Simo Särkkä

Aalto University, Finland (visiting at Oxford University, UK)

November 21, 2013
Contents

1. Introduction
2. Gaussian approximations
3. Linearization and sigma-point approximations
4. Taylor series of ODEs
5. Itô-Taylor series of SDEs
6. Stochastic Runge–Kutta, Weak Approximations
7. Summary
Overview of Numerical Methods

Gaussian approximations:
- Approximations of mean and covariance equations.
- Gaussian assumed density approximations.
- Statistical linearization.

Numerical simulation of SDEs:
- Itô–Taylor series.
- Euler–Maruyama method and Milstein’s method.
- Stochastic Runge–Kutta.

Other methods (not covered on this lecture):
- Approximations of higher order moments.
- Approximations of Fokker–Planck–Kolmogorov PDE.
Consider the stochastic differential equation (SDE)

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

The mean and covariance differential equations are

\[
\frac{dm}{dt} = E[f(\mathbf{x}, t)]
\]

\[
\frac{dP}{dt} = E \left[ f(\mathbf{x}, t) (\mathbf{x} - m)^T \right] + E \left[ (\mathbf{x} - m) f^T(\mathbf{x}, t) \right] + E \left[ L(\mathbf{x}, t) Q L^T(\mathbf{x}, t) \right]
\]

Note that the expectations are w.r.t. \( p(\mathbf{x}, t) \)!
Gaussian approximations [1/5]

- The **mean and covariance equations** explicitly:

\[
\frac{dm}{dt} = \int f(x, t) \, p(x, t) \, dx
\]

\[
\frac{dP}{dt} = \int f(x, t) (x - m)^T \, p(x, t) \, dx + \int (x - m) \, f^T(x, t) \, p(x, t) \, dx
\]

\[
+ \int L(x, t) \, Q \, L^T(x, t) \, p(x, t) \, dx.
\]

- In Gaussian **assumed density approximation** we assume

\[
p(x, t) \approx N(x | m(t), P(t)).
\]
Gaussian approximation to SDE can be obtained by integrating the following differential equations from the initial conditions \( m(0) = E[x(0)] \) and \( P(0) = \text{Cov}[x(0)] \) to the target time \( t \):

\[
\frac{d m}{d t} = \int f(x, t) \mathcal{N}(x | m, P) \, dx
\]

\[
\frac{d P}{d t} = \int f(x, t) (x - m)^T \mathcal{N}(x | m, P) \, dx
\]

\[
+ \int (x - m) f^T(x, t) \mathcal{N}(x | m, P) \, dx
\]

\[
+ \int L(x, t) Q L^T(x, t) \mathcal{N}(x | m, P) \, dx.
\]
If we denote the Gaussian expectation as

\[ E_N[g(x)] = \int g(x) \ N(x \mid m, P) \ dx \]

the mean and covariance equations can be written as

\[
\frac{dm}{dt} = E_N[f(x, t)]
\]

\[
\frac{dP}{dt} = E_N[(x - m)f^T(x, t)] + E_N[f(x, t)(x - m)^T] + E_N[L(x, t)QL^T(x, t)].
\]
Theorem

Let \( f(x, t) \) be differentiable with respect to \( x \) and let \( x \sim N(m, P) \). Then the following identity holds:

\[
\int f(x, t) (x - m)^T N(x \mid m, P) \, dx = \left[ \int F_x(x, t) \, N(x \mid m, P) \, dx \right] P,
\]

where \( F_x(x, t) \) is the Jacobian matrix of \( f(x, t) \) with respect to \( x \).
Gaussian approximations II

Gaussian approximation to SDE can be obtained by integrating the following differential equations from the initial conditions \(m(0) = \mathbb{E}[x(0)]\) and \(P(0) = \text{Cov}[x(0)]\) to the target time \(t\):

\[
\frac{dm}{dt} = \mathbb{E}_N[f(x, t)] \\
\frac{dP}{dt} = P \mathbb{E}_N[F_x(x, t)]^T + \mathbb{E}_N[F_x(x, t)] P + \mathbb{E}_N[L(x, t) Q L^T(x, t)],
\]

where \(\mathbb{E}_N[\cdot]\) denotes the expectation with respect to \(x \sim N(m, P)\).
We need to compute following kind of Gaussian integrals:

\[ E_N[g(x, t)] = \int g(x, t) \ N(x \mid m, P) \ dx \]

We can borrow methods from filtering theory.

Linearize the drift \( f(x, t) \) around the mean \( m \) as follows:

\[ f(x, t) \approx f(m, t) + F_x(m, t) \ (x - m), \]

Approximate the expectation of the diffusion part as

\[ L(x, t) \approx L(m, t). \]
Linearization approximation of SDE

Linearization based approximation to SDE can be obtained by integrating the following differential equations from the initial conditions \( m(0) = E[x(0)] \) and \( P(0) = \text{Cov}[x(0)] \) to the target time \( t \):

\[
\frac{dm}{dt} = f(m, t)
\]

\[
\frac{dP}{dt} = PF_x^T(m, t) + F_x(m, t)P + L(m, t)QL^T(m, t).
\]

- Used in extended Kalman filter (EKF).
Cubature integration \([1/3]\)

- Gauss–Hermite cubatures:

\[
\int f(x, t) \ N(x \mid m, P) \ dx \approx \sum_i W(i) \ f(x(i), t).
\]

- The sigma points (abscissas) \(x(i)\) and weights \(W(i)\) are determined by the integration rule.

- In multidimensional Gauss-Hermite integration, unscented transform, and cubature integration we select:

\[
x(i) = m + \sqrt{P} \xi_i.
\]

- The matrix square root is defined by \(P = \sqrt{P} \sqrt{P}^T\) (typically Cholesky factorization).

- The vectors \(\xi_i\) are determined by the integration rule.
In Gauss–Hermite integration the vectors and weights are selected as cartesian products of 1d Gauss–Hermite integration.

Unscented transform uses:

\[ \xi_0 = 0 \]
\[ \xi_i = \begin{cases} \sqrt{\lambda + ne_i}, & i = 1, \ldots, n \\ -\sqrt{\lambda + ne_{i-n}}, & i = n + 1, \ldots, 2n \end{cases} \]

and \( W^{(0)} = \lambda/(n + \kappa) \), and \( W^{(i)} = 1/[2(n + \kappa)] \) for \( i = 1, \ldots, 2n \).

Cubature method (spherical 3rd degree):

\[ \xi_i = \begin{cases} ne_i, & i = 1, \ldots, n \\ -ne_{i-n}, & i = n + 1, \ldots, 2n \end{cases} \]

and \( W^{(i)} = 1/(2n) \) for \( i = 1, \ldots, 2n \).
Cubature integration [3/3]

Sigma-point approximation of SDE

Sigma-point based approximation to SDE:

\[
\frac{d\mathbf{m}}{dt} = \sum_i W^{(i)} f(\mathbf{m} + \sqrt{\mathbf{P}} \xi_i, t)
\]

\[
\frac{d\mathbf{P}}{dt} = \sum_i W^{(i)} f(\mathbf{m} + \sqrt{\mathbf{P}} \xi_i, t) \xi_i^T \sqrt{\mathbf{P}}^T
\]

\[
+ \sum_i W^{(i)} \sqrt{\mathbf{P}} \xi_i f^T(\mathbf{m} + \sqrt{\mathbf{P}} \xi_i, t)
\]

\[
+ \sum_i W^{(i)} L(\mathbf{m} + \sqrt{\mathbf{P}} \xi_i, t) Q L^T(\mathbf{m} + \sqrt{\mathbf{P}} \xi_i, t).
\]

- Use in (continuous-time) unscented Kalman filter (UKF) and (continuous-time) cubature-based Kalman filters (GHKF, CKF, etc.).
Taylor series of ODEs vs. Itô-Taylor series of SDEs

- **Taylor series** expansions (in time direction) are classical methods for approximating solutions of deterministic ordinary differential equations (ODEs).
- Largely superseded by **Runge–Kutta** type of derivative free methods (whose theory is based on Taylor series).
- **Itô-Taylor series** can be used for approximating solutions of SDEs—direct generalization of Taylor series for ODEs.
- **Stochastic Runge–Kutta** methods are not as easy to use as their deterministic counterparts.
- It is easier to understand **Itô-Taylor series** by understanding Taylor series (for ODEs) first.
Consider the following ordinary differential equation (ODE):

\[
\frac{dx(t)}{dt} = f(x(t), t), \quad x(t_0) = \text{given},
\]

Integrating both sides gives

\[
x(t) = x(t_0) + \int_{t_0}^{t} f(x(\tau), \tau) \, d\tau.
\]

If the function \( f \) is differentiable, we can also write \( t \mapsto f(x(t), t) \) as the solution to the differential equation

\[
\frac{df(x(t), t)}{dt} = \frac{\partial f}{\partial t}(x(t), t) + \sum_i f_i(x(t), t) \frac{\partial f}{\partial x_i}(x(t), t).
\]
The integral form of this is

\[ f(x(t), t) = f(x(t_0), t_0) + \int_{t_0}^{t} \left[ \frac{\partial f}{\partial t}(x(\tau), \tau) + \sum_i f_i(x(\tau), \tau) \frac{\partial f}{\partial x_i}(x(\tau), \tau) \right] d\tau. \]

Let's define the linear operator

\[ \mathcal{L} g = \frac{\partial g}{\partial t} + \sum_i f_i \frac{\partial g}{\partial x_i} \]

We can now rewrite the integral equation as

\[ f(x(t), t) = f(x(t_0), t_0) + \int_{t_0}^{t} \mathcal{L} f(x(\tau), \tau) \, d\tau. \]
By substituting this into the original integrated ODE gives

\[ x(t) = x(t_0) + \int_{t_0}^{t} f(x(\tau), \tau) \, d\tau \]

\[ = x(t_0) + \int_{t_0}^{t} [f(x(t_0), t_0) + \int_{t_0}^{\tau} L f(x(\tau), \tau) \, d\tau] \, d\tau \]

\[ = x(t_0) + f(x(t_0), t_0) (t - t_0) + \int_{t_0}^{t} \int_{t_0}^{\tau} L f(x(\tau), \tau) \, d\tau \, d\tau. \]

The term \( L f(x(t), t) \) solves the differential equation

\[ \frac{d[L f(x(t), t)]}{dt} = \frac{\partial[L f(x(t), t)]}{\partial t} + \sum_i f_i(x(t), t) \frac{\partial[L f(x(t), t)]}{\partial x_i} \]

\[ = L^2 f(x(t), t). \]
In integral form this is

\[ L \mathbf{f}(\mathbf{x}(t), t) = L \mathbf{f}(\mathbf{x}(t_0), t_0) + \int_{t_0}^{t} L^2 \mathbf{f}(\mathbf{x}(\tau), \tau) \, d\tau. \]

Substituting into the equation of \( \mathbf{x}(t) \) then gives

\begin{align*}
\mathbf{x}(t) &= \mathbf{x}(t_0) + \mathbf{f}(\mathbf{x}(t_0), t)(t - t_0) \\
&\quad + \int_{t_0}^{t} \int_{t_0}^{\tau} \left[ L \mathbf{f}(\mathbf{x}(t_0), t_0) + \int_{t_0}^{\tau} L^2 \mathbf{f}(\mathbf{x}(\tau'), \tau') \, d\tau' \right] \, d\tau' \, d\tau \\
&= \mathbf{x}(t_0) + \mathbf{f}(\mathbf{x}(t_0), t_0)(t - t_0) + \frac{1}{2} L \mathbf{f}(\mathbf{x}(t_0), t_0)(t - t_0)^2 \\
&\quad + \int_{t_0}^{t} \int_{t_0}^{\tau} \int_{t_0}^{\tau'} L^2 \mathbf{f}(\mathbf{x}(\tau), \tau) \, d\tau' \, d\tau \, d\tau
\end{align*}
If we continue this procedure ad infinitum, we obtain the following Taylor series expansion for the solution of the ODE:

\[
x(t) = x(t_0) + f(x(t_0), t_0) (t - t_0) + \frac{1}{2!} \mathcal{L} f(x(t_0), t_0) (t - t_0)^2 \\
+ \frac{1}{3!} \mathcal{L}^2 f(x(t_0), t_0) (t - t_0)^3 + \ldots
\]

where

\[
\mathcal{L} = \frac{\partial}{\partial t} + \sum_i f_i \frac{\partial}{\partial x_i}
\]

The Taylor series for a given function \(x(t)\) can be obtained by setting \(f(t) = dx(t)/dt\).
Consider the following SDE
\[ d\mathbf{x} = f(\mathbf{x}(t), t) \, dt + L(\mathbf{x}(t), t) \, d\beta. \]

In integral form this is
\[ \mathbf{x}(t) = \mathbf{x}(t_0) + \int_{t_0}^{t} f(\mathbf{x}(\tau), \tau) \, d\tau + \int_{t_0}^{t} L(\mathbf{x}(\tau), \tau) \, d\beta(\tau). \]

Applying Itô formula to \( f(\mathbf{x}(t), t) \) gives
\[
\begin{align*}
    d\mathbf{f}(\mathbf{x}(t), t) &= \frac{\partial \mathbf{f}(\mathbf{x}(t), t)}{\partial t} \, dt + \sum_{u} \frac{\partial \mathbf{f}(\mathbf{x}(t), t)}{\partial x_u} \, f_u(\mathbf{x}(t), t) \, dt \\
    &\quad + \sum_{u} \frac{\partial \mathbf{f}(\mathbf{x}(t), t)}{\partial x_u} \left[ L(\mathbf{x}(t), t) \, d\beta(\tau) \right]_u \\
    &\quad + \frac{1}{2} \sum_{uv} \frac{\partial^2 \mathbf{f}(\mathbf{x}(t), t)}{\partial x_u \partial x_v} \left[ L(\mathbf{x}(t), t) \, Q \, L^T(\mathbf{x}(t), t) \right]_{uv} \, dt
\end{align*}
\]
Similarly for $L(x(t), t)$ we get via Itô formula:

$$
\begin{align*}
\mathrm{d}L(x(t), t) &= \frac{\partial L(x(t), t)}{\partial t} \, \mathrm{d}t + \sum_u \frac{\partial L(x(t), t)}{\partial x_u} \, f_u(x(t), t) \, \mathrm{d}t \\
&\quad + \sum_u \frac{\partial L(x(t), t)}{\partial x_u} \left[ L(x(t), t) \, \mathrm{d}\beta(\tau) \right]_u \\
&\quad + \frac{1}{2} \sum_{uv} \frac{\partial^2 L(x(t), t)}{\partial x_u \partial x_v} \left[ L(x(t), t) \, Q \, L^T(x(t), t) \right]_{uv} \, \mathrm{d}t
\end{align*}
$$
In integral form these can be written as

\[ f(x(t), t) = f(x(t_0), t_0) + \int_{t_0}^{t} \frac{\partial f(x(\tau), \tau)}{\partial t} \, d\tau + \int_{t_0}^{t} \sum_{u} \frac{\partial f(x(\tau), \tau)}{\partial x_u} f_u(x(\tau), \tau) \, d\tau \]

\[ + \int_{t_0}^{t} \sum_{u} \frac{\partial f(x(\tau), \tau)}{\partial x_u} [L(x(\tau), \tau) \, d\beta(\tau)]_u \]

\[ + \int_{t_0}^{t} \frac{1}{2} \sum_{uv} \frac{\partial^2 f(x(\tau), \tau)}{\partial x_u \partial x_v} [L(x(\tau), \tau) Q L^T(x(\tau), \tau)]_{uv} \, d\tau \]

\[ L(x(t), t) = L(x(t_0), t_0) + \int_{t_0}^{t} \frac{\partial L(x(\tau), \tau)}{\partial t} \, d\tau + \int_{t_0}^{t} \sum_{u} \frac{\partial L(x(\tau), \tau)}{\partial x_u} f_u(x(\tau), \tau) \, d\tau \]

\[ + \int_{t_0}^{t} \sum_{u} \frac{\partial L(x(\tau), \tau)}{\partial x_u} [L(x(\tau), \tau) \, d\beta(\tau)]_u \]

\[ + \int_{t_0}^{t} \frac{1}{2} \sum_{uv} \frac{\partial^2 L(x(\tau), \tau)}{\partial x_u \partial x_v} [L(x(\tau), \tau) Q L^T(x(\tau), \tau)]_{uv} \, d\tau \]
Let’s define operators

\[ \mathcal{L}_t \mathbf{g} = \frac{\partial \mathbf{g}}{\partial t} + \sum_u \frac{\partial \mathbf{g}}{\partial x_u} f_u + \frac{1}{2} \sum_{uv} \frac{\partial^2 \mathbf{g}}{\partial x_u \partial x_v} [\mathbf{L} \mathbf{Q} \mathbf{L}^T]_{uv} \]

\[ \mathcal{L}_{\beta, \nu} \mathbf{g} = \sum_u \frac{\partial \mathbf{g}}{\partial x_u} \mathbf{L}_{uv}, \quad \nu = 1, \ldots, n. \]

Then we can conveniently write

\[ \mathbf{f}(\mathbf{x}(t), t) = \mathbf{f}(\mathbf{x}(t_0), t_0) + \int_{t_0}^{t} \mathcal{L}_t \mathbf{f}(\mathbf{x}(\tau), \tau) \, d\tau + \sum_v \int_{t_0}^{t} \mathcal{L}_{\beta, \nu} \mathbf{f}(\mathbf{x}(\tau), \tau) \, d\beta_v(\tau) \]

\[ \mathbf{L}(\mathbf{x}(t), t) = \mathbf{L}(\mathbf{x}(t_0), t_0) + \int_{t_0}^{t} \mathcal{L}_t \mathbf{L}(\mathbf{x}(\tau), \tau) \, d\tau + \sum_v \int_{t_0}^{t} \mathcal{L}_{\beta, \nu} \mathbf{L}(\mathbf{x}(\tau), \tau) \, d\beta_v(\tau) \]
If we now substitute these into equation of $x(t)$, we get

$$x(t) = x(t_0) + f(x(t_0), t_0) (t - t_0) + L(x(t_0), t_0) (\beta(t) - \beta(t_0))$$

$$+ \int_{t_0}^{t} \int_{t_0}^{\tau} \mathcal{L}_{t, \tau} f(x(\tau), \tau) \, d\tau \, d\tau + \sum_{v} \int_{t_0}^{t} \int_{t_0}^{\tau} \mathcal{L}_{\beta, v} f(x(\tau), \tau) \, d\beta_v(\tau) \, d\tau$$

$$+ \int_{t_0}^{t} \int_{t_0}^{\tau} \mathcal{L}_{t, \tau} L(x(\tau), \tau) \, d\tau \, d\beta(\tau) + \sum_{v} \int_{t_0}^{t} \int_{t_0}^{\tau} \mathcal{L}_{\beta, v} L(x(\tau), \tau) \, d\beta_v(\tau) \, d\beta(\tau).$$

This can be seen to have the form

$$x(t) = x(t_0) + f(x(t_0), t_0) (t - t_0) + L(x(t_0), t_0) (\beta(t) - \beta(t_0)) + r(t)$$

$r(t)$ is a remainder term.

By neglecting the remainder we get the Euler–Maruyama method.
Euler–Maruyama method

Draw \( \hat{x}_0 \sim p(x_0) \) and divide time \([0, t]\) interval into \( K \) steps of length \( \Delta t \). At each step \( k \) do the following:

1. Draw random variable \( \Delta \beta_k \) from the distribution (where \( t_k = k \Delta t \))
   \[
   \Delta \beta_k \sim N(0, Q \Delta t).
   \]

2. Compute
   \[
   \hat{x}(t_{k+1}) = \hat{x}(t_k) + f(\hat{x}(t_k), t_k) \Delta t + L(\hat{x}(t_k), t_k) \Delta \beta_k.
   \]
Order of convergence

- **Strong** order of convergence $\gamma$:
  \[
  E \left[ |x(t_n) - \hat{x}(t_n)| \right] \leq K \Delta t^\gamma
  \]

- **Weak** order of convergence $\alpha$:
  \[
  | E [g(x(t_n))] - E [g(\hat{x}(t_n))] | \leq K \Delta t^\alpha,
  \]
  for any function $g$.

- **Euler–Maruyama method** has strong order $\gamma = 1/2$ and weak order $\alpha = 1$.

- The reason for $\gamma = 1/2$ is the following term in the remainder:
  \[
  \sum_v \int_{t_0}^t \int_{t_0}^\tau L_{\beta,\nu} L(x(\tau), \tau) \, d\beta_{\nu}(\tau) \, d\beta(\tau).
  \]
If we now expand the problematic term using Itô formula, we get

\[ x(t) = x(t_0) + f(x(t_0), t_0)(t - t_0) + L(x(t_0), t_0)(\beta(t) - \beta(t_0)) + \sum \mathcal{L}_{\beta,v} L(x(t_0), t_0) \int_{t_0}^{t} \int_{t_0}^{\tau} d\beta_v(\tau) \ d\beta(\tau) + \text{remainder}. \]

Notice the iterated Itô integral appearing in the equation:

\[ \int_{t_0}^{t} \int_{t_0}^{\tau} d\beta_v(\tau) \ d\beta(\tau). \]

Computation of general iterated Itô integrals is non-trivial.

We usually also need to approximate the iterated Itô integrals – different ways for strong and weak approximations.
Draw $\hat{x}_0 \sim p(x_0)$, and at each step $k$ do the following:

1. Jointly draw the following:
   
   $$\Delta \beta_k = \beta(t_{k+1}) - \beta(t_k)$$
   $$\Delta \chi_{v,k} = \int_{t_k}^{t_{k+1}} \int_{t_k}^{\tau} d\beta_v(\tau) \, d\beta(\tau).$$

2. Compute
   
   $$\hat{x}(t_{k+1}) = \hat{x}(t_k) + f(\hat{x}(t_k), t_k) \Delta t + L(\hat{x}(t_k), t_k) \Delta \beta_k$$
   
   $$+ \sum_v \left[ \sum_u \frac{\partial L}{\partial x_u}(\hat{x}(t_k), t_k) L_{uv}(\hat{x}(t_k), t_k) \right] \Delta \chi_{v,k}.$$
Milstein’s method [3/4]

- The strong and weak orders of the above method are both 1.
- The difficulty is in drawing the iterated stochastic integral jointly with the Brownian motion.
- If the noise is additive, that is, $L(x, t) = L(t)$ then Milstein’s algorithm reduces to Euler–Maruyama.
- Thus in additive noise case, the strong order of Euler–Maruyama is 1 as well.
- In scalar case we can compute the iterated stochastic integral:

$$
\int_{t_0}^{t} \int_{t_0}^\tau d\beta(\tau) \ d\beta(\tau) = \frac{1}{2} \left[ (\beta(t) - \beta(t_0))^2 - q(t - t_0) \right]
$$
Scalar Milstein’s method

Draw $\hat{x}_0 \sim p(x_0)$, and at each step $k$ do the following:

1. Draw random variable $\Delta \beta_k$ from the distribution (where $t_k = k \Delta t$)

   $\Delta \beta_k \sim N(0, q \Delta t)$.

2. Compute

   $$\hat{x}(t_{k+1}) = \hat{x}(t_k) + f(\hat{x}(t_k), t_k) \Delta t + L(x(t_k), t_k) \Delta \beta_k$$

   $$+ \frac{1}{2} \frac{\partial L}{\partial x}(\hat{x}(t_k), t_k) L(\hat{x}(t_k), t_k) (\Delta \beta_k^2 - q \Delta t).$$
Higher Order Methods

- By taking more terms into the expansion, can form methods of arbitrary order.
- The high order iterated Itô integrals will be increasingly hard to simulate.
- However, if $L$ does not depend on the state, we can get up to strong order 1.5 without any iterated integrals.
- For that purpose we need to expand the following terms using the Itô formula (see the lecture notes):

$$L_t f(x(t), t)$$
$$L_{\beta,v} f(x(t), t).$$
When $L$ and $Q$ are constant, we get the following algorithm. Draw $\hat{x}_0 \sim p(x_0)$, and at each step $k$ do the following:

1. Draw random variables $\Delta \zeta_k$ and $\Delta \beta_k$ from the joint distribution 
   \[
   \begin{pmatrix}
   \Delta \zeta_k \\
   \Delta \beta_k
   \end{pmatrix} \sim N \left( \begin{pmatrix}
   0 \\
   0
   \end{pmatrix}, \begin{pmatrix}
   Q \Delta t^3/3 & Q \Delta t^2/2 \\
   Q \Delta t^2/2 & Q \Delta t
   \end{pmatrix} \right). 
   \]

2. Compute 
   \[
   \hat{x}(t_{k+1}) = \hat{x}(t_k) + f(\hat{x}(t_k), t_k) \Delta t + L \Delta \beta_k + a_k \frac{(t - t_0)^2}{2} + \sum_v b_{v,k} \Delta \zeta_k 
   \]

   \[
   a_k = \frac{\partial f}{\partial t} + \sum_u \frac{\partial f}{\partial x_u} f_u + \frac{1}{2} \sum_{uv} \frac{\partial^2 f}{\partial x_u \partial x_v} [L Q L^T]_{uv} 
   \]

   \[
   b_{v,k} = \sum_u \frac{\partial f}{\partial x_u} L_{uv}. 
   \]
Stochastic versions of Stochastic Runge–Kutta methods are not as simple as in the case of deterministic equations.

In practice, stochastic Runge–Kutta methods can be derived, for example, by replacing the closed form derivatives in the Milstein’s method with finite differences.

We still cannot get rid of the iterated Itô integral occurring in Milstein’s method.

Stochastic Runge–Kutta methods cannot be derived as simple extensions of the deterministic Runge–Kutta methods.

A number of stochastic Runge–Kutta methods have also been presented by Kloeden et al. (1994); Kloeden and Platen (1999) as well as by Rößler (2006).
If we are only interested in the statistics of SDE solutions, weak approximations are enough.

We can sometimes use Girsanov theorem to enhance/ease the simulation.

In weak approximations iterated Itô integrals can be replaced with simpler approximations with right statistics.

For details, see Kloeden and Platen (1999).
Gaussian approximations of SDEs can be formed by assuming Gaussianity in the mean and covariance equations. The resulting equations can be numerically solved using linearization or cubature integration (sigma-point methods).

Itô–Taylor series is a stochastic counterpart of Taylor series for ODEs. With first order truncation of Itô–Taylor series we get Euler–Maruyama method.

Including additional stochastic term leads to Milstein’s method. Computation/approximation of iterated Itô integrals can be hard and needed for implementing the methods.

In additive noise case we get a simple 1.5 strong order method.

Stochastic Runge–Kutta methods involve the iterated Itô integrals as well – not simple extensions of the deterministic methods.

Weak approximations are simpler and enough for approximating the statistics of SDEs.