oscillator

A noiseless linear oscillator can be described by

$$\frac{dy}{dt} = i\omega y \qquad \Rightarrow \qquad y = y_0 e^{i\omega t}$$

With noise one obtains

$$dy = i\omega y \, dt + i\lambda y dW \tag{48}$$

[20] [15, Ch.4.3][25, Ch. 4.1]

where λdW represents fluctuations in the frequency of the oscillator.

For the first transformation $v(t) = e^{-i\omega t}y(t)$ Ito's formula results in the usual transformation since $d^2v/dy^2 = 0$,

$$dv = \left(-i\omega v + e^{-i\omega t}i\omega y\right)dt + e^{-i\omega t}y\, i\lambda dW$$

$$dv = i\lambda v\, dW$$

For regular functions one would have

$$\frac{dv}{v} = d\left(\ln v\right)$$

To solve this differential equation try a second transformation, $u = \ln v$, which yields

$$du = \left(\frac{1}{2}\left(-\frac{1}{v^2}\right)(i\lambda v)^2\right)dt + \frac{1}{v}i\lambda v dW$$

thus du is not simply given by dv/v, instead

$$du = \frac{1}{2}\lambda^2 dt + i\lambda dW \tag{49}$$

with the solution

$$u(t) - u(0) = \frac{1}{2}\lambda^2 t + i\lambda\left(W(t) - W(0)\right).$$

For y we have now

$$y = e^{i\omega t}e^{\frac{1}{2}\lambda^2 t + i\lambda W(t)} \tag{50}$$

Note:

- With this solution for the stochastic differential equation we have for each realization of the Wiener process ('random walk') a realization of the trajectory of the oscillator.

How about the mean of y?

Clearly

$$\langle u(t) - u(0)\rangle = \frac{1}{2}\lambda^2 t$$

and therefore

$$\langle y(t)\rangle = e^{i\omega t}e^{\frac{1}{2}\lambda^2 t}\left\langle e^{i\lambda(W(t)-W(0))}\right\rangle$$

$W(t)$ is Gaussian with 0 mean.
For Gaussian distributed variables z with vanishing mean one has

$$\langle e^z\rangle = e^{\frac{1}{2}\langle z^2\rangle}$$

One obtains this result by direct evaluation of

$$\langle e^z\rangle = \frac{1}{\sqrt{2\pi\Delta^2}}\int e^z e^{-\frac{1}{2}\frac{z^2}{\Delta^2}}dz$$

Thus

$$
\begin{aligned}
\langle y(t)\rangle &= e^{i\omega t}e^{\frac{1}{2}\lambda^2 t}e^{\frac{1}{2}\left\langle (i\lambda(W(t)-W(0)))^2\right\rangle} \\
&= e^{i\omega t}e^{\frac{1}{2}\lambda^2 t}e^{-\frac{1}{2}\lambda^2 t}
\end{aligned}
$$

$$
\langle y(t)\rangle = e^{i\omega_0 t}
$$

Note:

- Interpreted in the Ito sense the fluctuations do not affect the mean oscillation frequency or the oscillation amplitude of the Kubo oscillator (48).
- Interpreted in the Stratonovich sense (48) would describe fluctuations that lead to a damping of the oscillations. (see homework).
-

Ito's Formula for Multiple Variables

Consider a multi-dimensional Wiener process $\mathbf{W}(t) \equiv (W_1(t), \ldots, W_n(t))$ in which all components are statistically independent of each other and the stochastic differential equation for $\mathbf{x} \equiv (x_1(t), \ldots, x_n(t))$,

$$
d\mathbf{x} = \mathbf{A}(\mathbf{x}, t)dt + \mathbf{B}(\mathbf{x}, t)d\mathbf{W}(t)
$$

What stochastic differential equation does a scalar function of $\mathbf{x}$, $f = f(\mathbf{x})$, satisfy?

$$
\begin{aligned}
df &= \sum_i \frac{\partial f}{\partial x_i} dx_i + \frac{1}{2}\sum_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j} dx_i dx_j \\
&= \sum_i \frac{\partial f}{\partial x_i}\left(A_i dt + \sum_k B_{ik} dW_k\right) + \frac{1}{2}\sum_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j}\left(A_i dt + \sum_k B_{ik} dW_k\right)\left(A_j dt + \sum_l B_{jl} dW_l\right)
\end{aligned}
$$

Use

$$
\begin{aligned}
dW_i dW_j &= \delta_{ij} dt \qquad \text{statistically independent} \\
[dW_i]^{2+n} &= 0 \qquad n > 0 \\
dt^{1+n} &= 0 \qquad n > 0
\end{aligned}
$$

and get

$$
df = \left\{\sum_i \frac{\partial f}{\partial x_i} A_i + \frac{1}{2}\sum_{ijk}\frac{\partial^2 f}{\partial x_i \partial x_j}\underbrace{B_{ik}B_{jk}}_{(\mathbf{B}\mathbf{B}^t)_{ij}}\right\} dt + \sum_{ik}\frac{\partial f}{\partial x_i} B_{ik} dW_k
$$

5.4.2 Solvability by Variable Transform

Under what conditions can the stochastic differential equation

$$dy = f(y,t)dt + g(y,t)dW \tag{51}$$

be solved by such a variable transformation?

Assume an invertible variable transformation

$$v = v(y(t), t)$$

Then we have from Ito's formula

$$dv = \left(\frac{\partial v}{\partial t} + \frac{\partial v}{\partial y} f + \frac{1}{2} \frac{\partial^2 v}{\partial y^2} g^2 \right) dt + \frac{\partial v}{\partial y} g \, dW.$$

This transformation reduces the stochastic differential equation to simple integrations if the coefficients do not depend on v, i.e. if

$$dv = \alpha(t)dt + \beta(t)dW. \tag{52}$$

Thus the conditions are

$$\frac{\partial v}{\partial t} + \frac{\partial v}{\partial y} f + \frac{1}{2} \frac{\partial^2 v}{\partial y^2} g^2 = \alpha(t) \tag{53}$$

$$\frac{\partial v}{\partial y} g = \beta(t) \tag{54}$$

This will be possible only for certain combinations of the functions f and g, i.e. these two equations are not conditions for v but for f and g. We now obtain the corresponding condition.

From (54) we get

$$\frac{\partial v}{\partial y} = \frac{\beta(t)}{g(y(t),t)}. \tag{55}$$

Differentiating (53) with respect to y we get

$$\frac{\partial^2 v}{\partial y \partial t} + \frac{\partial}{\partial y} \left[\frac{\partial v}{\partial y} f + \frac{1}{2} \frac{\partial^2 v}{\partial y^2} g^2 \right] = 0. \tag{56}$$

We want a condition on f and g and not on v. Therefore we want to eliminate v from these conditions and still need expressions for the two terms with the second derivative

$$\begin{aligned} \frac{\partial^2 v}{\partial y \partial t} &= \frac{\partial}{\partial t} \left[\frac{\beta(t)}{g(y(t),t)} \right] \\ &= \frac{\beta'}{g} - \frac{\beta}{g^2} \frac{\partial g}{\partial t} \end{aligned}$$

$$\frac{\partial^2 v}{\partial y^2} = \frac{\partial}{\partial y} \left[\frac{\beta(t)}{g(y(t),t)} \right] = -\frac{\beta}{g^2} \frac{\partial g}{\partial y}.$$

Insert them into (56) to get

$$\frac{\beta'}{g} - \frac{\beta}{g^2}\frac{\partial g}{\partial t} + \frac{\partial}{\partial y}\left[\frac{\beta(t)}{g}f\right] - \frac{1}{2}\frac{\partial}{\partial y}\left[\frac{\beta}{g^2}\frac{\partial g}{\partial y}g^2\right] = 0$$

i.e.

$$\frac{\beta'(t)}{\beta(t)} = \frac{1}{g}\frac{\partial g}{\partial t} - g\frac{\partial}{\partial y}\left[\frac{f}{g}\right] + \frac{1}{2}g\frac{\partial^2 g}{\partial y^2} \tag{57}$$

Since the left-hand side depends only on t the condition on f and g is

$$\frac{\partial}{\partial y}\left\{\frac{1}{g}\frac{\partial g}{\partial t} - g\frac{\partial}{\partial y}\left[\frac{f}{g}\right] + \frac{1}{2}g\frac{\partial^2 g}{\partial y^2}\right\} = 0 \tag{58}$$

Thus

- the condition (58) guarantees that a y-independent β can be determined from (57).
- then $v(y,t)$ can be determined from (55)
- (57) guarantees that the expression (53) defining α is y-independent and therefore such an $\alpha = \alpha(t)$ can indeed be chosen.

Conclusion:

- if the coefficients in the stochastic differential equation (51) satisfies the condition (58) it can be transformed into the linear equation (52).

Example:
For the Kubo oscillator we have

$$f(y,t) = i\omega y \qquad g(y,t) = i\lambda y$$

Condition (58) is satisfied

$$\frac{\partial}{\partial y}\left\{-i\lambda y\frac{\partial}{\partial y}\left[\frac{\omega}{\lambda}\right] + \frac{1}{2}i\lambda y\frac{\partial^2}{\partial y^2}(i\lambda y)\right\} = 0$$

β is determined from

$$\frac{\beta'}{\beta} = 0 \qquad \Rightarrow \qquad \beta = \beta_0$$

For $v(y,t)$ we get the equation

$$\frac{\partial v}{\partial y} = \frac{\beta_0}{i\lambda y} \qquad \Rightarrow \qquad v = \frac{\beta_0}{i\lambda}\ln y + v_0(t)$$

implying

$$y = e^{i\frac{\lambda}{\beta_0}(v - v_0(t))}$$

and

$$\begin{aligned}
\alpha(t) &= v_0'(t) + \frac{\beta_0}{i\lambda}\frac{1}{y}\cdot i\omega y - \frac{1}{2}\frac{\beta_0}{i\lambda}\frac{1}{y^2}\cdot(i\lambda y)^2 \\
&= v_0'(t) + \frac{\beta_0\omega}{\lambda} - \frac{i}{2}\beta_0\lambda
\end{aligned}$$

We still can choose $v_0(t)$ and β_0 to simplify the equations.
For instance

$$\begin{aligned}
\beta_0 &= i\lambda \\
v_0(t) &= -i\omega t
\end{aligned}$$

Then

$$y = e^{v-i\omega t}$$

and

$$dv = \frac{1}{2}\lambda^2 dt + i\lambda dW$$

as before (cf. (49)).

5.4.3 Fokker-Planck Equation from the Ito Stochastic Differential Equation

Consider y satisfying

$$y = f\,dt + g\,dW$$

Use Ito's formula for $dh(y(t))$

$$\langle dh\rangle = \left\langle\left(\frac{dh}{dy}f + \frac{1}{2}\frac{d^2h}{dy^2}g^2\right)dt + \frac{dh}{dy}g\,dW\right\rangle$$

yielding

$$\frac{d}{dt}\langle h\rangle = \frac{\langle dh\rangle}{dt} = \left\langle\frac{dh}{dy}f + \frac{1}{2}\frac{d^2h}{dy^2}g^2\right\rangle$$

Using the probability distribution $P(y,t|y_0,t_0)$ for initial condition $y(t_0)=y_0$ we get

$$\frac{d}{dt}\langle h\rangle = \int h(y)\frac{\partial}{\partial t}P(y,t|y_0,t_0)\,dy = \int\left(\frac{dh}{dy}f + \frac{1}{2}\frac{d^2h}{dy^2}g^2\right)P(y,t|y_0,t_0)\,dy$$

As previously, integrate by parts to get

$$\int h(y)\left\{\frac{\partial}{\partial t}P(y,t|y_0,t_0) + \frac{\partial}{\partial y}\left(fP(y,t|y_0,t_0)\right) - \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(g^2P(y,t|y_0,t_0)\right)\right\}dy = 0$$

Since $h(y)$ is arbitraty we get again the FPE

$$\frac{\partial}{\partial t}P(y,t|y_0,t_0) = -\frac{\partial}{\partial y}\left(fP(y,t|y_0,t_0)\right) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(g^2P(y,t|y_0,t_0)\right) \tag{59}$$

5.5 Stratonovich's Stochastic Differential Equation

Make the connection between the Ito interpretation and the Stratonovich interpretation of the stochastic differential equation explicit.

Consider the Ito stochastic differential equation

$$dy = f(y,t)dt + g(y,t)dW$$

which has the formal integral

$$y = y(0) + \int_0^t f(y(t'),t')dt' + \int_0^t g(y(t'),t')dW(t')$$

Consider now the Stratonovich stochastic differential equation that has the same solution $y(t)$,

$$y = y(0) + \int_0^t \tilde{f}(y(t'),t')dt' + S\int_0^t \tilde{g}(y(t'),t')dW(t'),$$

where $S\int$ denotes the Stratonovich stochastic integral.

What is the connection between f, g and $\tilde{f}$, $\tilde{g}$?

$$\begin{aligned}
S\int_0^t \tilde{g}(y(t'),t')dW(t') &= \sum_i \tilde{g}\left(\frac{1}{2}\left[y_i + y_{i-1}\right], t_i\right)(W(t_i) - W(t_{i-1})) \\
&= \sum_i \tilde{g}\left(y_{i-1} + \frac{1}{2}dy_{i-1}, t_i\right)(W(t_i) - W(t_{i-1})) \\
&= \sum_i \left[\tilde{g}\left(y_{i-1},t_i\right) + \frac{1}{2}\tilde{g}'\left(y_{i-1},t_i\right)\left\{f(y_{i-1},t)\underbrace{dt}_{\to\mathcal{O}(dt\,dW)} + g(y_{i-1},t)dW\right\} + \mathcal{O}(dW^2)\right. \\
&\quad \cdot (W(t_i) - W(t_{i-1})) \\
&= \int \tilde{g}(y(t'),t')dW(t') + \frac{1}{2}\int g(y(t'),t')\frac{\partial\tilde{g}(y(t'),t')}{\partial y}dt'
\end{aligned}$$

Thus we have

$$\begin{aligned}
\int_0^t f(y(t'),t')dt' + \int_0^t g(y(t'),t')dW(t') &= \int_0^t \tilde{f}(y(t'),t')dt' + \int \tilde{g}(y(t'),t')dW(t') + \\
&\quad + \frac{1}{2}\int g(y(t'),t')\frac{\partial\tilde{g}(y(t'),t')}{\partial y}dt'
\end{aligned}$$

which gives the condition for the two equations to have the same solution

$$\begin{aligned}
\tilde{g}(y,t) &= g(y,t) \\
\tilde{f}(y,t) &= f(y,t) - \frac{1}{2}g(y,t)\frac{\partial g(y,t)}{\partial y}
\end{aligned}$$

i.e.

$$\begin{array}{llcll}
dy = f\,dt + g\,dW & [Ito] & \Leftrightarrow & dy = \left(f - \frac{1}{2}g\frac{\partial g}{\partial y}\right)dt + g\,dW & [Stratonovich] \\
dy = f\,dt + g\,dW & [Stratonovich] & \Leftrightarrow & dy = \left(f + \frac{1}{2}g\frac{\partial g}{\partial y}\right)dt + g\,dW & [Ito]
\end{array}$$

Variable Transformation:

Consider

$$dy = f\,dt + g\,dW \qquad [Stratonovich]$$

and

$$v = v(y) \qquad \Leftrightarrow \qquad y = u(v)$$

To use Ito's formula for the variable transformation we need to rewrite the differential equation first in the Ito sense

$$dy = \left(f + \frac{1}{2} g \frac{\partial g}{\partial y}\right) dt + g\,dW \quad [Ito]$$

then use Ito's formula

$$dv = \left[\frac{dv}{dy}\left(f + \frac{1}{2} g \frac{\partial g}{\partial y}\right) + \frac{1}{2}\frac{d^2 v}{dy^2} g^2\right] dt + \frac{dv}{dy} g\,dW$$

We need to write everything in terms of v

$$\frac{dv}{dy} = \frac{1}{\frac{du}{dv}}$$

and

$$\begin{aligned}
\frac{d^2 v}{dy^2} &= \frac{d}{dy}\left(\frac{1}{\frac{du}{dv}}\right) = \frac{d}{dv}\left(\frac{1}{\frac{du}{dv}}\right)\frac{dv}{dy} \\
&= -\frac{1}{u'^2}\frac{d^2 u}{dv^2}\frac{1}{u'} = -\frac{1}{u'^3}\frac{d^2 u}{dv^2}
\end{aligned}$$

Thus

$$dv = \left[\frac{1}{u'} f + \frac{1}{u'}\frac{1}{2} g \frac{\partial g}{\partial v}\frac{1}{u'} - \frac{1}{2}\frac{1}{u'^3}\frac{d^2 u}{dv^2} g^2\right] dt + \frac{1}{u'} g\,dW \qquad [Ito]$$

Now convert the Ito SDE back to a Stratonovich SDE

$$\begin{aligned}
dv &= \left[\frac{1}{u'} f + \frac{1}{u'}\frac{1}{2} g \frac{\partial g}{\partial v}\frac{1}{u'} - \frac{1}{2}\frac{1}{u'^3}\frac{d^2 u}{dv^2} g^2\right] dt + \frac{1}{u'} g\,dW \\
&\quad -\frac{1}{2}\frac{1}{u'} g \frac{d}{dv}\left(\frac{1}{u'} g\right) dt \\
&= \left[\frac{1}{u'} f + \frac{1}{u'}\frac{1}{2} g \frac{\partial g}{\partial v}\frac{1}{u'} - \frac{1}{2}\frac{1}{u'^3}\frac{d^2 u}{dv^2} g^2\right] dt + \frac{1}{u'} g\,dW \\
&\quad -\frac{1}{2}\frac{1}{u'} g\left(-\frac{1}{u'^2} u'' g + \frac{1}{u'}\frac{\partial g}{\partial v}\right) dt \\
&= [f\,dt + g\,dW]\frac{1}{u'} \\
dv &= \frac{dv}{dy}\,[f\,dt + g\,dW]
\end{aligned}$$

Thus:

- for the Stratonovich stochastic differential equation the usual variable transformations hold

5.6 Colored Noise: Ornstein-Uhlenbeck and its White-Noise Limit

So far we always considered only Gaussian white noise, i.e. the Wiener process, in the stochastic differential equation. How can one treat different types of noise, 'colored noise'?

Could develop the stochastic calculus for other types of noise. If the colored noise can be written as driven by white noise it is easier to introduce an additional equation.

We will find: The Ito SDE can in general not be considered as the white-noise limit of noise with finite correlation time.

Recall the Fokker-Planck equation for the Ornstein-Uhlenbeck process (22)

$$\frac{\partial}{\partial t}P(v,t) = \gamma\frac{\partial}{\partial v}\left(vP(v,t)\right) + \gamma\frac{kT}{M}\frac{\partial^2}{\partial v^2}P(v,t)$$

Comparing with (59) we get

$$f = -\gamma v \qquad g = \sqrt{\frac{2kT}{M}}\sqrt{\gamma} \equiv \kappa\sqrt{\gamma}$$

Consider a system driven by the Ornstein-Uhlenbeck process

$$\frac{dy}{dt} = ay + byv \tag{60}$$
$$dv = -\gamma v\,dt + \kappa\sqrt{\gamma}\,dW \tag{61}$$

Since v is a continuous function (rather than discontinous random kicks $L(t)$) the equation (60) for y is a conventional differential equation.

(61) has additive noise, therefore the additional term of Ito's formula does not arise and we get a standard variable transformation. Can therefore do usual integrating factor

$$v = e^{-\gamma t}u$$

$$du = \kappa\sqrt{\gamma}e^{\gamma t}dW$$

$$u = v_0 + \kappa\sqrt{\gamma}\int_0^t e^{\gamma t'}dW(t')$$

and

$$v(t) = v_0e^{-\gamma t} + \kappa\sqrt{\gamma}\int_0^t e^{-\gamma(t-t')}dW(t')$$

For simplicity set $v_0 = 0$.

For given $v(t)$ we can solve for y again by integrating factor

$$y(t) = y_0 e^{\int_0^t a+bv(t')dt'}$$

The Ornstein-Uhlenbeck process is a Gaussian process with finite correlation time $\tau = \gamma^{-1}$. For $\gamma \to \infty$ it turns into noise with vanishing correlation time. Determine y in that limit.

Consider first the autocorrelation function, assuming $t_2 \geq t_1$

$$
\begin{aligned}
C(t_1,t_2) = \left\langle \left(\kappa\sqrt{\gamma}\int_0^{t_1} e^{-\gamma(t_1-t')}dW(t')\right)\left(\kappa\sqrt{\gamma}\int_0^{t_2} e^{-\gamma(t_2-t'')}dW(t'')\right)\right\rangle &= \kappa^2\gamma\int_0^{t_1} e^{-\gamma(t_1-t')-\gamma(t_2-t')}dt' \\
&= \kappa^2\frac{1}{2}e^{-\gamma(t_2+t_1)}\left(e^{2\gamma t_1}-1\right) \\
&= \kappa^2\frac{1}{2}e^{-\gamma(t_2-t_1)} - \kappa^2\frac{1}{2}e^{-\gamma(t_2+t_1)}
\end{aligned}
$$

where we used the correlation formula and $\langle dW(t')dW(t'')\rangle = 0$ for $t' \neq t''$.

For large times we get

$$C(t_1,t_2) = \kappa^2\frac{1}{2}e^{-\gamma|t_2-t_1|}$$

We want that for $\gamma \to \infty$

$$C(t_1,t_2) \to \delta(t_2-t_1)$$

i.e.

$$1 = \int_{-\infty}^{+\infty} C(t,t')dt' = \kappa^2\frac{1}{2}\frac{2}{\gamma}$$

We need to choose $\kappa^2 = \gamma$.

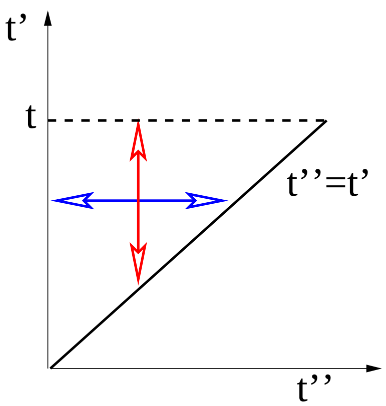
Note:

- increasing only the dissipation shortens the correlation time but decreases the overall amount of noise $\Rightarrow$ we need to increase the temperature $T = M\kappa^2/2k$ at the same to keep the noise level constant

Need to evaluate the integral for $v(t)$ for large γ.

$$\int_0^t v(t')dt' = \gamma\int_0^t dt'\int_0^{t'} e^{-\gamma(t'-t'')}dW(t'')$$

To exploit the limit $\gamma \to \infty$ we would like to do the t'-integral first



$$
\begin{aligned}
\lim_{\gamma\to\infty}\int_0^t v(t')dt' &= \lim_{\gamma\to\infty}\gamma\int_0^t dW(t'')\int_{t''}^t e^{-\gamma(t'-t'')}dt' = \\
&= \lim_{\gamma\to\infty}\gamma\int_0^t dW(t'')\frac{1}{\gamma}\left(1-e^{-\gamma(t-t'')}\right) = \\
&= \int_0^t dW(t'') = W(t)
\end{aligned}
$$

21

Thus we have

$$
y(t) = y_0 e^{at+bW(t)}
$$

Since for $\gamma \to \infty$ the Ornstein-Uhlenbeck process becomes δ-correlated, one might expect that one could replace the forcing in v directly by the Wiener process

$$
dy = ay\,dt + by\,dW \qquad [Ito]
$$

Comparing with (50) we get then

$$
y = y_0 e^{at-\frac{1}{2}b^2t+bW(t)}
$$

which does not agree with the limit $\gamma \to \infty$ of the Ornstein-Uhlenbeck process.

Notes:

- as was the case for the Fokker-Planck description before, the Ito SDE does not describe the limit of vanishing correlation time of smooth noise
- colored noise can be treated by using an additional equation.

[21]Aside: consider for an arbitrary test function $h(t)$ and large γ

$$
\begin{aligned}
\int_0^\infty h(t')\kappa\sqrt{\gamma}e^{-\gamma(t-t')}dt' &= \kappa\sqrt{\gamma}\int_0^\infty\left(h(0)+h'(0)t'+\frac{1}{2}h''(0)t'^2+\ldots\right)e^{-\gamma(t-t')}dt' \\
&= \kappa\sqrt{\gamma}\left\{\frac{1}{\gamma}h(0)+h'(0)\frac{d}{d\gamma}\left(\frac{1}{\gamma}\right)+\frac{1}{2}h''(0)\frac{d^2}{d\gamma^2}\left(\frac{1}{\gamma}\right)+\ldots\right\} \\
&= \frac{\kappa}{\sqrt{\gamma}}h(0)+\mathcal{O}\left(\frac{\kappa}{\gamma^{3/2}}\right)
\end{aligned}
$$

thus setting $\kappa = \sqrt{\gamma}$

$$
\lim_{\gamma\to\infty}\gamma e^{-\gamma(t-t')} = \delta(t-t')
$$

Then for $\gamma \to \infty$

$$
v(t) = \int_0^t \delta(t-t')dW(t')
$$

For y we need

$$
\int_0^t v(t')dt' = \int_0^t\left\{\int_0^{t'}\delta(t'-t'')dW(t'')\right\}dt' \quad \underbrace{=}_{\text{can be shown}} \quad W(t)
$$

6 Stochastic Resonance

There are climate variations on all kinds of time scales, from few years to 100,000 years.

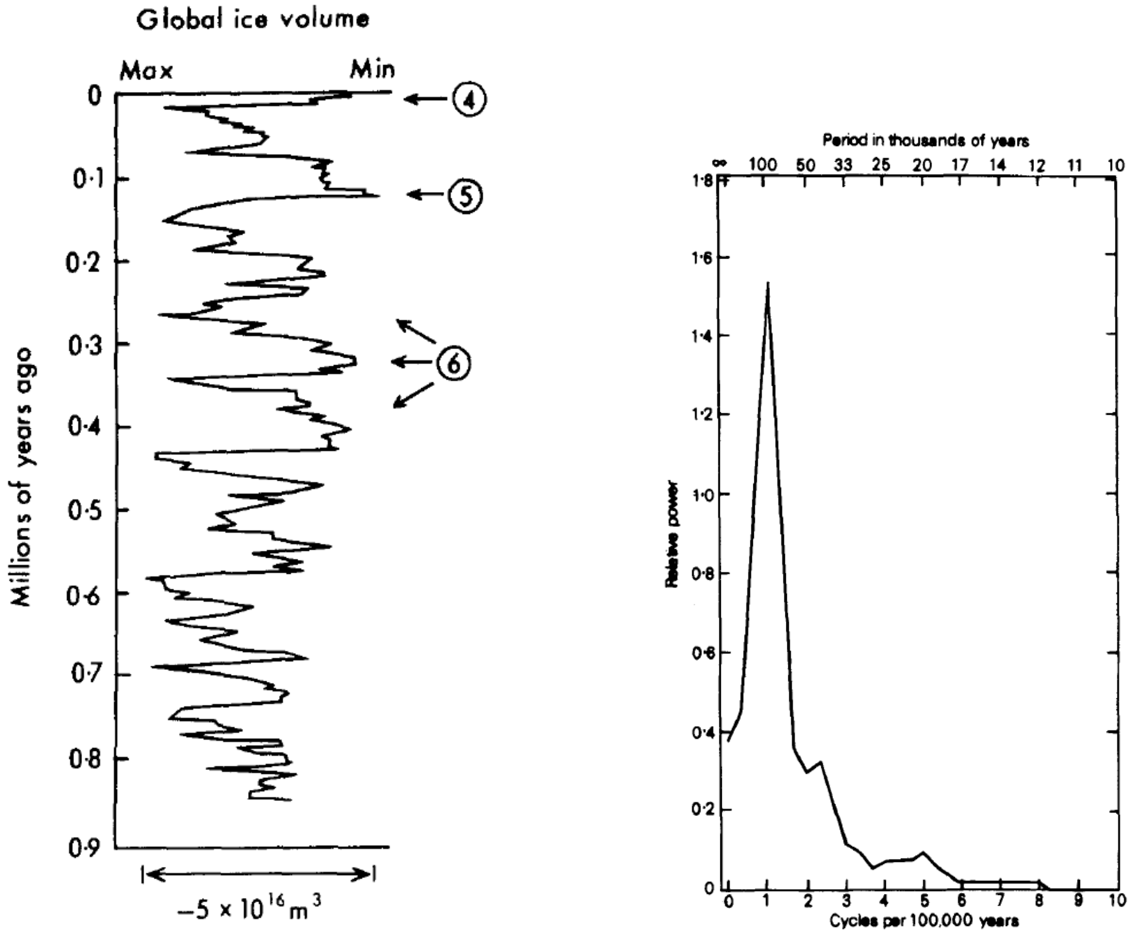


Figure 1: a) Fluctuation in global ice volume (or sea level) from observations of the oxygen-isotope content of fossil plankton in deep-sea core, which indicates fluctuations in gobal ice volume. b) Power spectrum of that time series. [20]

The earth 's orbit has secular variations (Milankovich cycles):

- the eccentricity of the orbit changes by 0.1% with a period of 96,000 years
- obliquity of axis of orbit varies with a period of 40,000 years
- precession of the longitude of the perihelion varies with a 21,000 year period.

However, the magnitude of the changes of the 96,000 year period is much too small:

- the temperature change induced directly by the variable heat flux is $\mathcal{O}(0.3K)$
- the observed changes in temperature are $\mathcal{O}(10K)$

It was suggested by Benzi et al. [4] that fluctuations in the climate ('noise') may trigger transitions between two states, are get resonantly enhanced by the small modulation: stochastic resonance. They considered a bistable system in which the barrier between the states was modulated slowly in time.

Transition rate is given by Kramers formula (27)

$$W = \frac{1}{2\pi\gamma}\omega_1\omega_2 e^{-\frac{2}{b\gamma}\Delta U}$$

Notes:

- Kramers formula valid for $\Delta U \gg b\gamma$
- if the potential is time-dependent the temporal variation has to be slow compared to the evolution inside the potential

$$\frac{1}{U}\frac{dU}{dt} \ll \omega_1$$

  so that the probability distribution in the well is (almost) equilibrated at all times.

Consider a double-well system as a *two-state* system[21]

$$n_+(t) = \int_{x_m}^{\infty} P(x',t)dx' \qquad n_-(t) = \int_{-\infty}^{x_m} P(x',t)dx'$$

Transition rates between the two states

$$W_\pm(t): \qquad n_\pm \rightarrow n_\mp$$

Evolution equation for the probability for the system to be in state $+$

$$\frac{dn_+}{dt} = -\frac{dn_-}{dt} = W_-(t)n_- - W_+(t)n_+$$

$$\frac{dn_+}{dt} = W_-(t) - \left(W_+(t) + W_-(t)\right)n_+$$

The transition rates are time-dependent because the barrier height depends on time.

$$\frac{d}{dt}\left(e^{\int^t W_+ + W_- dt'} n_+\right) = W_- e^{\int^t W_+ + W_- dt'}$$

$$n_+(t) = e^{-\int_{t_0}^t W_+ + W_- dt'} n_+(t_0) + e^{-\int_{t_0}^t W_+ + W_- dt'} \int_{t_0}^t W_-(t')\, e^{\int_{t_0}^{t'} W_+ + W_- dt''} dt'$$

Consider modulations of the barrier

$$\Delta U_\pm = \Delta U_0 \pm \Delta U_1 \cos\omega_s t$$

where ΔU_+ is the barrier height from the right potential well and ΔU_- from the left.

If the modulation amplitude is small, $\Delta U_1 \ll b\gamma$ one gets

$$\begin{aligned} W_\pm(t) &= \frac{1}{2\pi\gamma}\omega_1\omega_2 e^{-\frac{2}{b\gamma}(\Delta U_0 \pm \Delta U_1 \cos\omega_s t)} = \\ &= \frac{1}{2\pi\gamma}\omega_1\omega_2 e^{-\frac{2}{b\gamma}\Delta U_0}\left\{1 \mp \frac{2}{b\gamma}\Delta U_1 \cos\omega_s t\right\} \end{aligned}$$

which can be written as

$$W_\pm(t) = \frac{1}{2}\alpha_0 \mp \frac{1}{2}\epsilon\alpha_1 \cos\omega_s t + \mathcal{O}(\epsilon^2)$$

with

$$\alpha_0 = \frac{1}{\pi\gamma}\omega_1\omega_2 e^{-\frac{2}{b\gamma}\Delta U_0} \qquad \epsilon\alpha_1 = \frac{2\Delta U_1}{b\gamma}\alpha_0$$

Then

$$W_+ + W_- = \alpha_- + \mathcal{O}(\epsilon^2)$$

$$\begin{aligned} n_+ &= e^{-\alpha_0(t-t_0)}\left\{n_+(t_0) + \int_{t_0}^{t}\left(\frac{1}{2}\alpha_0 + \frac{1}{2}\epsilon\alpha_1\cos\omega_s t'\right)e^{\alpha_0(t'-t_0)}dt'\right\} = \\ &= e^{-\alpha_0(t-t_0)}\left\{n_+(t_0) + \frac{1}{2}\left(e^{\alpha_0(t-t_0)} - 1\right) + \frac{1}{2}\frac{\epsilon\alpha_1}{\sqrt{\omega_s^2+\alpha_0^2}}\left(\cos\left(\omega_s t - \phi\right)e^{\alpha_0(t-t_0)} - \cos\left(\omega_s t_0 - \phi\right)\right)\right\} \end{aligned}$$

with

$$\phi = \arctan\frac{\omega_s}{\alpha_0}$$

Assuming the system is initially either in the right or the left well one gets for the transition probability

$$\begin{aligned} n_+(t|x_0,t_0) &= \frac{1}{2}\left\{e^{-\alpha_0(t-t_0)}\left[2\underbrace{\delta_{x_0,+}}_{=1 \text{ for } x_0>0} - 1 - \frac{\epsilon\alpha_1}{\sqrt{\omega_s^2+\alpha_0^2}}\cos\left(\omega_s t_0 - \phi\right)\right] + 1 + \frac{\epsilon\alpha_1}{\sqrt{\omega_s^2+\alpha_0^2}}\cos(\omega_s t - \phi \right. \\ &= \frac{1}{2}\left\{e^{-\alpha_0(t-t_0)}\left[2\underbrace{\delta_{x_0,+}}_{=1 \text{ for } x_0>0} - 1 - \epsilon\tilde{\alpha}_1\cos\left(\omega_s t_0 - \phi\right)\right] + 1 + \epsilon\tilde{\alpha}_1\cos\left(\omega_s t - \phi\right)\right\} \end{aligned}$$

with

$$\tilde{\alpha}_1 = \frac{\alpha_1}{\sqrt{\omega_s^2+\alpha_0^2}}$$

To get the power spectrum determine the correlation function

Approximate the position

$$\langle x(t)\rangle = x_+ n_+(t) + x_- n_-(t) \equiv c\left(n_+ - n_-\right)$$

Then, using $\langle x(t)\rangle = 0$

$$\begin{aligned} \langle x(t)x(t+\tau)|x_0,t_0\rangle &= \sum_{s,s'=\pm1} sc\, s'c P(s,t+\tau;s',t|x_0,t_0) = \\ &= c^2\sum_{s,s'=\pm1} ss' P(s,t+\tau|s',t)P(s',t|x_0,t_0) \end{aligned}$$

$$
\begin{aligned}
c^{-2}\left\langle x(t)x(t+\tau)|x_0,t_0\right\rangle &= n_+(t+\tau|+c,t)n_+(t|x_0,t_0) - n_+(t+\tau|-c,t)\underbrace{n_-(t|x_0,t_0)}_{1-n_+(t|x_0,t_0)} \\
&\quad - \underbrace{n_-(t+\tau|+c,t)}_{1-n_+(t+\tau|+c,t)} n_+(t|x_0,t_0) + \underbrace{n_-(t+\tau|-c,t)}_{1-n_+(t+\tau|-c,t)}\underbrace{n_-(t|x_0,t_0)}_{1-n_+(t|x_0,t_0)}
\end{aligned}
$$

The contributions to this correlation can be read as different sequences of events like

$$
n_-(t+\tau|+,t)n_+(t|x_0,t_0) \qquad \Leftrightarrow \qquad \begin{pmatrix} x_0 \\ t_0 \end{pmatrix} \rightarrow \begin{pmatrix} x_+ \\ t \end{pmatrix} \rightarrow \begin{pmatrix} x_- \\ t+\tau \end{pmatrix}
$$

Using $n_- = 1 - n_+$ we get

$$
\begin{aligned}
c^{-2}\left\langle x(t)x(t+\tau)|x_0,t_0\right\rangle &= \left\{2n_+(t+\tau|+c,t) - 1 + 2n_+(t+\tau|-c,t) - 1\right\} n_+(t|x_0,t_0) \\
&\quad - \left\{2n_+(t+\tau|-c,t) - 1\right\}
\end{aligned}
$$

For large times ($t_0 \rightarrow -\infty$)

$$
\begin{aligned}
\frac{\left\langle x(t)x(t+\tau)|x_0,t_0\right\rangle}{c^2} &\rightarrow \Big\{e^{-\alpha_0\tau}\left[1 - \epsilon\tilde{\alpha}_1\cos\left(\omega_s t - \phi\right)\right] + 1 + \epsilon\tilde{\alpha}_1\cos\left(\omega_s\left(t+\tau\right) - \phi\right) - 1 \\
&\quad + e^{-\alpha_0\tau}\left[-1 - \epsilon\tilde{\alpha}_1\cos\left(\omega_s t - \phi\right)\right] + 1 + \epsilon\tilde{\alpha}_1\cos\left(\omega_s\left(t+\tau\right) - \phi\right) - 1\Big\} \cdot \\
&\quad \cdot \frac{1}{2}\left[1 + \epsilon\tilde{\alpha}_1\cos\left(\omega_s t - \phi\right)\right] \\
&\quad - \Big\{e^{-\alpha_0\tau}\left[-1 - \epsilon\tilde{\alpha}_1\cos\left(\omega_s t - \phi\right)\right] + 1 + \epsilon\tilde{\alpha}_1\cos\left(\omega_s\left(t+\tau\right) - \phi\right) - 1\Big\} \\
&= e^{-\alpha_0\tau} + \epsilon\tilde{\alpha}_1\left(0\right) + \\
&\quad + \epsilon^2\tilde{\alpha}_1^2\left[-e^{-\alpha_0\tau}\cos^2\left(\omega_s t - \phi\right) + \cos\left(\omega_s\left(t+\tau\right) - \phi\right)\cdot\cos\left(\omega_s t - \phi\right)\right]
\end{aligned}
$$

Notes:

- the correlation function depends on τ and t even for $t_0 \rightarrow -\infty$ due to the periodic modulation, which defines a phase
- the dependence on t is only be apparent if the averaging is done at fixed values of the phase relative to the modulation.

Consider averaging over multiple experiments, each starting at different times in the cycle $\Rightarrow$ average over t.

Typically, even for single run the correlation function is determined as an average over time

$$
\left\langle x(t)x(t+\tau)|x_0,t_0\right\rangle \rightarrow \int_0^\infty x(t)x(t+\tau)dt
$$

Therefore look at average over one period

$$
\frac{\omega_s}{2\pi}\int_0^{\frac{2\pi}{\omega_s}}\cos\left(\omega_s\left(t+\tau\right) - \phi\right)\cdot\cos\left(\omega_s t - \phi\right)dt = \cos\omega_s\tau\cdot\frac{1}{2} - \sin\omega_s\tau\cdot 0 = \frac{1}{2}\cos\omega_s\tau
$$

$$
\frac{\omega_s}{2\pi}\int_0^{\frac{2\pi}{\omega_s}}\frac{\left\langle x(t)x(t+\tau)|x_0,t_0\right\rangle}{c^2}dt = \underbrace{e^{-\alpha_0|\tau|}\left\{1 - \frac{1}{2}\epsilon^2\tilde{\alpha}_1^2\right\}}_{\text{decaying component}} + \underbrace{\frac{1}{2}\epsilon^2\tilde{\alpha}_1^2\cos\omega_s\tau}_{\text{periodic component}}
$$

Note:

- for general τ the exponential has $|\tau|$. In our derivation we had assumed $\tau > 0$

Averaged power spectrum

$$\frac{1}{c^2}\left\langle S(\Omega)\right\rangle_t = \left\{1 - \frac{1}{2}\epsilon^2\tilde{\alpha}_1^2\right\} \underbrace{\frac{2\alpha_0}{\alpha_0^2+\Omega^2}}_{\text{Lorentzian shape}} + \frac{1}{2}\pi\epsilon^2\tilde{\alpha}_1^2\left\{\delta(\Omega-\omega_s)+\delta(\Omega+\omega_s)\right\}$$

We are interested in the dependence of the signal-to-noise ratio at $\Omega = \omega_s$, i.e. the ratio of the δ-function to the smooth spectrum $\Omega = \omega_s$

$$\begin{aligned} R &= \frac{\pi\epsilon^2\tilde{\alpha}_1^2}{\left\{1-\frac{1}{2}\epsilon^2\tilde{\alpha}_1^2\right\}\frac{2\alpha_0}{\alpha_0^2+\omega_s^2}} = \frac{\pi\epsilon^2\tilde{\alpha}_1^2\left(\alpha_0^2+\omega_s^2\right)}{2\alpha_0} + \mathcal{O}(\epsilon^4) = \\ &= \frac{\pi\epsilon^2\alpha_1^2}{2\alpha_0} = \frac{\pi 4\left(\Delta U_1\right)^2}{2b^2\gamma^2}\frac{1}{\pi\gamma}\omega_1\omega_2 e^{-\frac{2}{b\gamma}\Delta U_0} \end{aligned}$$

Thus

$$R = k_0\left(\Delta U_1\right)^2\frac{e^{-\frac{2}{b\gamma}\Delta U_0}}{b^2}$$

Notes:

- for small noise b the signal-to-noise ratio increase with increasing noise level: the number of jumps between the two wells increases and because the escape time depends so sensitively on the barrier height these jumps occur predominantly during a relatively well-defined period when the relevant barrier is minimal
- for large noise levels the jumping can also occur at any other times: signal becomes noisy again
- essential is the bistability of the system: the system has to be nonlinear

6.1 Examples

Stochastic resonance has been found and investigated in a wide range of physical and biological systems (for a review see e.g. [14])

- climate models
- various set-ups of lasers
- SQUIDs (superconducting quantum interference device): superconducting loop interrupted by a Josephson junction
- chemical reactions
- neuronal systems
- psycho-physical experiments

Here just a couple of examples.

6.1.1 Ring Laser [22]

In the ring laser the beam can travel in one of two opposite directions ⇒ bistability

By an acousto-optic coupler one of the two directions can be preferred: the preferred direction depends on the frequency of the modulation ⇒ modulating the frequency periodically a periodic switching of directions can be induced

Combine a periodic and noisy modulation of the ultrasound frequency to get stochstic resonance.

This paper triggered a huge interest in stochastic resonance.

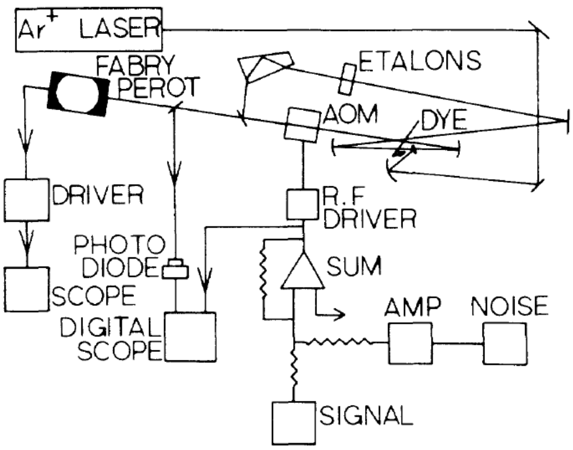


Figure 2: Experimental set-up of ring laser **[22]**

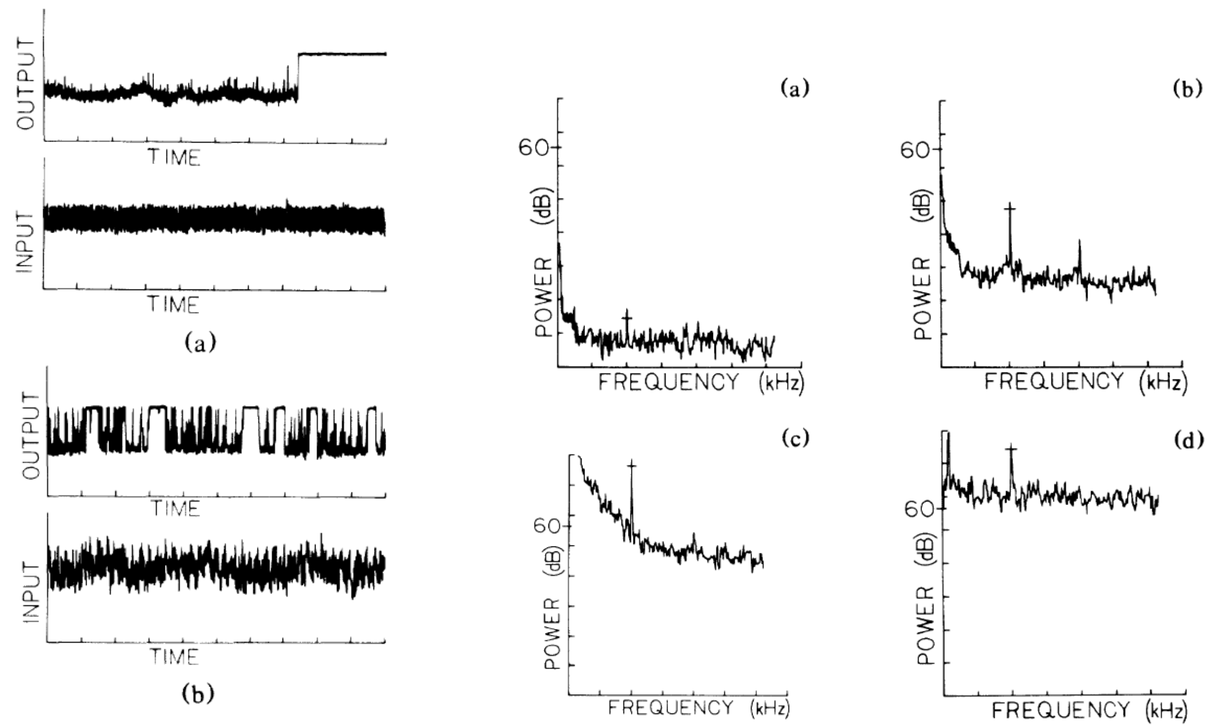


Figure 3: Left: Input and output of laser for two noise strengths. Right: Power spectrum for increasing noise strength [22].

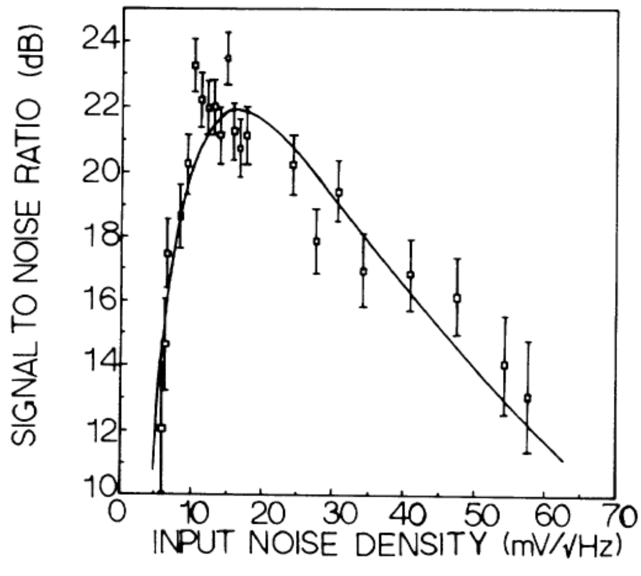


Figure 4: Dependence of the signal-to-noise ratio on the noise strength **[22].**

6.1.2 Mechanoreceptors in Crayfish[12]

Stimulate mechanical mechanoreceptors on the tailfan of the crayfish ⇒ sensory neuron generates spikes triggered by the stimulation

Stimulus: weak periodic signal + noise of variable strength

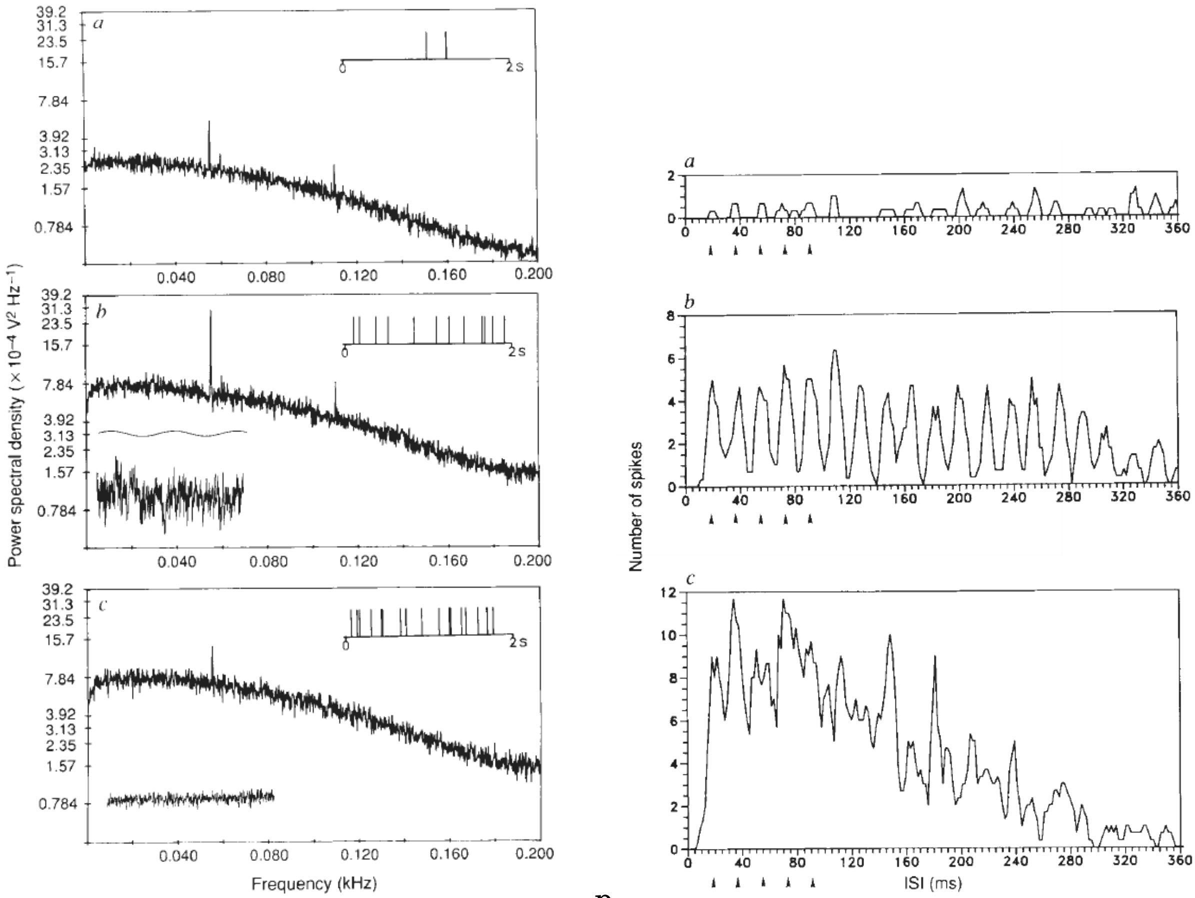


Figure 5: a) Power spectrum from spiking activity of crayfish mechanoreceptor for 3 different noise levels with fixed weak periodic signal. b) Interspike interval histograms for different noise levels [12].

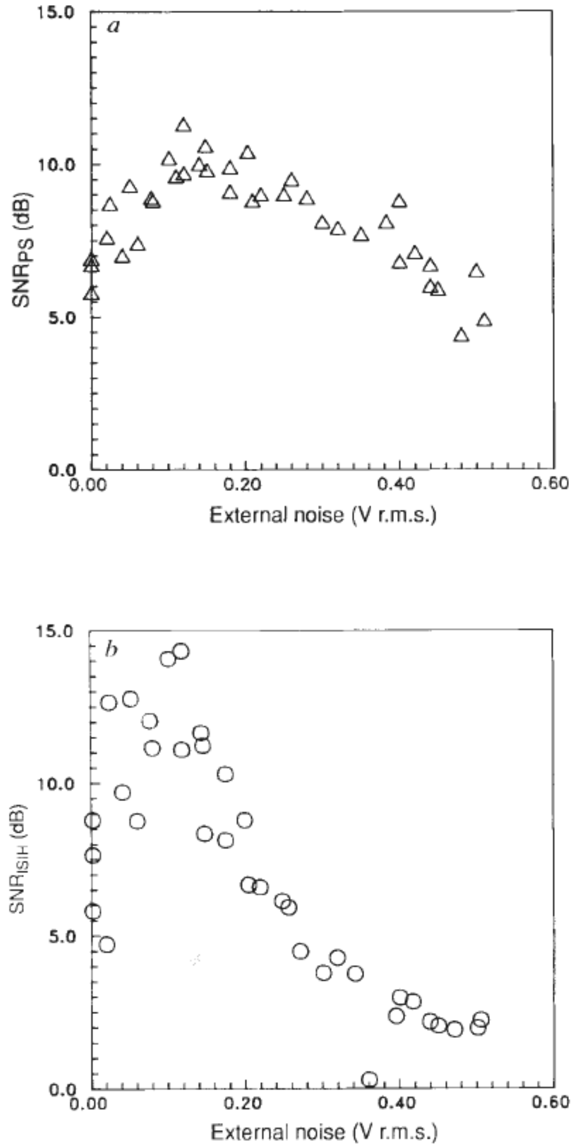


Figure 6: Dependence of the signal-to-noise ratio as determined from the power spectrum (top) and the interspike interval distribution (bottom) on the strength of the external noise [12].

6.1.3 Tactile Sensation[9]

Human psychophysics experiment: apply small indentations to the tip of a subject's middle digit

There is a minimal indentation that is needed for the person to perceive it.

To be discriminated:

- subthreshold stimulus plus noise
- no stimulus plus noise

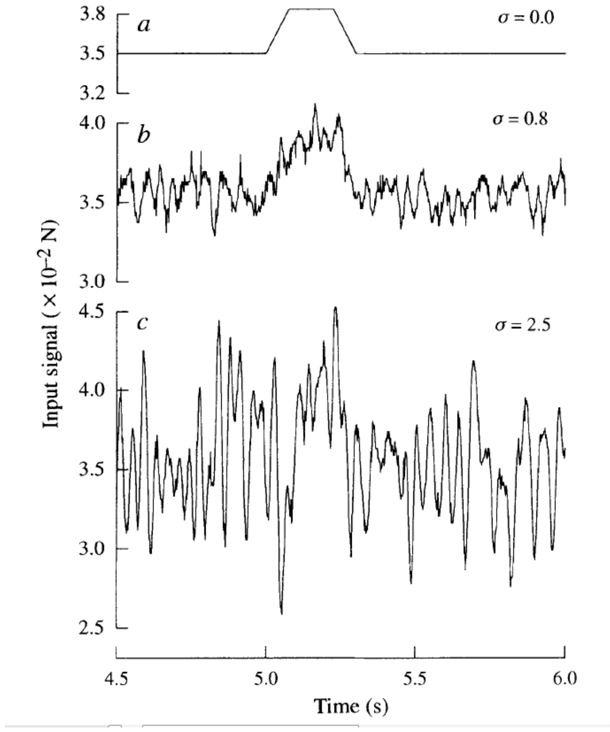


Figure 7: Indentation stimulus given to the finger tip. Three different noise strengths [9].

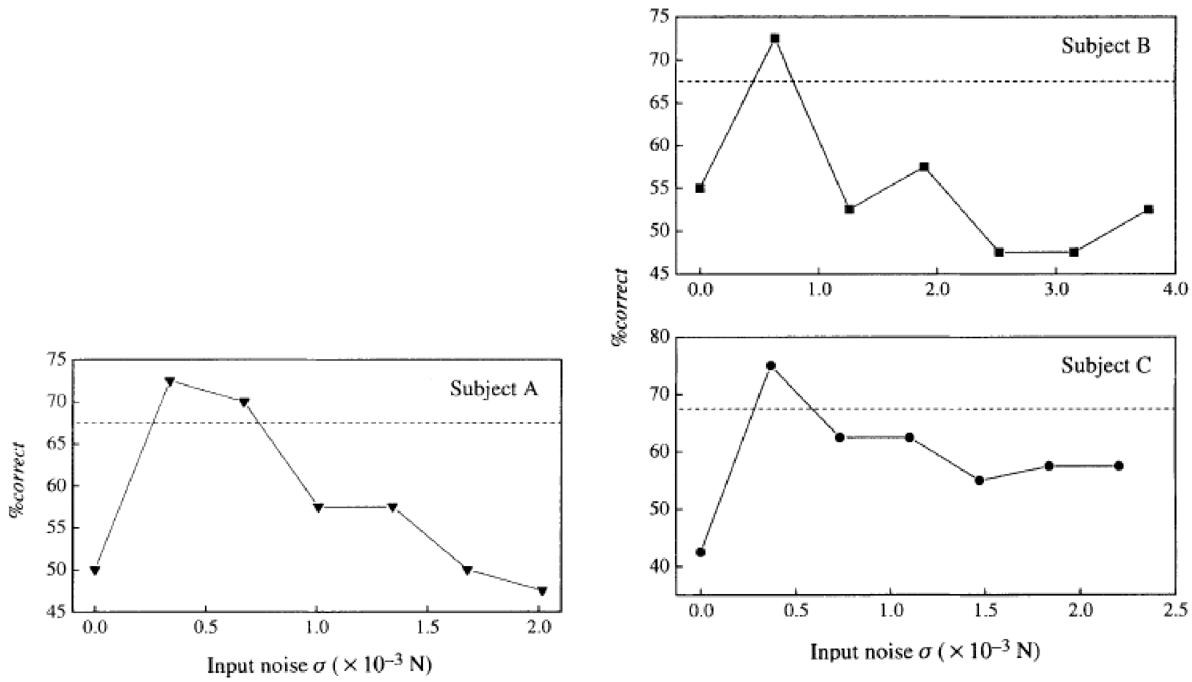


Figure 8: Percent of correct discrimination of the stimulus for three different subjects as a function of noise strength [9].

For another interesting psychophysics experiment see [26].

7 Sketch of Numerical Methods for Stochastic Differential Equations

For numerical solutions of stochastic differential equations two goals are possible

1. **Strong approximation**: pathwise approximation
   for any given realization $W(t)$ of the noise the numerical solution approximates the exact solution $\tilde{y}$
   $$E_s(t) = \langle |y(t) - \tilde{y}(t)| \rangle = \frac{1}{N} \sum_{k=1}^{N} |y_k(t) - \tilde{y}_k(t)|.$$
   Here $y_k(t)$ is the numerical result one obtains for the k^{th}−realization of the noise. To get a good estimate of the error the number N of realizations has to be sufficiently large. Add up absolute value of error to avoid error cancellation.

2. **Weak approximation**: approximation of expectation values
   for any $f(y)$ in a class of *test functions* the mean value obtained with the numerical solution approximates the mean value of $f(y)$
   $$E_m(t) =< f(\tilde{y}(t)) > - < f(y(t)) >$$
   Typically one would require convergence of the

   (a) mean: $f(y) = y$

   (b) variance: $f(y) = y^2$

Notes:

- to obtain a strong approximation the numerical realizations $W(t)$ of the noise have to approximate the exact realizations
- for a weak approximation the numerical noise can be quite different than the exact noise as long as it yields a $y(t)$ for which sufficiently many expectation values agree, e.g. mean, variance, higher moments $\langle (y(t))^m \rangle$.

7.1 Strong Approximation

7.1.1 Euler-Maruyama Scheme

Consider
$$dy = f(y,t)dt + g(y,t)dW \tag{62}$$
Discretize time and integrate over a short time interval Δt
$$\int_t^{t+\Delta t} dy = \int_t^{t+\Delta t} f(y,t')dt' + \int_t^{t+\Delta t} g(y,t')dW(t')$$

Using a left-endpoint rule with only a single interval one obtains the Euler-Maruyama scheme

$$y_{n+1} = y_n + f(y_n, t_n)\Delta t + g(y_n, t_n)\Delta W_n \tag{63}$$

where

$$\Delta W_n = W(t_n + \Delta t) - W(t_n).$$

The ΔW_n are Gaussian distributed with variance Δt. They are $\delta-$correlated

$$\langle \Delta W_n \Delta W_{n'} \rangle = \delta_{n,n'}.$$

Using a normally distributed variable $\Delta\tilde{W}$ that has variance 1 we get ΔW_n by setting

$$\Delta W_n = \sqrt{\Delta t}\, \Delta\tilde{W} \qquad \text{with} \qquad P(\Delta\tilde{W}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\Delta\tilde{W}^2}{2}} \tag{64}$$

The noise strength is characterized by $g(y_n, t_n)$.

Notes:

- For each time step generate a new random number ΔW_n that obeys Gaussian statistics (normal distribution); in matlab this is done with `randn`.
- If y is a vector, usually the random processes for different components of y are independent: for each component of y one has to generate a different independent random number
- To check convergence of the strong approximation:
  need to compare solutions with different time steps Δt for the **same realization** $W(t)$
  1. Starting with smallest Δt, generate increments $\Delta W_n^{(0)}$ for all time steps t_n using a random number generator for a normal (Gaussian) distribution with variance Δt according to (64).
  2. Increase the time step to $\Delta t^{(2)} \equiv 2\Delta t$, generate the Wiener process with increments $\Delta W_n^{(1)}$with larger time step by adding pairs of successive increments,
  $$\Delta W_n^{(1)} = \Delta W_{2n}^{(0)} + \Delta W_{2n+1}^{(0)} \qquad n = 0, 1, ... \tag{65}$$
  3. Continue to add up the appropriate increments of the Wiener process with increments $\Delta W_n^{(l)}$ to generate Wiener processes with increments $\Delta W_n^{(l+1)}$, $l = 1, 2, \ldots$ corresponding to time steps $\Delta t^{(l+1)} = 2^{l+1}\Delta t$. These compound increments have the variance
  $$\langle \left(\Delta W_n^{(l)}\right)^2 \rangle = 2^l$$
  since the variances are additive
  $$\left\langle (x_1 + x_2)^2 \right\rangle = \int (x_1 + x_2)^2 \, e^{-\frac{x_1^2}{2\sigma^2}} e^{-\frac{x_2^2}{2\sigma^2}} dx_1 dx_2 = \left\langle x_1^2 \right\rangle + \underbrace{\langle 2x_1 x_2 \rangle}_{=0} + \left\langle x_2^2 \right\rangle$$

4. Use $2^l\Delta t$ and $\Delta W_n^{(l)}$, $l = 1, 2, \ldots$, for successively less accurate approximations in (63).

Note:

- For $f(y_n, t_n) = 0$ and $g(y_n, t_n) = 1$ the Euler scheme (63) generates $W(t)$ exactly for all l. Changing the time step for a given realization makes no difference since the coarsened Brownian motion adds the steps of the finer representation of the realization.

Order:

- One can show that, in general, the Euler-Maruyama scheme is of order $\Delta t^{\frac{1}{2}}$.
- If $\frac{dg}{dy} = 0$ then the Euler-Maruyama scheme is of order $\mathcal{O}(\Delta t^1)$ as in the deterministic case (see (68)).

7.1.2 Milstein Scheme

In particular for multiplicative noise one may want to get a higher-order approximation.

Write the stochastic differential equation again in integral form

$$\int_t^{t+\Delta t} dy = \int_t^{t+\Delta t} f(y)dt' + \int_t^{t+\Delta t} g(y)dW(t') \tag{66}$$

Note:

- for simplicity assume that f and g do not depend explicitly on time.

To obtain a higher-order approximation we need to go beyond the left-end-point rule: use a better approximation of the integrand.

Use Ito's formula for a function $v(y)$ with y satisfying the stochastic differential equation (62)

$$dv = \left(\frac{dv}{dy}f + \frac{1}{2}\frac{d^2v}{dy^2}g^2\right)\,dt + \frac{dv}{dy}g\,dW$$

Rewrite as integral

$$v(y(t')) = v(y(t)) + \int_t^{t'}\left(\frac{dv}{dy}f + \frac{1}{2}\frac{d^2v}{dy^2}g^2\right)\,dt'' + \int_t^{t'}\frac{dv}{dy}g\,dW(t'')$$

Use Ito's formula in integral form to rewrite the two integrands in (66)

$$
\begin{aligned}
y(t+\Delta t) &= y(t) + \int_t^{t+\Delta t} \left[f(y(t)) + \int_t^{t'} \left(\frac{df}{dy} f + \frac{1}{2}\frac{d^2 f}{dy^2} g^2 \right) dt'' + \int_t^{t'} \frac{df}{dy} g \, dW(t'') \right] dt' \\
&\quad + \int_t^{t+\Delta t} \left[g(y(t)) + \int_t^{t'} \left(\frac{dg}{dy} f + \frac{1}{2}\frac{d^2 g}{dy^2} g^2 \right) dt'' + \int_t^{t'} \frac{dg}{dy} g \, dW(t'') \right] dW(t') \quad \textbf{(67)} \\
&= \underbrace{y(t) + \Delta t \, f(y(t)) + g(y(t)) \, \Delta W}_{\text{Euler-Maruyama}} + h.o.t.
\end{aligned}
$$

We want the leading-order terms in the h.o.t.:
since $dW^2 = dt$ the dominant term is

$$
\int_t^{t+\Delta t} \int_t^{t'} \frac{dg(y(t''))}{dy} g(t'') \, dW(t'') \, dW(t') = \frac{dg(y(t))}{dy} g(t) \int_t^{t+\Delta t} \left\{ \int_t^{t'} dW(t'') \right\} dW(t') + ...
$$

Use the Ito integration rule (47) for polynomials

$$
\int_0^t W^n(t') dW(t') = \frac{1}{n+1} \left(W(t)^{n+1} - W(0)^{n+1} \right) - \frac{1}{2} n \int_0^t W(t')^{n-1} dt'
$$

to evaluate the integral

$$
\begin{aligned}
\int_t^{t+\Delta t} \left\{ \int_t^{t'} dW(t'') \right\} dW(t') &= \int_t^{t+\Delta t} \left[W(t') - W(t) \right] dW(t') \\
&= \frac{1}{2} \left(W(t+\Delta t)^2 - W(t)^2 - \frac{1}{2}\Delta t \right) - W(t) \left(W(t+\Delta t) - W(t) \right) \\
&= \frac{1}{2} \left(W(t+\Delta t) - W(t) \right)^2 - \frac{1}{2}\Delta t = \frac{1}{2}\Delta W^2 - \frac{1}{2}\Delta t
\end{aligned}
$$

Note:

- the integral does not vanish since in a given realization one has in general $\Delta W^2 \neq \Delta t$; only in the mean one has $\langle \Delta W^2 \rangle = \Delta t$.

Thus one obtains

$$
y_{n+1} = y_n + \Delta t \, f(y_n) + g(y_n) \, \Delta W + \frac{1}{2} \frac{dg(y_n)}{dy} g(y_n) \left(\Delta W_n^2 - \Delta t \right) \quad \textbf{(68)}
$$

Notes:

- For the convergence test proceed as for the Euler-Maruyama scheme and generate increments ΔW_n with variance Δt for the smallest Δt to be used and then generate the compound increments as in (65).

- Milstein scheme has strong convergence of order Δt^1

- for additive noise one has $\frac{dg}{dy} = 0$: the Euler-Maruyama scheme is then identical to the Milstein scheme and also becomes strong of $\mathcal{O}(\Delta t)$ (see simulations).

Sample Code:

```
function milstein
% pre-assign memory for speed
    nstepmax=1e5; time(1:nstepmax)=0; x(1:nstepmax)=0; realization(1:nstepmax)=0;
    noise(1:nstepmax)=0;
% physical parameters
    tmax=0.5; amp=.1; field=0.1; x(1)=0;
% numerical parameters
    ntjmin=3;ntjmax=12;
    nstep_max=2^ntjmax; dtmin=tmax/nstep_max;
    nconf=100;
log_plot_sol=0;
% control scheme:
%forward euler: milstein=0 %milstein: milstein=1
    milstein=1;
errormean(1:ntjmax-1)=0;
for iconf=1:nconf
    for i=1:nstepmax
        realization(i)=sqrt(dtmin)*randn;
    end
    for ntj=ntjmax:-1:ntjmin
        nt=2^ntj;
        dt(ntj)=tmax/nt;
        time(1)=0;
        for i=1:nt
            if (ntj==ntjmax)
                noise(i)=realization(i);
            else
                noise(i)=noise(2*i-1)+noise(2*i);
```

```
            end
            x(i+1)=x(i)+dt(ntj)*F(x(i),amp,field)+...
            x(i)*noise(i)+...
            1/2*milstein*x(i)*(noise(i)^2-dt(ntj));
            time(i+1)=time(i)+dt(ntj);
        end
        if (log_plot_sol==1)
            figure(1)
            plot(time(1:nt+1),x(1:nt+1));
            if ntj==1
                figure(1)
                hold all
                xlabel('time');
                ylabel('x');
            end
        end
        xfinal(ntj)=x(nt+1);
    end
    if (log_plot_sol==1)
        hold off
    end
    error(iconf,1:ntjmax-1)=abs(xfinal(2:ntjmax)-xfinal(1:ntjmax-1));
    errormean=errormean+error(iconf,:);
    figure(3)
        slope1x=[0.001,0.01];
        slope1y=[0.0001,0.001];
        slope05x=[0.001,0.1];
        slope05y=[0.0001,0.001];
        loglog(dt(1:ntjmax-1),error(iconf,1:ntjmax-1),'-o');
        hold all
        loglog(slope1x,slope1y);
        loglog(slope05x,slope05y);
        xlabel('dt');
        ylabel('difference between succ. approx');
end
hold off
errormean=errormean/nconf;
figure(4)
    loglog(dt(1:ntjmax-1),errormean(1:ntjmax-1),'-o');
    axis([1e-4 1e-1 1e-6 1e-2]);
    hold all
```

```
        loglog(slope1x,slope1y);
        loglog(slope05x,slope05y);
        xlabel('dt');
        ylabel('difference between succ. approx');
        hold off

    end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    function [F]b=F(x,amp,field)

        %xmin2=5;
        %F=-amp*x*(1-x)*(xmin2-x)/xmin2;
        F=-amp*2*x+field;

    end
```

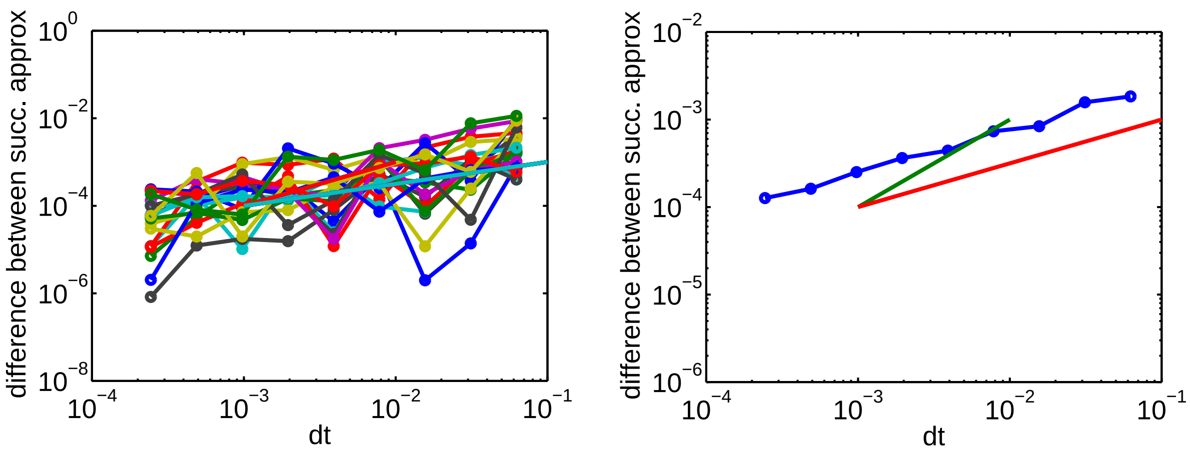


Figure 9: Euler-Maruyama scheme for $dy = -0.2y\,dt + y\,dW$: a) errors for 20 individual realizations of the Wiener process as a function of Δt. b) mean error.

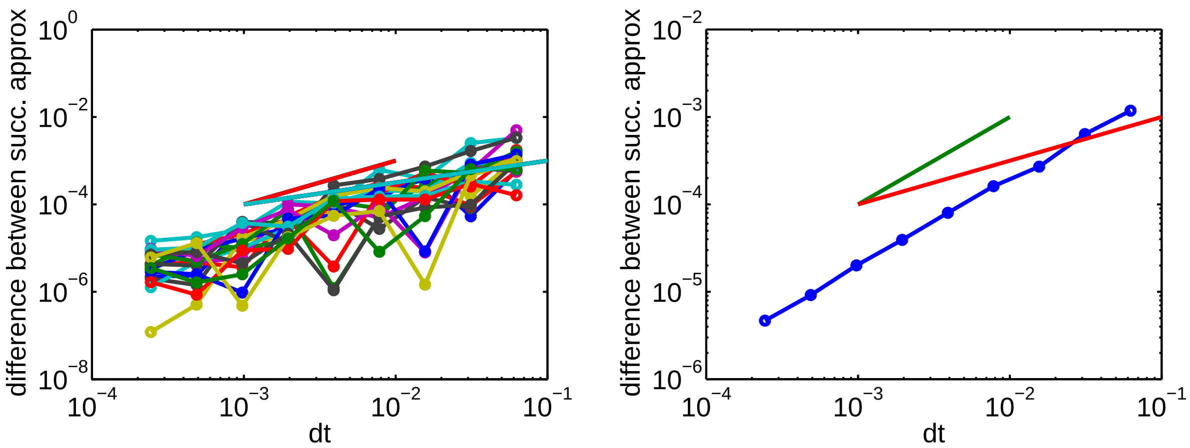


Figure 10: Milstein scheme for $dy = -0.2y\,dt + y\,dW$: a) errors for 20 individual realizations of the Wiener process as a function of Δt. b) mean error.

To go beyond $\mathcal{O}(\Delta t)$ we would have to deal with the integrals

$$\int_t^{t+\Delta t}\int_t^{t'} dW(t'')dt' \qquad \int_t^{t+\Delta t}\int_t^{t'} dt'' dW(t')$$

They cannot be expressed simply in terms of ΔW and Δt. One would have to introduce an additional random variable

$$\Delta z = \int_t^{t+\Delta t}\int_t^{t'} dW(t'')dt'.$$

One can show

$$\langle \Delta z\rangle = 0 \qquad \langle (\Delta z)^2\rangle = \frac{1}{3}\Delta t^3 \qquad \langle \Delta z \Delta W\rangle = \frac{1}{2}\Delta t^2$$

Note:

- expect Δz to be of $\mathcal{O}(\Delta t^{\frac{3}{2}})$, very loosely speaking.
- with Δz included the scheme would become of $\mathcal{O}(\Delta t^{\frac{3}{2}})$.

7.1.3 Implicit Schemes

As in deterministic case stability may require implicit scheme.

If one tries to implement a fully implicit scheme one runs into difficulties:

as example, consider backward Euler for

$$dy = aydt + bydW$$

$$y_{n+1} = \frac{y_n}{1 - a\Delta t - b\Delta W_n}$$

Since ΔW is unbounded the denominator can vanish for some ΔW_n.

$\Rightarrow$ treat only the deterministic term implicitly

i) Backward Euler

$$y_{n+1} = y_n + \Delta t F(y_{n+1}, t_{n+1}) + g(y_n, t_n)\Delta W$$

ii) Backward Milstein

$$y_{n+1} = y_n + \Delta t F(y_{n+1}, t_{n+1}) + g(y_n, t_n)\Delta W + \frac{1}{2}\frac{dg(y_n)}{dy} g(y_n)\left(\Delta W^2 - \Delta t\right)$$

7.2 Weak Approximation

If only averages and moments like $\langle y^n \rangle$ are of interest the strong approximation is not needed, i.e. any given run need not converge to the exact solution corresponding to a given realization of the noise.

In particular:

The noise used in the simulation need not be the same noise as in the stochastic differential equation.

i) Forward Euler

$$y_{n+1} = y_n + F(y_n)\Delta t + g(y_n)\Delta\tilde{W}$$

where the process $\Delta\tilde{W}$ needs to satisfy

$$\begin{aligned}
\left\langle \Delta\tilde{W} \right\rangle &= \mathcal{O}(\Delta t^2) \\
\left\langle \left(\Delta\tilde{W}\right)^2 \right\rangle &= \Delta t + \mathcal{O}(\Delta t^2) \\
\left\langle \left(\Delta\tilde{W}\right)^3 \right\rangle &= \mathcal{O}(\Delta t^2)
\end{aligned}$$

Notes:

- the noise $\Delta\tilde{W}$ need not be the Wiener process, i.e. $\Delta\tilde{W}$ need not be $\int_t^{t+\Delta t} dW(t')$. Specifically, it may differ from the Wiener process at order $\mathcal{O}(\Delta t^2)$. The conditions for the process $\Delta\tilde{W}$ can be combined to

$$|\langle \Delta\tilde{W} \rangle| + |\langle \Delta\tilde{W}^3 \rangle| + |\langle \Delta\tilde{W}^2 \rangle - \Delta t| \leq K\Delta t^2$$

- the noise could be a simple coin toss with $\Delta\tilde{W} = \pm\sqrt{\Delta t}$and

$$P(\Delta\tilde{W} = \pm\sqrt{\Delta t}) = \frac{1}{2}$$

  (it seems however that in matlab there is no random number generator for such a dichotomic noise: need to generate uniformly distributed numbers in the interval $[0, 1]$ and then check whether the number is larger or smaller than $\frac{1}{2}$. This seems to be slower in matlab than the Gaussian distribution)

- this weak Euler scheme has weak convergence of $\mathcal{O}(\Delta t)$

ii) Order-2 Weak Taylor Scheme

by keeping all the integrals in (67) and keeping also a term coming from $dW^2 dt$ at next order one gets

$$\begin{aligned} y_{n+1} &= y_n + F_n\Delta t + g_n\Delta W + \frac{1}{2}g_n\frac{dg_n}{dy}\left(\Delta W^2 - \Delta t\right) + \\ &+\frac{dF_n}{dy}g_n\Delta z + \frac{1}{2}\left(F_n\frac{dF_n}{dy} + \frac{1}{2}\frac{d^2F_n}{dy^2}g_n^2\right)\Delta t^2 + \\ &+\left(F_n\frac{dg_n}{dy} + \frac{1}{2}\frac{d^2g_n}{dy^2}\right)(\Delta W\Delta t - \Delta z) \end{aligned}$$

with

$$\Delta z = \int_t^{t+\Delta t}\int_t^{t'} dW(t'')dt'$$

For weak convergence ΔW can be replaced by $\Delta\tilde{W}$ and Δz by $\frac{1}{2}\Delta\tilde{W}\Delta t$ if

$$|\langle\Delta\tilde{W}\rangle| + |\langle\Delta\tilde{W}^3\rangle| + |\langle\Delta\tilde{W}^5\rangle| + |\langle\Delta\tilde{W}^2\rangle - \Delta t| + |\langle\Delta\tilde{W}^4\rangle - 3\Delta t^2| \le K\Delta t^3$$

Notes:

- the conditions are satisfied by Gaussian random variable and also by three-state discrete random variable with

$$P(\Delta\tilde{W} = \pm\sqrt{3\Delta t}) = \frac{1}{6} \qquad P(\Delta\tilde{W} = 0) = \frac{2}{3}$$

With that replacement get simplified scheme

$$\begin{aligned} y_{n+1} &= y_n + F_n\Delta t + g_n\Delta\tilde{W} + \frac{1}{2}g_n\frac{dg_n}{dy}\left(\Delta\tilde{W}^2 - \Delta t\right) + \frac{1}{2}\left(F_n\frac{dF_n}{dy} + \frac{1}{2}\frac{d^2F_n}{dy^2}g_n^2\right)\Delta t^2 \\ &+\frac{1}{2}\left(\frac{dF_n}{dy}g_n + F_n\frac{dg_n}{dy} + \frac{1}{2}\frac{d^2g_n}{dy^2}\right)\Delta\tilde{W}\Delta t \end{aligned}$$

Note:

- generate $\Delta\tilde{W}$: generate a uniformly distributed variable $\xi \in [0,1]$ (using `rand` in matlab)

$$\begin{aligned} 0 \le \xi \le \frac{1}{6} \qquad \Delta\tilde{W} &= +\sqrt{3\Delta t} \\ \frac{1}{6} < \xi < \frac{1}{3} \qquad \Delta\tilde{W} &= -\sqrt{3\Delta t} \\ \frac{1}{3} < \xi \le 1 \qquad \Delta\tilde{W} &= 0 \end{aligned}$$

8 Projects

Suggested Projects:

1. Thermal Ratchets [10, 2]
2. Ratchets and games [3]
3. Bifurcations [18]
4. Coherence resonance [23]
5. Coherence resonance in SIR [19]
6. Reduction of Kramers equation to Smoluchowski equation [5]
7. Stability of incoherent spiking in neuronal networks [6]
8. Localization of waves in random media [11]
9. Noise-induced neuronal network oscillations [24]
10. Noisy Kuramoto model: Fokker-Planck for oscillators [27, can be downloaded from class web site]
11. Exit from non-smooth potentials [29]
12. Black-Scholes equation for options pricing [31, 16, chapters from these books]
13. Coagulation [7, chapter III.6]