Machine Learning with Signal Processing

Part II: Stochastic Differential Equations

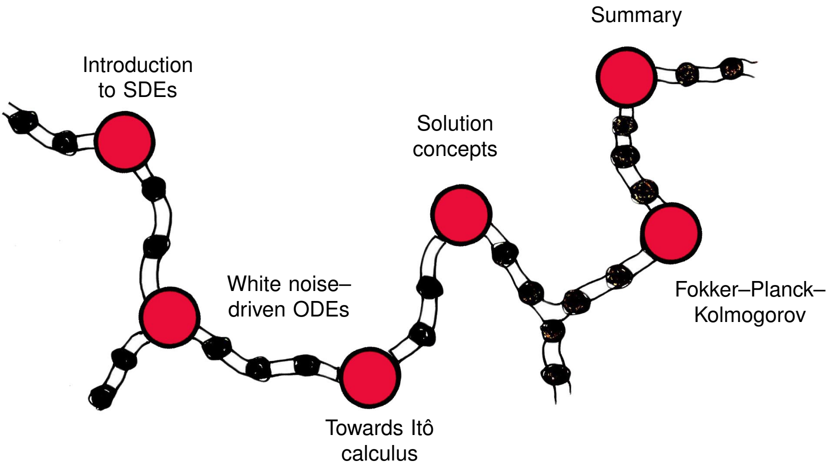
Arno Solin
Assistant Professor in Machine Learning
Department of Computer Science
Aalto University

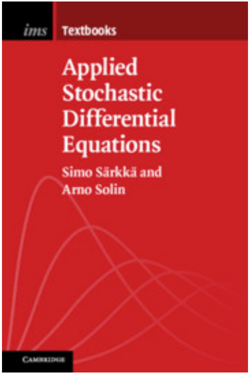
ICML 2020 Tutorial

@arnosolin

arno.solin.fi

Outline


Introduction to SDEs
White noise–driven ODEs
Towards Itô calculus
Solution concepts
Fokker–Planck–Kolmogorov
Summary



S. Särkkä and A. Solin (2019). Applied Stochastic Differential Equations. Cambridge University Press. Cambridge, UK.
Book PDF and codes for replicating examples available online.

Differential equations model how things change

- **Ordinary differential equations (ODEs)**
  (deterministic)
- **Stochastic differential equations (SDEs)**
  (stochastic)

What is a stochastic differential equation (SDE)?

- Consider an ordinary differential equation (ODE):

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}, t)$$

- Then we add white noise to the right hand side:

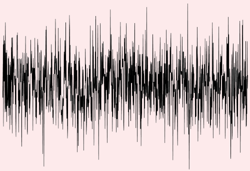
$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}, t) + \mathbf{L}(\mathbf{x}, t)\,\mathbf{w}(t)$$

- $\mathbf{f}(\mathbf{x}, t)$ is the drift function and $\mathbf{L}(\mathbf{x}, t)$ is the dispersion matrix (diffusion term)
- Now we have a stochastic differential equation (SDE)

White noise

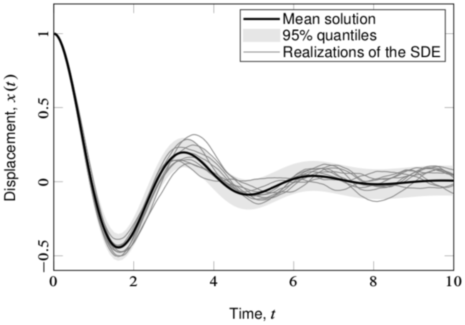
1. $\mathbf{w}(t_1)$ and $\mathbf{w}(t_2)$ are independent if $t_1 \neq t_2$
2. $t \mapsto \mathbf{w}(t)$ is a Gaussian process with mean and covariance:

$$\mathbb{E}[\mathbf{w}(t)] = \mathbf{0},$$
$$\mathbb{E}[\mathbf{w}(t)\,\mathbf{w}^{\mathsf{T}}(s)] = \delta(t - s)\,\mathbf{Q}$$



- $\mathbf{Q}$ is the spectral density of the process
- The sample path $t \mapsto \mathbf{w}(t)$ is discontinuous almost everywhere
- White noise is unbounded and it takes arbitrarily large positive and negative values at any finite interval

What does a solution of an SDE look like?



Solution paths of a stochastic spring model

$$\frac{d^2 x(t)}{dt^2} + \gamma \frac{dx(t)}{dt} + \nu^2 x(t) = w(t)$$

SDEs as white noise–driven differential equations

- Treating SDEs as white noise–driven differential equations has its limits

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}, t) + \mathbf{L}(\mathbf{x}, t)\,\mathbf{w}(t)$$

- For linear equations the approach works
- But this interpretation breaks down in the general setting:
  - The chain rule of calculus starts giving wrong answers!
  - With non-linear differential equations the behaviour becomes unexpected
  - Trying to prove the existence of solutions becomes tricky
- The source of all the problems is the everywhere discontinuous white noise $\mathbf{w}(t)$
- So how should we really formulate SDEs?

Equivalent integral equation

- We have a differential equation of the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t) + \mathbf{L}(\mathbf{x}, t)\,\mathbf{w}(t)$$

- Integrating the differential equation from t_0 to t gives:

$$\mathbf{x}(t) - \mathbf{x}(t_0) = \int_{t_0}^{t} \mathbf{f}(\mathbf{x}(t), t)\,dt + \int_{t_0}^{t} \mathbf{L}(\mathbf{x}(t), t)\,\mathbf{w}(t)\,dt$$

- The first integral is just a Riemann/Lebesgue integral
- The second integral is the problematic one due to the white noise (this is the interesting part!)

Attempt 1: Riemann integral

- In the Riemannian sense the integral would be defined as

$$\int_{t_0}^{t} \mathbf{L}(\mathbf{x}(t), t)\, \mathbf{w}(t)\, \mathrm{d}t = \lim_{n \to \infty} \sum_k \mathbf{L}(\mathbf{x}(t_k^*), t_k^*)\, \mathbf{w}(t_k^*)\, (t_{k+1} - t_k),$$

  where $t_0 < t_1 < \ldots < t_n = t$ and $t_k^* \in [t_k, t_{k+1}]$
- Upper and lower sums are defined as the selections of t_k^* such that the integrand $\mathbf{L}(\mathbf{x}(t_k^*), t_k^*)\, \mathbf{w}(t_k^*)$ has its maximum and minimum values, respectively
- The Riemann integral exists if the upper and lower sums converge to the same value

! Because white noise is discontinuous everywhere, the Riemann integral does not exist

Attempt 2: Stieltjes integral [1/2]

- A Stieltjes integral is more general and allows for discontinuous integrands
- We can interpret the increment $\mathbf{w}(t)\,\mathrm{d}t$ as increments of another process $\beta(t)$ such that

$$\int_{t_0}^{t} \mathbf{L}(\mathbf{x}(t), t)\,\mathbf{w}(t)\,\mathrm{d}t = \int_{t_0}^{t} \mathbf{L}(\mathbf{x}(t), t)\,\mathrm{d}\beta(t).$$

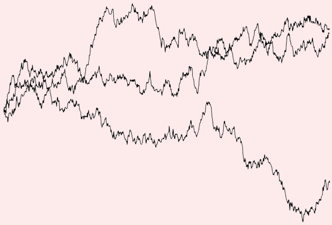
- It turns out that a suitable process for this purpose is Brownian motion...

Brownian motion

1. Gaussian increments:

$$\Delta\beta_k \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}\,\Delta t_k),$$

   where $\Delta\beta_k = \beta(t_{k+1}) - \beta(t_k)$ and $\Delta t_k = t_{k+1} - t_k$
2. Non-overlapping increments are independent



- $\mathbf{Q}$ is the diffusion matrix of the Brownian motion.
- Brownian motion $t \mapsto \beta(t)$ has discontinuous derivative everywhere
- White noise can be considered the formal derivative of Brownian motion $\mathbf{w}(t) = \mathrm{d}\beta(t)/\mathrm{d}t$

Attempt 2: Stieltjes integral [2/2]

- Stieltjes integral is defined as a limit of the form

$$\int_{t_0}^{t} \mathbf{L}(\mathbf{x}(t), t)\, \mathrm{d}\beta = \lim_{n \to \infty} \sum_{k} \mathbf{L}(\mathbf{x}(t_k^*), t_k^*)\, [\beta(t_{k+1}) - \beta(t_k)],$$

  where $t_0 < t_1 < \ldots < t_n$ and $t_k^* \in [t_k, t_{k+1}]$
- The limit t_k^* should be independent of the position on the interval $t_k^* \in [t_k, t_{k+1}]$
- For integration with respect to Brownian motion this is not the case

! Thus, the Stieltjes integral definition does not work either

Attempt 3: Lebesgue integral

- In a Lebesgue integral we could interpret $\beta(t)$ to define a 'stochastic measure'
- Essentially, this will also lead to the definition

$$\int_{t_0}^{t} \mathbf{L}(\mathbf{x}(t), t)\,\mathrm{d}\beta = \lim_{n\to\infty} \sum_{k} \mathbf{L}(\mathbf{x}(t_k^*), t_k^*)\,[\beta(t_{k+1}) - \beta(t_k)],$$

  where $t_0 < t_1 < \ldots < t_n$ and $t_k^* \in [t_k, t_{k+1}]$.
- Again, the limit should be independent of the choice $t_k^* \in [t_k, t_{k+1}]$
- Also our 'measure' is not really a sensible measure

! The Lebesgue integral does not work either

Attempt 4: Itô integral

- The solution to the problem is the Itô stochastic integral
- The idea is to fix the choice to $t_k^* = t_k$, and define the integral as

$$\int_{t_0}^{t} \mathbf{L}(\mathbf{x}(t), t) \, \mathrm{d}\beta(t) = \lim_{n \to \infty} \sum_{k} \mathbf{L}(\mathbf{x}(t_k), t_k) \, [\beta(t_{k+1}) - \beta(t_k)]$$

- This Itô stochastic integral turns out to be a sensible definition of the integral
- However, the resulting integral does not obey the computational rules of ordinary calculus
- Instead of ordinary calculus we have Itô calculus

Itô stochastic differential equations

- Consider the white noise–driven ODE

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}, t) + \mathbf{L}(\mathbf{x}, t)\,\mathbf{w}(t)$$

- This is actually defined as the Itô integral equation

$$\mathbf{x}(t) - \mathbf{x}(t_0) = \int_{t_0}^{t} \mathbf{f}(\mathbf{x}(t), t)\,\mathrm{d}t + \int_{t_0}^{t} \mathbf{L}(\mathbf{x}(t), t)\,\mathrm{d}\beta(t),$$

  which should be true for arbitrary t_0 and t

- Which can be written (considering the limits 'small') as

$$\boxed{\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}(\mathbf{x}, t)\,\mathrm{d}\beta}$$

- This is the canonical form of an Itô SDE

Connection with white noise–driven ODEs

- Let's formally divide by $\mathrm{d}t$, which gives

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}, t) + \mathbf{L}(\mathbf{x}, t) \frac{\mathrm{d}\boldsymbol{\beta}}{\mathrm{d}t}$$

- Thus we can interpret $\mathrm{d}\boldsymbol{\beta}/\mathrm{d}t$ as white noise $\mathbf{w}$ (not an entity as such, only the formal derivative)
- Note that we cannot define more general equations

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}(t), \mathbf{w}(t), t),$$

because we cannot re-interpret this as an Itô integral equation

Non-linear SDEs

- There is no general solution method for non-linear SDEs

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}(\mathbf{x}, t)\,\mathrm{d}\boldsymbol{\beta}$$

- However, numerical simulation of solution trajectories is usually possible (*e.g.*, with stochastic Runge–Kutta)
- The simplest alternative is the Euler–Maruyama method:

$$\hat{\mathbf{x}}(t_{k+1}) = \hat{\mathbf{x}}(t_k) + \mathbf{f}(\hat{\mathbf{x}}(t_k), t_k)\,\Delta t + \mathbf{L}(\hat{\mathbf{x}}(t_k), t_k)\,\Delta\boldsymbol{\beta}_k,$$

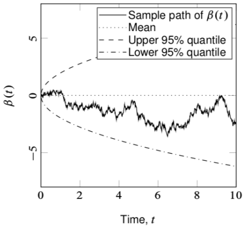
where $\Delta\boldsymbol{\beta}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}\,\Delta t)$

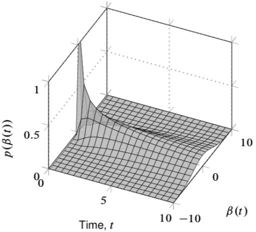
Solution concepts in SDEs

- Path of a Brownian motion which is solution to stochastic differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = w(t)$$

- Strong vs. weak solutions
- Evolution of the probability density of the solution trajectories is given by the Fokker–Planck–Kolmogorov PDE





Fokker–Planck–Kolmogorov PDE

The probability density $p(\mathbf{x}, t)$ of the solution of the SDE

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\,\mathrm{d}t + \mathbf{L}(\mathbf{x}, t)\,\mathrm{d}\beta$$

solves the Fokker–Planck–Kolmogorov PDE

$$\frac{\partial p(\mathbf{x}, t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}[f_i(x, t)\,p(\mathbf{x}, t)] + \frac{1}{2}\sum_{ij} \frac{\partial^2}{\partial x_i\,\partial x_j}\left\{[\mathbf{L}(\mathbf{x}, t)\,\mathbf{Q}\,\mathbf{L}^{\mathsf{T}}(\mathbf{x}, t)]_{ij}\,p(\mathbf{x}, t)\right\}$$

- In physics literature it is called the Fokker–Planck equation
- In stochastics it is the forward Kolmogorov equation

Summary

- Stochastic differential equations (SDE) can be seen as differential equations with a stochastic driving force
- SDEs are typical in physics, engineering, and finance applications
- A heuristic white noise formulation has problems with the chain rule, non-linearities, and solution existence
- Instead, use the Itô stochastic integral (calculus)
- Various solution concepts; in general, non-linear SDEs are tricky to solve (good schemes for simulation exist though)

Up next

3

- Three views into Gaussian processes
- (one of which is in terms of linear SDEs)

Bibliography

These references are sources for finding a more detailed overview on the topics of this part:

- S. Särkkä and A. Solin (2019). *Applied Stochastic Differential Equations*. Cambridge University Press. Cambridge, UK.
- B. Øksendal (2003). *Stochastic Differential Equations: An Introduction with Applications*. Springer, New York.
- P. E. Kloeden and E. Platen (1999). *Numerical Solution to Stochastic Differential Equations*. Springer, New York.