

# Lecture 3: Probability Distributions and Statistics of SDEs

Simo Särkkä

Aalto University

November 11, 2014

# Contents

- 1 Introduction
- 2 Fokker-Planck-Kolmogorov Equation
- 3 Mean and Covariance of SDE
- 4 Higher Order Moments of SDEs
- 5 Mean and covariance of linear SDEs
- 6 Markov Properties and Transition Densities SDEs
- 7 Summary

- Consider the **stochastic differential equation (SDE)**

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

- Each  $\mathbf{x}(t)$  is **random variable**, and we denote its **probability density** with  $p(\mathbf{x}, t)$  – or sometimes with  $p(\mathbf{x}(t))$ .
- The probability density is solution to a *partial differential equation* called **Fokker–Planck–Kolmogorov equation**.
- The mean  $\mathbf{m}(t)$  and covariance  $\mathbf{P}(t)$  are solutions of certain **ordinary differential equations** (with a catch. . .).
- For **linear SDEs** we get quite **explicit results**.

## Fokker-Planck-Kolmogorov equation

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

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

solves the **Fokker-Planck-Kolmogorov** partial differential equation

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

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

## FPK Example: Diffusion equation

Brownian motion can be defined as solution to the SDE

$$dx = d\beta.$$

If we set the diffusion constant of the Brownian motion to be  $q = 2D$ , then the FPK reduces to the diffusion equation

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2}$$

# Fokker-Planck-Kolmogorov PDE: Derivation [1/5]

- Let  $\phi(\mathbf{x})$  be an arbitrary **twice differentiable function**.
- The Itô differential of  $\phi(\mathbf{x}(t))$  is, by the **Itô formula**, given as follows:

$$\begin{aligned}d\phi &= \sum_i \frac{\partial \phi}{\partial x_i} f_i(\mathbf{x}, t) dt + \sum_i \frac{\partial \phi}{\partial x_i} [\mathbf{L}(\mathbf{x}, t) d\beta]_i \\ &+ \frac{1}{2} \sum_{ij} \left( \frac{\partial^2 \phi}{\partial x_i \partial x_j} \right) [\mathbf{L}(\mathbf{x}, t) \mathbf{Q} \mathbf{L}^\top(\mathbf{x}, t)]_{ij} dt.\end{aligned}$$

- **Taking expectations** and formally dividing by  $dt$  gives the following equation, which we will **transform into FPK**:

$$\begin{aligned}\frac{d\mathbf{E}[\phi]}{dt} &= \sum_i \mathbf{E} \left[ \frac{\partial \phi}{\partial x_i} f_i(\mathbf{x}, t) \right] \\ &+ \frac{1}{2} \sum_{ij} \mathbf{E} \left[ \left( \frac{\partial^2 \phi}{\partial x_i \partial x_j} \right) [\mathbf{L}(\mathbf{x}, t) \mathbf{Q} \mathbf{L}^\top(\mathbf{x}, t)]_{ij} \right].\end{aligned}$$

- The left hand side can now be written as follows:

$$\begin{aligned}\frac{dE[\phi]}{dt} &= \frac{d}{dt} \int \phi(\mathbf{x}) p(\mathbf{x}, t) d\mathbf{x} \\ &= \int \phi(\mathbf{x}) \frac{\partial p(\mathbf{x}, t)}{\partial t} d\mathbf{x}.\end{aligned}$$

- Recall the multidimensional integration by parts formula

$$\int_C \frac{\partial u(\mathbf{x})}{\partial x_i} v(\mathbf{x}) d\mathbf{x} = \int_{\partial C} u(\mathbf{x}) v(\mathbf{x}) n_i dS - \int_C u(\mathbf{x}) \frac{\partial v(\mathbf{x})}{\partial x_i} d\mathbf{x}.$$

- In this case, the boundary terms vanish and thus we have

$$\int \frac{\partial u(\mathbf{x})}{\partial x_i} v(\mathbf{x}) d\mathbf{x} = - \int u(\mathbf{x}) \frac{\partial v(\mathbf{x})}{\partial x_i} d\mathbf{x}.$$

- Currently, our equation **looks like this**:

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

- For the **first term on the right**, we get via **integration by parts**:

$$\begin{aligned} \mathbb{E} \left[ \frac{\partial \phi}{\partial x_i} f_i(\mathbf{x}, t) \right] &= \int \frac{\partial \phi}{\partial x_i} f_i(\mathbf{x}, t) p(\mathbf{x}, t) d\mathbf{x} \\ &= - \int \phi(\mathbf{x}) \frac{\partial}{\partial x_i} [f_i(\mathbf{x}, t) p(\mathbf{x}, t)] d\mathbf{x} \end{aligned}$$

- We now have only one term left.

- For the remaining term we use **integration by parts twice**, which gives

$$\begin{aligned} & \mathbb{E} \left[ \left( \frac{\partial^2 \phi}{\partial x_i \partial x_j} \right) [\mathbf{L}(\mathbf{x}, t) \mathbf{Q} \mathbf{L}^\top(\mathbf{x}, t)]_{ij} \right] \\ &= \int \left( \frac{\partial^2 \phi}{\partial x_i \partial x_j} \right) [\mathbf{L}(\mathbf{x}, t) \mathbf{Q} \mathbf{L}^\top(\mathbf{x}, t)]_{ij} \rho(\mathbf{x}, t) \, d\mathbf{x} \\ &= - \int \left( \frac{\partial \phi}{\partial x_j} \right) \frac{\partial}{\partial x_i} \left\{ [\mathbf{L}(\mathbf{x}, t) \mathbf{Q} \mathbf{L}^\top(\mathbf{x}, t)]_{ij} \rho(\mathbf{x}, t) \right\} \, d\mathbf{x} \\ &= \int \phi(\mathbf{x}) \frac{\partial^2}{\partial x_i \partial x_j} \left\{ [\mathbf{L}(\mathbf{x}, t) \mathbf{Q} \mathbf{L}^\top(\mathbf{x}, t)]_{ij} \rho(\mathbf{x}, t) \right\} \, d\mathbf{x} \end{aligned}$$

# Fokker-Planck-Kolmogorov PDE: Derivation [5/5]

- Our equation now **looks like this**:

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

- This can also be **written as**

$$\int \phi(\mathbf{x}) \left[ \frac{\partial p(\mathbf{x}, t)}{\partial t} + \sum_i \frac{\partial}{\partial x_i} [f_i(\mathbf{x}, t) p(\mathbf{x}, t)] \right. \\ \left. - \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \{ [\mathbf{L}(\mathbf{x}, t) \mathbf{Q} \mathbf{L}^T(\mathbf{x}, t)]_{ij} p(\mathbf{x}, t) \} \right] d\mathbf{x} = 0.$$

- But the function is  $\phi(\mathbf{x})$  **arbitrary** and thus the **term in the brackets must vanish**  $\Rightarrow$  Fokker-Planck-Kolmogorov equation.

## FPK Example: Benes SDE

The FPK for the SDE

$$dx = \tanh(x) dt + d\beta$$

can be written as

$$\begin{aligned}\frac{\partial p(x, t)}{\partial t} &= -\frac{\partial}{\partial x} (\tanh(x) p(x, t)) + \frac{1}{2} \frac{\partial^2 p(x, t)}{\partial x^2} \\ &= (\tanh^2(x) - 1) p(x, t) - \tanh(x) \frac{\partial p(x, t)}{\partial x} + \frac{1}{2} \frac{\partial^2 p(x, t)}{\partial x^2}.\end{aligned}$$

# Mean and Covariance of SDE [1/2]

- Using **Itô formula** for  $\phi(\mathbf{x}, t)$ , taking **expectations** and **dividing** by  $dt$  gives

$$\begin{aligned}\frac{d\mathbb{E}[\phi]}{dt} &= \mathbb{E}\left[\frac{\partial\phi}{\partial t}\right] + \sum_i \mathbb{E}\left[\frac{\partial\phi}{\partial x_i} f_i(\mathbf{x}, t)\right] \\ &\quad + \frac{1}{2} \sum_{ij} \mathbb{E}\left[\left(\frac{\partial^2\phi}{\partial x_i \partial x_j}\right) [\mathbf{L}(\mathbf{x}, t) \mathbf{Q} \mathbf{L}^T(\mathbf{x}, t)]_{ij}\right]\end{aligned}$$

- If we **select** the function as  $\phi(\mathbf{x}, t) = x_u$ , then we get

$$\frac{d\mathbb{E}[x_u]}{dt} = \mathbb{E}[f_u(\mathbf{x}, t)]$$

- In **vector form** this gives the **differential equation for the mean**:

$$\frac{d\mathbf{m}}{dt} = \mathbb{E}[\mathbf{f}(\mathbf{x}, t)]$$

## Mean and Covariance of SDE [2/2]

- If we **select**  $\phi(\mathbf{x}, t) = x_u x_v - m_u(t) m_v(t)$ , then we get differential equation for the **components of covariance**:

$$\begin{aligned} & \frac{d E[x_u x_v - m_u(t) m_v(t)]}{dt} \\ &= E[(x_v - m_v(t)) f_u(\mathbf{x}, t)] + E[(x_u - m_u(t)) f_v(\mathbf{x}, t)] \\ & \quad + [\mathbf{L}(\mathbf{x}, t) \mathbf{Q} \mathbf{L}^T(\mathbf{x}, t)]_{uv}. \end{aligned}$$

- The final **mean and covariance differential equations** are

$$\begin{aligned} \frac{d\mathbf{m}}{dt} &= E[\mathbf{f}(\mathbf{x}, t)] \\ \frac{d\mathbf{P}}{dt} &= E[\mathbf{f}(\mathbf{x}, t) (\mathbf{x} - \mathbf{m})^T] + E[(\mathbf{x} - \mathbf{m}) \mathbf{f}^T(\mathbf{x}, t)] \\ & \quad + E[\mathbf{L}(\mathbf{x}, t) \mathbf{Q} \mathbf{L}^T(\mathbf{x}, t)] \end{aligned}$$

- Note that the **expectations** are w.r.t.  $p(\mathbf{x}, t)$ !

- To **solve the equations**, we need to know  $p(\mathbf{x}, t)$ , the solution to the FPK.
- In **linear-Gaussian case** the first two moments indeed characterize the solution.
- Useful starting point for **Gaussian approximations of SDEs**.

# Mean and Covariance of SDE: Example

## Example (Moments of an Ornstein–Uhlenbeck process)

The Ornstein–Uhlenbeck process is

$$dx = -\lambda x dt + d\beta, \quad x(0) = x_0,$$

We have  $f(x) = -\lambda x$  and thus the differential equations for the mean and variance are thus given as

$$\begin{aligned} \frac{dm}{dt} &= -\lambda m \\ \frac{dP}{dt} &= -2\lambda P + q, \end{aligned}$$

with  $m(0) = x_0$ ,  $P(0) = 0$ . The whole state distribution is

$$p(x, t) \triangleq p(x(t)) = N(x(t) \mid m(t), P(t)).$$

# Higher Order Moments

- It is also possible to derive differential equations for the **higher order moments** of SDEs.
- But with state dimension  $n$ , we have  $n^3$  **third order moments**,  $n^4$  **fourth order moments** and so on.
- Recall that a **given scalar function**  $\phi(x)$  satisfies

$$\frac{dE[\phi(x)]}{dt} = E\left[\frac{\partial\phi(x)}{\partial x} f(x)\right] + \frac{g}{2} E\left[\frac{\partial^2\phi(x)}{\partial x^2} L^2(x)\right].$$

- If we apply this to  $\phi(x) = x^n$ :

$$\frac{dE[x^n]}{dt} = n E[x^{n-1} f(x, t)] + \frac{g}{2} n(n-1) E[x^{n-2} L^2(x)]$$

- This, in principle, is an equation for **higher order moments**.
- To actually use this, we need to use **moment closure methods**.

# Mean and covariance of linear SDEs

- Consider a **linear stochastic differential equation**

$$d\mathbf{x} = \mathbf{F}(t) \mathbf{x}(t) dt + \mathbf{u}(t) dt + \mathbf{L}(t) d\beta(t), \quad \mathbf{x}(t_0) \sim N(\mathbf{m}_0, \mathbf{P}_0).$$

- The **mean and covariance equations** are now given as

$$\begin{aligned} \frac{d\mathbf{m}(t)}{dt} &= \mathbf{F}(t) \mathbf{m}(t) + \mathbf{u}(t) \\ \frac{d\mathbf{P}(t)}{dt} &= \mathbf{F}(t) \mathbf{P}(t) + \mathbf{P}(t) \mathbf{F}^T(t) + \mathbf{L}(t) \mathbf{Q} \mathbf{L}^T(t), \end{aligned}$$

- The **general solutions** are given as

$$\begin{aligned} \mathbf{m}(t) &= \boldsymbol{\Psi}(t, t_0) \mathbf{m}(t_0) + \int_{t_0}^t \boldsymbol{\Psi}(t, \tau) \mathbf{u}(\tau) d\tau \\ \mathbf{P}(t) &= \boldsymbol{\Psi}(t, t_0) \mathbf{P}(t_0) \boldsymbol{\Psi}^T(t, t_0) \\ &\quad + \int_{t_0}^t \boldsymbol{\Psi}(t, \tau) \mathbf{L}(\tau) \mathbf{Q}(\tau) \mathbf{L}^T(\tau) \boldsymbol{\Psi}^T(t, \tau) d\tau \end{aligned}$$

# Mean and covariance of LTI SDEs

- In **LTI SDE case**

$$d\mathbf{x} = \mathbf{F} \mathbf{x}(t) dt + \mathbf{L} d\beta(t),$$

we have similarly

$$\begin{aligned}\frac{d\mathbf{m}(t)}{dt} &= \mathbf{F} \mathbf{m}(t) \\ \frac{d\mathbf{P}(t)}{dt} &= \mathbf{F} \mathbf{P}(t) + \mathbf{P}(t) \mathbf{F}^T + \mathbf{L} \mathbf{Q} \mathbf{L}^T\end{aligned}$$

- The **explicit solutions** are

$$\mathbf{m}(t) = \exp(\mathbf{F}(t - t_0)) \mathbf{m}(t_0)$$

$$\mathbf{P}(t) = \exp(\mathbf{F}(t - t_0)) \mathbf{P}(t_0) \exp(\mathbf{F}(t - t_0))^T$$

$$+ \int_{t_0}^t \exp(\mathbf{F}(t - \tau)) \mathbf{L} \mathbf{Q} \mathbf{L}^T \exp(\mathbf{F}(t - \tau))^T d\tau.$$

- Let the matrices  $\mathbf{C}(t)$  and  $\mathbf{D}(t)$  solve the **LTI differential equation**

$$\begin{pmatrix} d\mathbf{C}(t)/dt \\ d\mathbf{D}(t)/dt \end{pmatrix} = \begin{pmatrix} \mathbf{F} & \mathbf{LQ}\mathbf{L}^T \\ \mathbf{0} & -\mathbf{F}^T \end{pmatrix} \begin{pmatrix} \mathbf{C}(t) \\ \mathbf{D}(t) \end{pmatrix}$$

- Then  $\mathbf{P}(t) = \mathbf{C}(t)\mathbf{D}^{-1}(t)$  solves the differential equation

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{F}\mathbf{P}(t) + \mathbf{P}(t)\mathbf{F}^T + \mathbf{LQ}\mathbf{L}^T$$

- Thus we can solve the **covariance with matrix exponential** as well:

$$\begin{pmatrix} \mathbf{C}(t) \\ \mathbf{D}(t) \end{pmatrix} = \exp \left\{ \begin{pmatrix} \mathbf{F} & \mathbf{LQ}\mathbf{L}^T \\ \mathbf{0} & -\mathbf{F}^T \end{pmatrix} t \right\} \begin{pmatrix} \mathbf{C}(t_0) \\ \mathbf{D}(t_0) \end{pmatrix}.$$

# Markov properties of SDEs

- Definition of a **Markov process**:

$$p(\mathbf{x}(t) \mid \mathcal{X}_s) = p(\mathbf{x}(t) \mid \mathbf{x}(s)), \quad \text{for all } t \geq s.$$

where

$$\mathcal{X}_t = \{\mathbf{x}(\tau) : 0 \leq \tau \leq t\}.$$

- All **Itô processes** are **Markov processes**.
- $p(\mathbf{x}(t) \mid \mathbf{x}(s))$  is the **transition density** of the process.
- The **transition density** is also a solution to the **Fokker–Planck–Kolmogorov equation**.
- **Finite-dimensional distributions** can be constructed as

$$p(\mathbf{x}(t_0), \mathbf{x}(t_1), \dots, \mathbf{x}(t_n)) = p(\mathbf{x}(t_0)) \prod_{i=1}^n p(\mathbf{x}(t_i) \mid \mathbf{x}(t_{i-1})).$$

# Transition Densities of LTI SDEs

- The **solution of LTI SDE starting** from  $\mathbf{x}(s) \sim \mathbf{N}(\mathbf{m}(s), \mathbf{P}(s))$ :

$$\mathbf{m}(t) = \exp(\mathbf{F}(t-s)) \mathbf{m}(s)$$

$$\mathbf{P}(t) = \exp(\mathbf{F}(t-s)) \mathbf{P}(s) \exp(\mathbf{F}(t-s))^{\top} + \int_s^t \exp(\mathbf{F}(t-\tau)) \mathbf{L} \mathbf{Q} \mathbf{L}^{\top} \exp(\mathbf{F}(t-\tau))^{\top} d\tau.$$

- Starting **exactly at  $\mathbf{x}(s)$**  corresponds to  $\mathbf{m}(s) = \mathbf{x}(s)$ ,  $\mathbf{P}(s) = \mathbf{0}$ .
- Thus we have

$$p(\mathbf{x}(t) | \mathbf{x}(s)) = \mathbf{N}(\mathbf{x}(t) | \mathbf{m}(t | s), \mathbf{P}(t | s)),$$

where

$$\mathbf{m}(t | s) = \exp(\mathbf{F}(t-s)) \mathbf{x}(s)$$

$$\mathbf{P}(t | s) = \int_s^t \exp(\mathbf{F}(t-\tau)) \mathbf{L} \mathbf{Q} \mathbf{L}^{\top} \exp(\mathbf{F}(t-\tau))^{\top} d\tau.$$

# Transition Densities of LTI SDEs (cont.)

- Let  $p(\mathbf{x}(t_{k+1}) | \mathbf{x}(t_k))$  and  $\Delta t_k = t_{k+1} - t_k$ , which gives:

$$\begin{aligned} & p(\mathbf{x}(t_{k+1}) | \mathbf{x}(t_k)) \\ &= N\left(\mathbf{x}(t_{k+1}) | \exp(\mathbf{F} \Delta t_k) \mathbf{x}(t_k), \int_0^{\Delta t_k} \exp(\mathbf{F}(\Delta t_k - \tau)) \mathbf{L} \mathbf{Q} \mathbf{L}^T \exp(\mathbf{F}(\Delta t_k - \tau))^T d\tau\right) \end{aligned}$$

- This is now **equivalent to a discrete-time system**

$$\mathbf{x}(t_{k+1}) = \mathbf{A}_k \mathbf{x}(t_k) + \mathbf{q}_k, \quad \mathbf{q}_k \sim N(\mathbf{0}, \Sigma_k)$$

where

$$\mathbf{A}_k = \exp(\mathbf{F} \Delta t_k)$$

$$\Sigma_k = \int_0^{\Delta t_k} \exp(\mathbf{F}(\Delta t_k - \tau)) \mathbf{L} \mathbf{Q} \mathbf{L}^T \exp(\mathbf{F}(\Delta t_k - \tau))^T d\tau.$$

# Transition Densities of LTI SDEs: Example

## Example (Discretized Wiener velocity model)

$$\frac{d^2x(t)}{dt^2} = w(t).$$

In more rigorous Itô SDE form this model can be written as

$$\begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}}_{\mathbf{F}} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} dt + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{\mathbf{L}} d\beta(t),$$

We now get (note that  $\mathbf{F}^2 = 0$ ):

$$\mathbf{A} = \exp(\mathbf{F} \Delta t) = \mathbf{I} + \mathbf{F} \Delta t = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix}$$

$$\Sigma = \int_0^{\Delta t} \begin{pmatrix} 1 & \Delta t - \tau \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & q \end{pmatrix} \begin{pmatrix} 1 & \Delta t - \tau \\ 0 & 1 \end{pmatrix}^T d\tau = \begin{pmatrix} \frac{1}{3} \Delta t^3 & \frac{1}{2} \Delta t^2 \\ \frac{1}{2} \Delta t^2 & \Delta t \end{pmatrix} q.$$

# Summary

- The **probability density** of SDE solution  $\mathbf{x}(t)$  solves the **Fokker–Planck–Kolmogorov (FKP) partial differential equation**.
- The **mean  $\mathbf{m}(t)$  and covariance  $\mathbf{P}(t)$**  of the solution solve a pair of *ordinary differential equations*.
- In non-linear case, the **expectations in the mean and covariance equations** cannot be solved without knowing the **whole probability density**.
- For **higher moment moments** we can derive (theoretical) differential equations as well—can be approximated with **moment closure**.
- SDEs are **Markov processes** and can be characterised via **transition densities**.
- For **linear SDEs**, we can solve the **probability density, transition densities**, and all the **moments** explicitly.

(Wiener velocity model)