Open access • Journal Article • DOI:10.1103/PHYSREVLETT.95.200201

Linear theory for control of nonlinear stochastic systems. — Source link

Hilbert J. Kappen

Institutions: Radboud University Nijmegen

Published on: 07 Nov 2005 - Physical Review Letters (American Physical Society)

Topics: Stochastic control, Optimal control, Monte Carlo integration, Linear system and Nonlinear control

Linear Theory for Control of Nonlinear Stochastic Systems

Hilbert J. Kappen*

Department of Medical Physics & Biophysics, Radboud University, Geert Grooteplein 21 6525 EZ Nijmegen, The Netherlands†

(Received 12 November 2004; published 7 November 2005)

We address the role of noise and the issue of efficient computation in stochastic optimal control problems. We consider a class of nonlinear control problems that can be formulated as a path integral and where the noise plays the role of temperature. The path integral displays symmetry breaking and there exists a critical noise value that separates regimes where optimal control yields qualitatively different solutions. The path integral can be computed efficiently by Monte Carlo integration or by a Laplace approximation, and can therefore be used to solve high dimensional stochastic control problems.

DOI: 10.1103/PhysRevLett.95.200201 PACS numbers: 02.50.Ey, 02.30.Yy, 05.45.−a, 45.80.+r

Optimal control of nonlinear systems in the presence of noise is a very general problem that occurs in many areas of science and engineering. It underlies autonomous system behavior, such as the control of movement and planning of actions of animals and robots, but also, for instance, the optimization of financial investment policies and control of chemical plants. The problem is simply stated: given that the system is in this configuration at this time, what is the optimal course of action to reach a goal state at some future time? The cost of each time course of action consists typically of a path contribution that specifies the amount of work or other cost of the trajectory, and an end cost, that specifies to what extent the trajectory reaches the goal state.

In the absence of noise, the optimal control problem can be solved in two ways: using the Pontryagin minimum principle (PMP) [1], which is a pair of ordinary differential equations that are similar to the Hamilton equations of motion or using the Hamilton-Jacobi-Bellman (HJB) equation, which is a partial differential equation [2].

In the presence of (Wiener) noise, the PMP formalism is replaced by a set of stochastic differential equations which become difficult to solve (see, however, [3]). The inclusion of noise in the HJB framework is mathematically quite straightforward, yielding the so-called stochastic HJB equation [4]. Its solution, however, requires a discretization of space and time and the computation becomes intractable in both memory requirement and CPU time in high dimensions. As a result, deterministic control can be computed efficiently using the PMP approach, but stochastic control is intractable due to the curse of dimensionality.

For small noise, one expects that optimal stochastic control resembles optimal deterministic control, but for larger noise, the optimal stochastic control can be entirely different from the deterministic control [5], but there is currently no good understanding of how noise affects optimal control.

In this Letter, we address both the issue of efficient computation and the role of noise in stochastic optimal control. We consider a class of nonlinear stochastic control problems, which can be formulated as a statistical mechanics problem. This class of control problems includes arbitrary dynamical systems, but with a limited control mechanism. It contains linear-quadratic [4] control as a special case. We show that under certain conditions on the noise, the HJB equation can be written as a *linear* partial differential equation,

$$-\partial_t \psi = H\psi, \tag{1}$$

with H a (non-Hermitian) operator. Equation (1) must be solved subject to a boundary condition at the end time. As a result of the linearity of Eq. (1), the solution can be obtained in terms of a diffusion process evolving forward in time, and can be written as a path integral. The path integral has a direct interpretation as a free energy, where noise plays the role of temperature.

This link between stochastic optimal control and a free energy has two immediate consequences. (1) Phenomena that allow for a free energy description, typically display phase transitions. We argue that for stochastic optimal control one can identify a critical noise value that separates regimes where the optimal control is qualitatively different and illustrate this with a simple example. (2) Since the path integral appears in other branches of physics, such as statistical mechanics and quantum mechanics, we can borrow approximation methods from those fields to compute the optimal control approximately. We show how the Laplace approximation can be combined with Monte Carlo (MC) sampling to efficiently compute the optimal control.

Let $\vec{x}$ be an n-dimensional stochastic variable that is subject to the stochastic differential equation

$$d\vec{x} = (\vec{b}(\vec{x}, t) + \vec{u})dt + d\vec{\xi} \tag{2}$$

with $d\vec{\xi}$ a Wiener process with $\langle d\xi_i d\xi_j \rangle = \nu_{ij} dt$, and ν_{ij} independent of $\vec{x}, \vec{u}, t$. $\vec{b}(\vec{x}, t)$ is an arbitrary n-dimensional function of $\vec{x}$ and t, and $\vec{u}$ an n-dimensional vector of control variables. Given $\vec{x}$ at an initial time t, the stochastic optimal control problem is to find the control path $\vec{u}(\cdot)$ that minimizes

$$C(\vec{x},t,\vec{u}(\cdot)) = \left\langle \phi(\vec{x}(t_f)) + \int_t^{t_f} d\tau \left(\frac{1}{2}\vec{u}(\tau)^T \mathbf{R}\vec{u}(\tau) + V(\vec{x}(\tau),\tau) \right) \right\rangle_{\vec{x}} \quad (3)$$

with $\mathbf{R}$ a matrix, $V(\vec{x},t)$ a time-dependent potential, and $\phi(\vec{x})$ the end cost. The brackets $\langle\rangle_{\vec{x}}$ denote expectation value with respect to the stochastic trajectories (2) that start at $\vec{x}$.

One defines the *optimal cost-to-go function* from any time t and state $\vec{x}$ as

$$J(\vec{x},t) = \min_{\vec{u}(\cdot)} C(\vec{x},t,\vec{u}(\cdot)). \quad (4)$$

J satisfies the stochastic HJB equation which takes the form

$$-\partial_t J = \min_{\vec{u}} \left(\frac{1}{2}\vec{u}^T \mathbf{R}\vec{u} + V + (\vec{b}+\vec{u})^T \vec{\nabla} J + \frac{1}{2}\mathrm{Tr}(\boldsymbol{\nu}\vec{\nabla}^2 J) \right) = -\frac{1}{2}(\vec{\nabla}J)^T \mathbf{R}^{-1}\vec{\nabla}J + V + \vec{b}^T\vec{\nabla}J + \frac{1}{2}\mathrm{Tr}(\boldsymbol{\nu}\vec{\nabla}^2 J) \quad (5)$$

with $\mathrm{Tr}(\boldsymbol{\nu}\vec{\nabla}^2 J) = \sum_{ij}\nu_{ij}\partial^2 J/\partial x_i \partial x_j$ and

$$\vec{u} = -\mathbf{R}^{-1}\vec{\nabla}J(\vec{x},t) \quad (6)$$

the optimal control at $\vec{x}, t$. The HJB equation is nonlinear in J and must be solved with end boundary condition $J(\vec{x},t_f) = \phi(\vec{x})$.

Define $\psi(\vec{x},t)$ through [6]

$$J(\vec{x},t) = -\lambda \log \psi(\vec{x},t) \quad (7)$$

and assume there exists a scalar λ such that

$$\lambda \delta_{ij} = (\mathbf{R}\boldsymbol{\nu})_{ij}, \quad (8)$$

with δ_{ij} the Kronecker delta. In the one-dimensional case, such a λ can always be found. In the higher dimensional case, this restricts the matrices $\mathbf{R} \propto \boldsymbol{\nu}^{-1}$ [8]. Equation (8) reduces the dependence of optimal control on the n-dimensional noise matrix to a scalar value λ that will play the role of temperature. Equation (5) reduces to the linear Eq. (1) with

$$H = -\frac{V}{\lambda} + \vec{b}^T\vec{\nabla} + \frac{1}{2}\mathrm{Tr}(\boldsymbol{\nu}\vec{\nabla}^2). \quad (9)$$

Let $\rho(\vec{y},\tau|\vec{x},t)$ with $\rho(\vec{y},t|\vec{x},t) = \delta(\vec{y}-\vec{x})$ describe a diffusion process for $\tau > t$ defined by the Fokker-Planck equation

$$\partial_\tau \rho = H^\dagger \rho = -\frac{V}{\lambda}\rho - \vec{\nabla}^T(\vec{b}\rho) + \frac{1}{2}\mathrm{Tr}(\boldsymbol{\nu}\vec{\nabla}^2)\rho \quad (10)$$

with $H^\dagger$ the Hermitian conjugate of H. Then $A(\tau) = \int d\vec{y}\rho(\vec{y},\tau|\vec{x},t)\psi(\vec{y},\tau)$ is independent of τ and in particular $A(t) = A(t_f)$. It immediately follows that

$$\psi(\vec{x},t) = \int d\vec{y}\rho(\vec{y},t_f|\vec{x},t)\exp(-\phi(\vec{y})/\lambda). \quad (11)$$

We arrive at the important conclusion that $\psi(\vec{x},t)$ can be computed either by backward integration using Eq. (1) or by forward integration of a diffusion process given by Eq. (10).

We can write the integral in Eq. (11) as a path integral. We use the standard argument [9] and divide the time interval $t \to t_f$ in n_1 intervals and write $\rho(\vec{y},t_f|\vec{x},t) = \Pi_{i=1}^{n_1}\rho(\vec{x}_i,t_i|\vec{x}_{i-1},t_{i-1})$ and let $n_1 \to \infty$. The result is

$$\psi(\vec{x},t) = \int [d\vec{x}]_{\vec{x}} \exp\left(-\frac{1}{\lambda} S(\vec{x}(t\to t_f)) \right) \quad (12)$$

with $\int [d\vec{x}]_{\vec{x}}$ an integral over all paths $\vec{x}(t\to t_f)$ that start at $\vec{x}$ and with

$$S(\vec{x}(t\to t_f)) = \phi(\vec{x}(t_f)) + \int_t^{t_f} d\tau \left(\frac{1}{2}\left(\frac{d\vec{x}(\tau)}{d\tau} - \vec{b}(\vec{x}(\tau),\tau) \right)^T \times \mathbf{R}\left(\frac{d\vec{x}(\tau)}{d\tau} - \vec{b}(\vec{x}(\tau),\tau) \right) + V(\vec{x}(\tau),\tau) \right) \quad (13)$$

the action associated with a path. From Eqs. (7) and (12), the cost to go $J(x,t)$ becomes a log partition sum (i.e., a free energy) with temperature λ.

The path integral Eq. (12) can be estimated by stochastic integration from t to t_f of the diffusion process Eq. (10) in which particles get annihilated at a rate $V(\vec{x},t)/\lambda$:

$$\vec{x} = \vec{x} + \vec{b}(\vec{x},t)dt + d\vec{\xi}, \quad \text{with probability } 1 - Vdt/\lambda,$$
$$\vec{x} = \dagger, \quad \text{with probability } Vdt/\lambda, \quad (14)$$

where $\dagger$ denotes that the particle is taken out of the simulation. Denote the trajectories by $\vec{x}_\alpha(t\to t_f)$, $\alpha = 1,\ldots,N$. Then, $\psi(\vec{x},t)$ and $\vec{u}$ are estimated as

$$\hat{\psi}(\vec{x},t) = \sum_{\alpha\in\text{alive}} w_\alpha, \quad (15)$$

$$\hat{\vec{u}}dt = \frac{1}{\hat{\psi}(\vec{x},t)} \sum_{\alpha\in\text{alive}}^{N} w_\alpha d\vec{\xi}_\alpha(t),$$
$$w_\alpha = \frac{1}{N}\exp[-\phi(\vec{x}_\alpha(t_f))/\lambda], \quad (16)$$

where "alive" denotes the subset of trajectories that do not get destroyed along the way by the $\dagger$ operation. The normalization $1/N$ ensures that the annihilation process is properly taken into account. Equation (16) states that optimal control at time t is obtained by averaging the initial directions of the noise component of the trajectories $d\vec{\xi}_\alpha(t)$, weighted by their success at t_f.

The above sampling procedure can be quite inefficient, when many trajectories get annihilated. One of the simplest procedures to improve it is by importance sampling. We replace the diffusion process that yields $\rho(\vec{y},t_f|\vec{x},t)$ by

another diffusion process, that will yield $\rho'(\vec{y}, t_f|\vec{x}, t) = \exp(-S'/\lambda)$. Then Eq. (12) becomes,

$$\psi(\vec{x}, t) = \int [d\vec{x}]_{\vec{x}} \exp(-S'/\lambda) \exp(-(S - S')/\lambda).$$

The idea is to chose ρ' such as to make the sampling of the path integral as efficient as possible. Here, we use the Laplace approximation, which is given by the k deterministic trajectories $x_\beta(t \to t_f)$ that minimize the action

$$J(\vec{x}, t) \approx -\lambda \log \sum_{\beta=1}^{k} \exp[-S(\vec{x}_\beta(t \to t_f))/\lambda]. \qquad (17)$$

The Laplace approximation ignores all fluctuations around the modes and becomes exact in the limit $\lambda \to 0$. The Laplace approximation can be computed efficiently, requiring $\mathcal{O}(n^2 m^2)$ operations, where m is the number of time discretization.

For each Laplace trajectory, we define a diffusion processes ρ'_β according to Eq. (14) with $\vec{b}(\vec{x}, t) = \dot{\vec{x}}_\beta(t)$. The estimators for ψ and $\vec{u}$ are given again by Eqs. (15) and (16), but with weights

$$w_\alpha = \frac{1}{N} \exp(-[S(\vec{x}_\alpha(t \to t_f)) - S'_\beta(\vec{x}_\alpha(t \to t_f))]/\lambda). \qquad (18)$$

S is the original action equation (13) and S'_β is the new action for the Laplace guided diffusion. When there are multiple Laplace trajectories one should include all of these in the sample.

We give a simple one-dimensional example of a double slit to illustrate the effectiveness of the Laplace guided MC method and to show how the optimal cost to go undergoes symmetry breaking as a function of the noise.

Consider a stochastic particle that moves with constant velocity from $t = 0$ to $t_f = 2$ in the horizontal direction and where there is deflecting noise in the x direction:

$$dx = udt + d\xi.$$

The cost is given by Eq. (3) with $\phi(x) = \frac{1}{2}x^2$ and $V(x, t_1)$ implements a slit at an intermediate time $t_1 = 1$ (Fig. 1). Solving the cost to go by means of the forward computation using Eq. (11) can be done in closed form. The exact result, the Laplace approximation equation (17) and the Laplace guided importance sampling result using Eq. (18) are plotted for $t = 0$ as a function of x in Fig. 2. For each x, the Laplace approximation consists of the two deterministic trajectories, each being piecewise linear, starting at $t = 0$ in x and ending at $t = 2$ in $x = 0$. We see that the Laplace approximation is quite good for this example, in particular, when one takes into account that a constant shift in J does not affect the optimal control. The MC importance sampler has maximal error of order 0.1 and is significantly better than the Laplace approximation. Naive MC sampling using Eq. (14) (not shown) fails for this problem, because most trajectories get destroyed by the infinite potential. Numerical simulations using $N = 100\,000$ trajectories yield estimation errors in J up to approximately 6 for certain values of x.

We show an example how optimal stochastic control exhibits spontaneous symmetry breaking. For two slits of width ϵ at $x = \pm 1$, the cost to go becomes to lowest order in ϵ:

$$J(x, t) = \frac{R}{T}\left(\frac{1}{2}x^2 - \nu T \log 2 \cosh\frac{x}{\nu T}\right) + \text{const}, \qquad t < t_1,$$

where the constant diverges as $\mathcal{O}(\log\epsilon)$ independent of x and $T = t_1 - t$ the time to reach the slits. The expression between brackets is a typical free energy with inverse

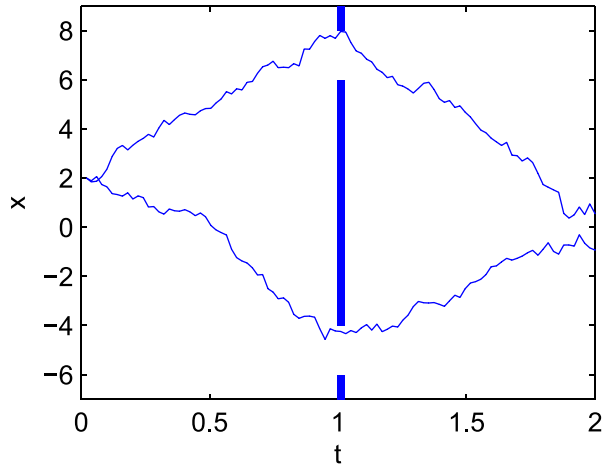

FIG. 1 (color online). A double slit is placed at $t = 1$ with openings at $-6 < x < -4$ and $6 < x < 8$. $V = \infty$ for $t = 1$ outside the openings, and zero otherwise. Also shown are two example trajectories under optimal control.

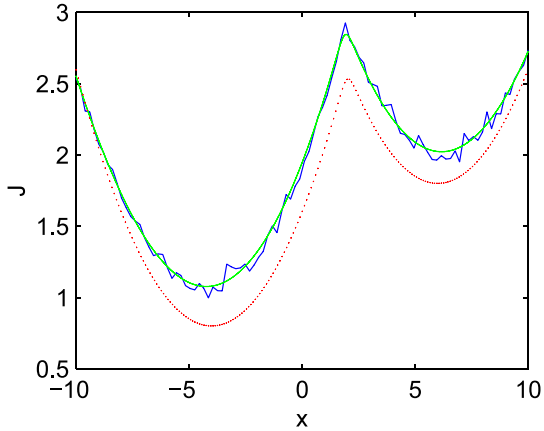

FIG. 2 (color online). Comparison of Laplace approximation (dotted line) and Monte Carlo importance sampling (solid jagged line) of $J(x, t = 0)$ with exact result (solid smooth line) for the double slit problem. The importance sampler used $N = 100$ trajectories for each x. $R = 0.1$, $\nu = 1$, $dt = 0.02$.

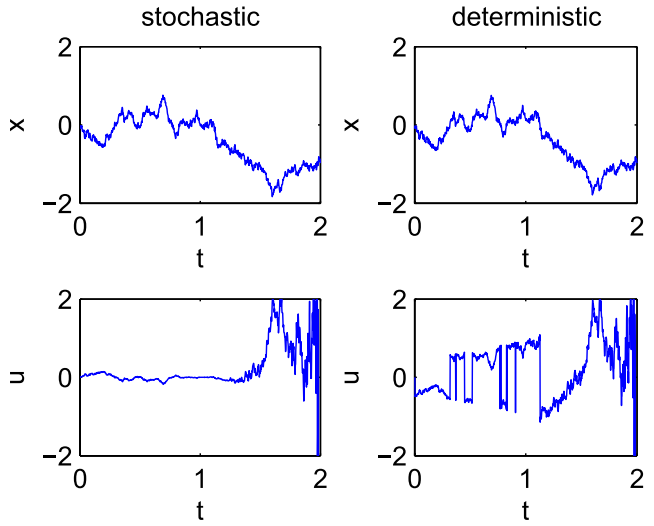

FIG. 3 (color online). Symmetry breaking in J as a function of T implies a ''delayed choice'' mechanism for optimal stochastic control. When the target is far in the future, the optimal policy is to steer between the targets. Only when $T < 1/\nu$ should one aim for one of the targets. Sample trajectories (top row) and controls (bottom row) under stochastic control (left column) and deterministic control (right column). $\nu = R = 1$, $t_1 = 2$.

temperature $\beta = 1/\nu T$. It displays a symmetry breaking at $\nu T = 1$. The optimal control is given by the gradient of J:

$$u = \frac{1}{T}\left(\tanh\frac{x}{\nu T} - x\right). \tag{19}$$

For $T > 1/\nu$ (far in the past) optimal control steers towards $x = 0$ (between the targets) and delays the choice of which slit to aim for until later. The reason why this is optimal is that the expected diffusion alone of size $\sqrt{\nu T}$ is likely to reach any of the slits without control (although it is not clear yet which slit). Only sufficiently late in time ($T < 1/\nu$) should one make a choice.

Figure 3 depicts two trajectories and their controls under stochastic optimal control [Eq. (19)] and deterministic optimal control [Eq. (19) with $\nu = 0$], using the same realization of the noise. Note, that at early times the deterministic control drives x away from zero, whereas in the stochastic control drives x towards zero and is smaller in size. The stochastic control delays the choice for which slit to aim until $T \approx 1$.

In summary, we have shown that stochastic optimal control involves symmetry breaking with qualitatively different solutions for high and low noise levels. This property is expected to be true also for more general stochastic control problems. The path integral formulation allows for an efficient solution of the HJB equation because it replaces the intractable n-dimensional numerical integration by a Monte Carlo sampling, which is known to be often much more efficient. This approach will thus be of direct practical value for the control of high dimensional, strongly nonlinear systems, such as, for instance, robot arms, navigation of autonomous systems, and chemical reactions. For realistic applications, naive sampling should be replaced by more advanced sampling schemes, such as importance sampling or a Metropolis method, and should be combined with efficient discretization such as splines, wavelets, or a Fourier basis [10,11].

I would like to thank Hans Maassen for useful discussions. This work is supported in part by the Dutch Technology Foundation and the BSIK/ICIS project.

*Email address: B.Kappen@science.ru.nl
†Electronic address: http://www.snn.kun.nl/~bert

[1] L. Pontryagin, V. Boltyanskii, R. Gamkrelidze, and E. Mishchenko, *The Mathematical Theory of Optimal Processes* (Interscience, New York, 1962).
[2] R. Bellman and R. Kalaba, *Selected Papers on Mathematical Trends in Control Theory* (Dover, New York, 1964).
[3] J. Yong and X. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations* (Springer, New York, 1999).
[4] R. Stengel, *Optimal Control and Estimation* (Dover, New York, 1993).
[5] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach* (Prentice Hall, Englewood Cliffs, NJ, 2003).
[6] The log transform goes back to Schrödinger and was first used in control theory by [7].
[7] W. Fleming, Applied Mathematics and Optimization **4**, 329 (1977).
[8] For example, if both $\mathbf{R}$ and $\boldsymbol{\nu}$ are diagonal matrices in a direction with low noise, control is expensive (R_{ii} large) and only small control steps are permitted. In noisy directions the reverse is true: control is cheap and large control values are permitted. As another example, consider a one-dimensional second order system subject to additive control $\ddot{x} = b(x, t) + u$. The stochastic formulation is of the form

$$dx = ydt, \qquad dy = (b(x, t) + u)dt + d\vec{\xi}.$$

Equation (8) states that due to the absence of a control term in the equation for dx, the noise in this equation should be zero.
[9] H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics* (World Scientific, Singapore, 1995), 2nd ed.
[10] W. Miller, J. Chem. Phys. **63**, 1166 (1975).
[11] D. Freeman and J. Doll, J. Chem. Phys. **80**, 5709 (1984).