
A Gaussian approximation for stochastic non-linear dynamical processes with annihilation

Wim Wiegerinck **Bert Kappen**
SNN, Radboud University Nijmegen
Geert Grooteplein 21, int 126
6525 EZ Nijmegen, The Netherlands
{w.wiegerinck,b.kappen}@science.ru.nl

Abstract

We consider a stochastic nonlinear dynamical process with annihilation of particles. This process can be viewed as the continuous time version of the extended Kalman filter/smoother. It also plays an important role in stochastic optimal control theory. We derive a Gaussian approximation for this process. With the use of the path integral formalism we derive Euler-Lagrange equations for the mode. Furthermore, we derive a linear noise approximation to estimate the size of the fluctuations around the mode, and estimates of the partition function, based on the mode and Gaussian corrections. Numerical experiments confirm the validity of the approximation method. In addition, they show that the Gaussian correction provides a significant improvement of the estimate of the partition function.

1 A stochastic dynamical process with annihilation

We consider the following stochastic simulation of particles with nonlinear dynamics, additive Gaussian noise and stochastic annihilation of particles,

$$\begin{aligned} dx &= b(x, t)dt + d\xi \\ x &= x + dx, \quad \text{with probability } 1 - V(x, t)dt \\ x &= \dagger, \quad \text{with probability } V(x, t)dt \end{aligned} \tag{1}$$

where $d\xi$ is the noise process with strength ν , (i.e., $d\xi = N(0, \nu dt)$), and \dagger denotes that the particle is taken out of the simulation. (Note that, for convenience, we restrict to 1-D throughout this paper. Results are easily generalized to higher dimensional spaces.)

We initialize at t_0, x_0 , and we can describe the simulation by the unnormalized density $\rho(x, t|x_0, t_0)$. The density evolves in time according to

$$\partial_t \rho(x, t|x_0, t_0) = -\partial_x (b(x, t)\rho(x, t|x_0, t_0)) + \frac{1}{2}\nu \partial_x^2 \rho(x, t|x_0, t_0) - V(x, t)\rho(x, t|x_0, t_0). \tag{2}$$

If we would set $V = 0$ in this equation, it would reduce to the Fokker-Planck equation, modeling a process of drift and diffusion, due to the terms with $b(x, t)$ and ν respectively. These terms conserve the total probability of the density. The extra term with the potential V makes that probability is not conserved (since particles are annihilated at a rate $V(x, t)dt$).

1.1 Motivation from machine learning

Equation (2) has a number of relations with the subject of machine learning. For instance, the process (2) can be viewed as a continuous time limit of the conventional discrete time non-linear dynamical systems with Gaussian noise (see e.g. [1]), i.e., it is a continuous time version of the

extended Kalman filter in which drift and diffusion can be thought of as predictor, and the annihilation potential $V(x, t)$ as corrector. With prior $\tilde{\rho}(x_0, t_0)$ and corrections up to time t_f the smoothed estimate of the state distribution at time $t_0 < t \leq t_f$ is given by

$$\tilde{\rho}(x, t) = \frac{1}{Z(V)} \int \rho(y, t_f | x, t) dy \int \rho(x, t | x_0, t_0) \tilde{\rho}(x_0, t_0) dx_0. \quad (3)$$

The process also has a central role in a recent stochastic optimal control theory [2, 3]. Under certain conditions, the optimal cost-to-go $J(x_0, t_0)$ in a control problem with horizon time t_f can be shown to be proportional to minus the log partition function of ρ , i.e.,

$$J(x_0, t_0) = -\lambda \log \int \rho(y, t_f | x, t) dy = -\lambda \log Z(x, t). \quad (4)$$

If the partition function – and hence J – is known, the control u simply follows by its gradient. The problem is that the computation of this partition function is intractable, and therefore approximations are needed.

2 The Gaussian approximation

Our goal is to approximate the density of particles at given final time t_f . To do this, we will consider the density of particles that survive the process up to the final time. At each intermediate time, we will approximate this density by a Gaussian shaped density, which is described by its mode and width σ . The mode as a function of time is the trajectory from initial time to final time that has maximum probability. The width $\sigma(t)$ as a function of time describes a band of typical fluctuations around the optimal trajectory.

2.1 The path integral formulation

In order to find the optimal trajectory, we will write the density $\rho(y, t_f | x_0, t_0)$, obeying Eq. 2, as a path integral. In order to make this construct, we first consider the transition density from a state x to a state y with an infinitesimal time step Δt

$$\rho(y, t + \Delta t | x, t) \propto \exp \left(- \left[\frac{(y - x - b(x, t)\Delta t)^2}{2\nu\Delta t} + V(x, t)\Delta t \right] \right). \quad (5)$$

This could be seen as a factor in an extended Kalman filter that is the combination of a predictor term modelled by the stochastic state transition $y = x + b\Delta t + d\xi$ and a corrector term modelled by $V(x, t)\Delta t$. The transition probability for a finite time step is the product of n of these infinitesimal steps, $t_f = t_n = t_0 + n\Delta t$, $y = x_n$,

$$\rho(x_n, t_n | x_0, t_0) \propto \int \prod_{i=1}^{n-1} dx_i \exp \left(-\Delta t \sum_{i=0}^{n-1} \left[\frac{1}{2} \left(\frac{x_{i+1} - x_i}{\Delta t} - b(x_i, t_i) \right)^2 + V(x_{i+1}, t_{i+1}) \right] \right). \quad (6)$$

In the limit of $\Delta t \rightarrow 0$, the sum in the exponent becomes an integral: $\Delta t \sum_{i=0}^{n-1} \rightarrow \int_t^{t_f} d\tau$, and the repeated integral at discrete times $\int \prod_{i=1}^{n-1} dx_i$ becomes an integral over paths that start at x and end at y , denoted as $\int [dx]_x^y$ (or simply $\int [dx]$ later in the paper). Thus we write

$$\rho(y, t_f | x_0, t_0) = \int [dx]_x^y \exp(-S[x]) \quad (7)$$

$$S[x] \equiv S(x(t_0) \rightarrow t_f) = \int_{t_0}^{t_f} \left(\frac{(\dot{x}(\tau) - b(x(\tau), \tau))^2}{2\nu} + V(x(\tau), \tau) \right) d\tau \quad (8)$$

with $x(t_0 \rightarrow t_f)$ a path with $x(\tau = t_0) = x_0$, $x(\tau = t_f) = y$, and $\dot{x}(\tau) = \frac{dx(\tau)}{d\tau}$. The functional $S[x]$ is called the action associated with the path x .

2.2 Euler-Lagrange equations

In this subsection, we aim to find the mode of the process. This is the path $x(t_0 \rightarrow t_f)$, starting at given x_0 and ending at arbitrary y , that minimizes the action S . This optimal path x_{opt} can be found by variational calculus. We write the action in a slightly more general (and more conventional) way

$$S[x] = \int_{t_0}^{t_f} L(t, x, \dot{x}) dt . \quad (9)$$

in which L is known as the Lagrangian. Let $x(t)$ be the minimizing path that we need to determine, and let $\epsilon z(t)$ be an independent fluctuation of the path at time t . Fluctuations are allowed at $\tau = t_f$. We start at $x(t_0) = x_0$, therefore $z(t_0) = 0$. For small ϵ we can Taylor expand $S[x + \epsilon z]$, which yields

$$S[x + \epsilon z] = S[x] + \epsilon \int_{t_0}^{t_f} z \partial_x L(t, x, \dot{x}) + \dot{z} \partial_{\dot{x}} L(t, x, \dot{x}) dt . \quad (10)$$

Since x is an extremum, the first order Taylor term must vanish, i.e.

$$0 = \int_{t_0}^{t_f} z \partial_x L(t, x, \dot{x}) + \dot{z} \partial_{\dot{x}} L(t, x, \dot{x}) dt \quad (11)$$

$$= \int_{t_0}^{t_f} z \left[\partial_x L - \frac{d}{dt} \partial_{\dot{x}} L \right] dt + z(t_f) \partial_{\dot{x}} L|_{t_f} . \quad (12)$$

This must hold for any z , which implies the well-known Euler-Lagrange equations with end-condition,

$$\frac{d}{dt} \partial_{\dot{x}} L = \partial_x L \quad (13)$$

$$\partial_{\dot{x}} L|_{t_f} = 0 . \quad (14)$$

It is convenient to define the momentum $p(t) = \partial_{\dot{x}} L(t, x, \dot{x})$. Since

$$L(t, x, \dot{x}) = \frac{(\dot{x} - b(x, t))^2}{2\nu} + V(x, t) , \quad (15)$$

we have $p(t) = \nu^{-1}(\dot{x} - b(x, t))$. The optimal path is described by the coupled ordinary differential equations,

$$\dot{x}(t) = b(x, t) + \nu p(t) \quad (16)$$

$$\dot{p}(t) = \partial_x V, \quad (17)$$

with a begin condition for x (as in the problem definition) and an end condition for p (which followed from the variational computation),

$$x(t_0) = x_0 \quad (18)$$

$$p(t_f) = 0 . \quad (19)$$

Some remarks

- To compute the optimal path, these equations have to be solved. In general a closed form solution will not be feasible, and numerical methods should be used. Solving this ordinary differential equation numerically, however, will be much easier than finding numerical solutions of the original partial differential equation (2) describing the evolution ρ .
- The solution $x_{\text{opt}}(t)$ can conceptually be found by the following forward backward algorithm. We consider all possible initial momenta p_0 . Then we prepare the system in $x(t_0) = x_0$ and $p(t_0) = p_0$. Then we integrate forwards in time. This yield a set of (not necessarily optimal) paths parametrized by p at time t . Each path ends at a final state x_f and final momentum p_f . Then we select the optimal path, which is the one with $p_f = 0$, and we follow this particular path backward in time, from which we find $x(t)$ for all intermediate times. This is conceptually similar to the forward-backward methods that are used in discrete-time systems, such as the Viterbi algorithm for hidden Markov models [4, 1].

- By varying the initial condition x_0 , we obtain a family of optimal paths, which can be parametrized as $x_{\text{opt}}(t + s|x, s)$ by points x at given time $t_0 \leq s \leq t_f$. The top row of figure 1 in the next section illustrates this notion. The time derivatives of these paths

$$\beta(x, t) = \left. \frac{d}{ds} x_{\text{opt}}(s + t|x, t) \right|_{s=0}, \quad (20)$$

where in fact $\beta = b + \nu p$, define a velocity field in the state-space.

- Note that the contribution of the momentum p to the velocity is proportional to the noise ν . This makes sense, since fluctuations are actually the cause of the deviation of the (surviving) particles from its original drift b . Thanks to the fluctuations the surviving particles avoided from running into regions of high annihilation rate and escaped to regions with lower annihilation rate.
- Another remark concerns the positive sign in front of V in Eq. 15, and hence in (17). This sign is opposite to the sign in front of the potential in the Lagrangian describing a classical mechanical system. Hence the behavior of the system is completely different. In classical mechanics, one usually starts at an initial point with zero momentum. Then the momentum increases in the direction of the negative gradient of the potential (i.e. down-hill). The result is that at later times the particle is down-hill, but with larger momentum. In the system described in this paper the particle aims to be down-hill with low momentum. Therefore it starts at the initial point with a large momentum in the down-hill direction. Then the momentum changes in the direction of the positive gradient (i.e up-hill), so that the particle effectively slows-down and ends down-hill, with zero momentum.

2.3 Linear noise approximation

Now that we have the optimal path, $x_{\text{opt}}(t)$, we need to estimate the size $\sigma(t)$ of the fluctuations around this path. Here we give arguments leading to a linear noise approximation a la van Kampen [5]. To approximate the size of the fluctuations around a path, we consider a particle at time t at position x (which defines an optimal path through this point at this time). At time t there is a probability distribution of future annihilation, which we express by an effective potential,

$$P_{\text{future annihilation}}(x) \propto \exp(-V^{\text{eff}}(x)). \quad (21)$$

Now consider the transition probability $P(y, t + dt|x, t)$, taking the effective potential into account. Typical jumps $\xi = y - x$ are small, so we do a linear expansion of the effective potential around x ,

$$V^{\text{eff}}(x + \xi) \approx V_0^{\text{eff}} + \xi V_1^{\text{eff}}. \quad (22)$$

The probability of a jump due to the drift/diffusion while taking the effective annihilation into account is then approximately

$$P(\xi, t) \propto \exp\left(-\left[\frac{(\xi - b(x, t)dt)^2}{2\nu dt} + \xi V_1^{\text{eff}}\right]\right). \quad (23)$$

This probability is maximized by a step in the direction of the optimal path, so we conclude that it must be of the form

$$P(\xi, t) \propto \exp\left(-\frac{(\xi - \beta(x, t)dt)^2}{2\nu dt}\right), \quad (24)$$

with β as in (20). So a Gaussian approximation of the the diffusion/annihilation process is described by an effective drift/diffusion process with diffusion ν and drift $\beta(x, t)$ following the optimal paths. For such a process, the size of the fluctuations obey

$$\partial_t \sigma^2(t) = 2 \partial_x \beta(x(t), t) \sigma^2(t) + \nu \quad (25)$$

the solution is stable if $\partial_x \beta(x(t), t) < 0$, which means that the flow lines of the paths are converging [5].

2.4 The partition function

The partition function is the fraction of particles that survives the process,

$$Z = \frac{1}{Z_0} \int [dx] \exp(-S[x]) \quad (26)$$

where $Z_0 = \int d[x] \exp(-S_0[x])$, the standard normalization of the drift/diffusion process without annihilation. With the Gaussian approximation, we can make an estimate of the partition function. In lowest order, the path integral is dominated by the optimal path x_{opt} , i.e.,

$$\int [dx] \exp(-S[x]) \approx \int [dx] \delta[x - x_{\text{opt}}] \exp(-S[x]) = \exp(-S[x_{\text{opt}}]) . \quad (27)$$

The resulting estimate of the partition function can be understood as the fraction of particles that survives a run along the path $x_{\text{opt}}(t_i \rightarrow t_f)$ with annihilation rate given by the Lagrangian, i.e.,

$$V^{\text{path}}(x, t) = \frac{(\beta(x, t) - b(x, t))^2}{2\nu} + V(x, t), \quad (28)$$

with β as in (20). In the next order, the path integral is dominated by the paths within the Gaussian around the optimal path,

$$\int [dx] \exp(-S[x]) \approx \int [dx] P_{\text{Gaussian}}[x|x_{\text{opt}}, \sigma^2] \exp(-S[x]) . \quad (29)$$

This estimate corresponds to the fraction of particles that survives runs fluctuating around the optimal path, with the same annihilation rate. This can be rewritten as a corrected path annihilation rate for particles running along the optimal path

$$V^{\text{path, corrected}}(x, t) = \left\langle \frac{(\beta(x, t) - b(x, t))^2}{2\nu} + V(x, t) \right\rangle_{[x_{\text{opt}}, \sigma^2]} . \quad (30)$$

By numerical integration of the annihilation rates (28) and/or (30) along the optimal path, estimates of the partition function can be computed.

3 Numerical example

We consider a system with $V(x) = \frac{1}{2}\alpha x^2$ and $b(x, t) = 0$. Then the Euler Lagrange equations lead to a linear system, which can be solved explicitly. Assuming that $x(t_f) = x_f$, the system can be solved backward in time, leading to $x(t) = \cosh((\alpha\nu)^{1/2}(t_f - t))x_f$. Eliminating x_f in terms of initial condition x_0 leads to the solution

$$x(t) = \frac{\cosh((\alpha\nu)^{1/2}(t_f - t))}{\cosh((\alpha\nu)^{1/2}(t_f))} x_0 \quad (31)$$

In figure 1, we plotted a number of optimal paths, starting at different initial states (for $\alpha\nu = 1$ (left panel) and $\alpha\nu = 5$ (right panel)). These paths may help to get some insight in the nature of the Euler-Lagrange equations, in particular the sign in front of the potential and the end-condition of zero momentum, discussed earlier in the paper.

In the lower row of the figure, we plotted paths generated from stochastic simulations of the diffusion/annihilation process (1) and the optimal path $\pm\sigma(t)$, where $\sigma(t)$ is computed according to the linear noise approximation (details omitted in this paper). From these graphs we see that the approximations provide a reasonable fit. Simulations with smaller time discretization give similar results (not shown here). The computation-time to compute the mode and noise estimates is an order of magnitude smaller than the time to do the stochastic sampling.

In figure 2, we plotted estimates of the partition function Z (i.e. fraction of surviving particles) based on the mode and with Gaussian corrections as function of the noise ν . All processes started at $x = 3$. Estimates are compared with results of stochastic simulations, each noise level starting with 100000 particles. The results show that in particular for large noises the Gaussian correction provides a significant improvement of the estimate.

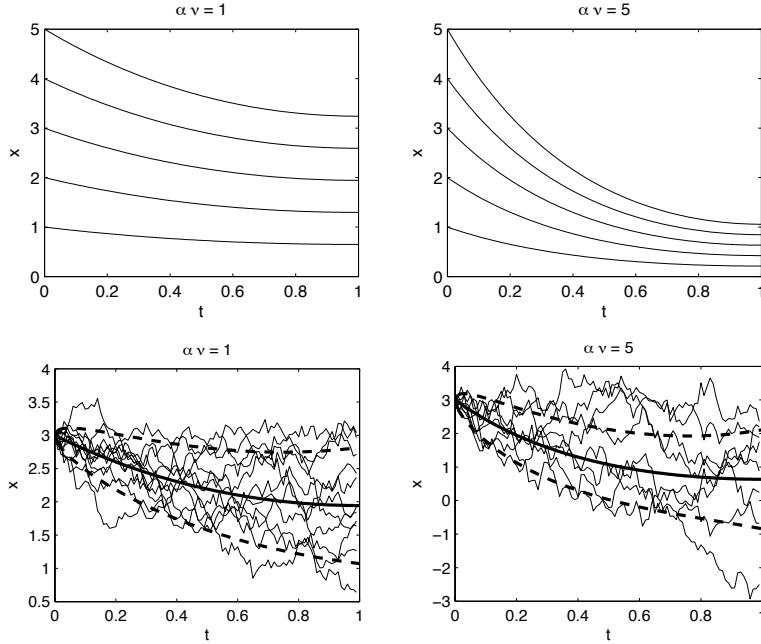


Figure 1: Top: optimal paths starting at different initial points x_0 with $\alpha\nu = 1$ (left) and $\alpha\nu = 5$ (right). Bottom: optimal paths (fat lines) starting at $x_0 = 3$, plus indications of estimated noise $\sigma(t)$ (fat dashed) and some random paths, sampled according to (1), with $\nu = 1$ (left) and $\nu = 5$ (right). $\alpha = 1$ in both cases. Note: The simulations with $\nu = 1$ started with 500 particles. The one with $\nu = 5$ started with 200 particles.

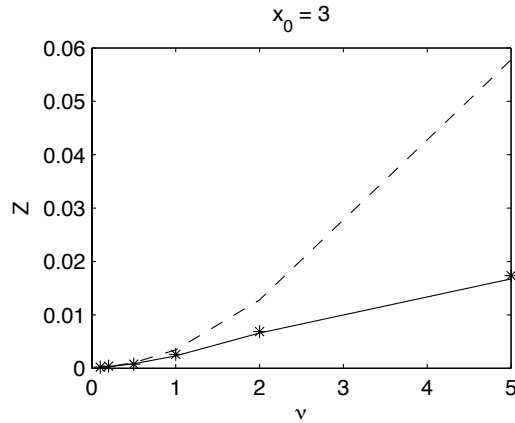


Figure 2: Estimate of the partition function (i.e. fraction of surviving particles) Z based on the mode (dashed) and with Gaussian corrections (drawn) as function of the noise ν . All processes started at $x = 3$. Estimates are compared with results of stochastic simulations, each starting with 100000 particles (stars).

4 Discussion

We considered a stochastic nonlinear dynamical process with annihilation of particles. This process can be viewed as the continuous time version of the extended Kalman filter/smoothers. It also plays an important role in stochastic optimal control theory.

Since the computation of the process, e.g. the computation of the smoothed estimate, is intractable, approximations are needed. In this paper we derived a Gaussian approximation. The mode is derived using variational calculus, leading to the well-known Euler-Lagrange equations. These are a set of ordinary differential equations that are to be solved under begin and end condition. Conceptually, they can be solved by a forward-backward algorithm. Furthermore we derived a linear noise approximation that provides estimates of the fluctuations around the path. The model and the Gaussian corrections were used to estimate the partition function, i.e. the fraction of particles that survive the process. For a simple model, the approximations have been compared with stochastically generated samples from the process and seem to agree very well. The Gaussian corrections provide a significantly better estimate of the partition function.

The applicability of the method to more realistic models is subject of future work. One of the model extensions that we want to study is to allow discrete time observations. In such a model, the process has annihilation probabilities in addition to annihilation rates. In the Euler-Lagrange formalism, this leads to jumps in the momentum, proportional to the gradient of the discrete-time potentials related to these annihilation probabilities. The resulting equations will be a chain of differential equations connected via these relations at the observation times.

Current applications of dynamic Bayesian models are mostly restricted to discrete time. An important reason for this restriction is the lack of efficient inference methods for continuous time models. In our paper, we aimed to make a contribution in the development of new approximate inference methods in order to make dynamic Bayesian models in continuous time practically feasible.

Acknowledgments

This research is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant BSIK03024.

References

- [1] C. Bishop. *Pattern recognition and machine learning*. Springer, 2006.
- [2] H. J. Kappen. Linear theory for control of nonlinear stochastic systems. *Physical Review Letters*, 95(20):200201, November 2005.
- [3] H. J. Kappen. Path integrals and symmetry breaking for optimal control theory. *Journal of statistical mechanics: theory and experiment*, page P11011, November 2005.
- [4] A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269, 1967.
- [5] N. van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam, 1992.