Elements of Stochastic Optimal Control: ICDNS

MSci/MSc

Nick Jones

nick.jones@imperial.ac.uk

Aspects of Stochastic Optimal Control

In this lecture we will particularly briskly investigate selected topics in Stochastic Optimal Control. The treatment here is remarkably brief and selective.

We will have one more lecture on topics in Control in which we will investigate the bridge between topics in inference and control. This will use up a practical slot. We then turn to driving in our final week.

Tasks

Beyond asking you to check the more obvious statements in the handouts I have made the following requests. The objective here is to ensure you've actually followed up on the material and understand it. Beautiful expositions are not required - just demonstrations of understanding.

1. I Research how rejection sampling performs for higher dimensional problems and how the number of rejections depends on Q, P^* and c. You'll find answers to this in MacKay. Output: This can be summarized in a page or less.
2. I It also helps understand the following two papers: Libby et al [4] and Kobayashi et al [5] which you are expected to read (but you are by no means expected to understand these fully). Output: read them.
3. I These reviews are reasonably easy read and also combine to give an introduction to Bayesian cognitive science more generally: a) Probabilistic brains: knowns and unknowns. Alexandre Pouget, Jeffrey Beck, Wei Ji Ma and Peter Latham, Nature Neuroscience, 2013. b) Statistically optimal perception and learning: from behavior to neural representations. József Fiser, Pietro Berkes, Gergő Orbán and Máté Lengyel, Trends in Cognitive Sciences, 2010. Please read them (they are pretty interesting). Output: read them.
4. I Find and understand a brief proof of Landauer's principle (if you can't find one by the time we hit control theory ask me). Output: half a page or less.
5. I Gibbs sampling is called Glauber dynamics in the physics literature (go and have a very brief read about Glauber dynamics). Output: Just a few sentences.
6. C Read the introduction to Sontag [11] and, *in particular*, convince yourself of the role of PID control in stabilizing an inverted pendulum. Output: A page or less of explanation of this system.
7. C "For the linear system described to be controllable we require the matrix with columns $\tilde{\mathbf{A}}^i\mathbf{b}$ where $i = 0, .., n-1$ (and is an exponent not an index) with $x \in \mathbb{R}^n$ to be invertible. Look at the treatment of controllability by Zabczyk [9] (or in the other sources provided) and prove that this holds." Output: A precis of the proof in a page.
8. C Read about, and be able to explain, the Ott, Grebogi, and Yorke algorithm for stabilizing chaotic dynamics (you'll find an account of it in ref. [7]). Output: A precis of the method in a page or less.
9. C Please convince yourself of the proof of Bode's Integral Formula - you'll find it provided in [6]. Output: A precis of the proof in a page.
10. C Be able to derive the optimal control and cost-to-go of Linear Quadratic Gaussian control [1]. Output: A precis of the proof in half a page. Read [12] you do not need to understand fully. Output: Read it.

Noisy dynamics

We now consider Gaussian perturbations to our dynamical system.

$$dx = f(x, u, t)dt + d\xi$$

These can be distinct for different state variables and element of $d\xi$ can be correlated between them: $< d\xi_i d\xi_j >= \nu_{ij}(t, x, u)dt$ (though uncorrelated in time). Note that the control and noise need not be independent.

Expected cost and optimal cost-to-go

$$dx = f(x, u, t)dt + d\xi$$

$$C(x_0, u(0 \to T)) = \langle W(x_T) + \int_0^T dt\, w(x, u, t) \rangle$$

Any particular control $u(0 \to T)$ (with i.c. x_0) thus specifies an ensemble of trajectories with a corresponding distribution of costs from which we can calculate an expected cost.

$$J(t, x) = \min_u w(x, u, t)dt + \langle J(t + dt, x_{t+dt}) \rangle$$

Where the expectation occurs because, while we know our location and control precisely, we do not know precisely where this will take us. I'll discuss $u'(x, t)$ here. Even though we've precomputed $u'(x, t)$, it's useful when buffeted by (specifically unanticipated but generally modeled) noise.

Stochastic Hamilton Jacobi Bellman equation

$$J(t,x) = \min_u w(x,u,t)dt + \langle J(t+dt, x_{t+dt})\rangle$$

We can Taylor expand $J(t+dt, x_{t+dt})$. We have to go to second order in dx since $< dx^2 >= \mathcal{O}(dt)$ (basics of SDEs).

$$\langle J(t+dt, x_{t+dt})\rangle = J(x,t) + dt\partial_t J(t,x) + \langle dx\rangle \partial_x J(t,x) + \frac{1}{2}\langle dx^2\rangle \partial_x^2 J(t,x)$$

Noting that $< dx >= f(x,u,t)dt$ and $< dx^2 >= \nu(t,x,u)dt$ we obtain:

$$-\partial_t J(t,x) = \min_u (w(x,u,t) + f(x,u,t)\partial_x J(t,x) + \frac{1}{2}\nu(t,x,u)\partial_x^2 J(t,x))$$

This is the Stochastic HJB equation. We have picked up a diffusion-like term in our dynamics.

Linear Quadratic Gaussian Control

Linear dynamics, quadratic costs, Gaussian (white) noise.
Linear Dynamics with Gaussian noise:

$$dx = [Ax + Bu]dt + Fd\xi$$

Quadratic Cost:

$$w(x, u, t) = \frac{1}{2}u^T Ru + \frac{1}{2}x^T Qx$$

Final Cost:

$$W(x) = \frac{1}{2}x^T Q_T x$$

Where we can have A, B, F, R, Q time varying.

Linear Quadratic Gaussian Control II

One can show (using the stochastic HJB) that the optimal cost-to-go is of the form:

$$J(x,t) = \frac{1}{2}x^T V(t)x + a(t)$$

(with V symmetric) and the corresponding optimal control:

$$u'(x,t) = -R^{-1}B^T V(t)x$$

where $V(t)$ is obtained by solving:
$-\dot{V} = Q + A^T V + VA - VBR^{-1}B^T V$:
and $-\dot{a} = \frac{1}{2}Tr(FF^T V)$ with conditions $V(T) = Q_T$ and $a(T) = 0$.
Please prove the above - I've modeled my notation on Todorov [1] so the proof presented there should be particularly straightforward.

Linear Quadratic Gaussian Control III

Notably the optimal control is independent of F and noise effects can thus only modulate the total cost.

Dynamics: $dx = [Ax + Bu]dt + Fd\xi$

Cost: $w(x, u, t) = \frac{1}{2}u^T Ru + \frac{1}{2}x^T Qx$

Final Cost: $W(x) = \frac{1}{2}x^T Q_T x$

Optimal cost-to-go: $J(x, t) = \frac{1}{2}x^T V(t)x + a(t)$

Optimal controls: $u'(x, t) = -R^{-1}B^T V(t)x$

Dynamics of V: $-\dot{V} = Q + A^T V + VA - VBR^{-1}B^T V$

Dynamics of a: $-\dot{a} = \frac{1}{2}Tr(FF^T V)$

Non-exam extension

This is non examinable.

A more general class of system

Dynamics: $dx = (b(x,t) + Bu)dt + d\xi$
Uncontrolled dynamics: arbitrary. Noise: Gaussian and uncorrelated (assumed independent of u). Control: limited to be linear.
Cost: $w(x,u,t) = \frac{1}{2}u^T R u + Q(x,t)$
Costs are quadratic in the control but are arbitrary otherwise.
Final cost: $W(x) = \phi(x_T)$
Constraint: $\nu = \lambda B R^{-1} B$
The constraint has two notable features that we can observe in the case $B = I$ and R, ν diagonal. First we cannot have control in dimension i if $\nu_{ii} = 0$. Second if my passive dynamics/noise is large in dimension i (ν_{ii} large) then it's cheap for me to push this system around in direction i (R_{ii} small). This is natural for systems generically - we might think that it is easier to control a system to something close to where it might have gone anyway [12].

A more general class of system II

Dynamics: $dx = (b(x,t) + Bu)dt + d\xi$
Cost: $w(x,u,t) = \frac{1}{2}u^T Ru + Q(x,t)$
Final cost: $W(x) = \phi(x_T)$
Constraint: $\nu = \lambda BR^{-1}B$

We thus have a simple control (relatively simple noise) but a complex system we'd like to control with complex costs. We have also introduced a constraint which is strong but natural. I recommend reading Ref. [3] (it helps understand the non-examinable practical). It turns out one can make analytical progress with this system (in the examinable material we'll look at a discrete time variant of this system).

A more general class of system III

Dynamics: $dx = (b(x,t) + Bu)dt + d\xi$
Cost: $w(x,u,t) = \frac{1}{2}u^T Ru + Q(x,t)$
Final cost: $W(x) = \phi(x_T)$
Constraint: $\nu = \lambda BR^{-1}B$
We can thus write the stochastic HJB as:
$-\partial_t J(t,x) = \min_u(\frac{1}{2}u^T Ru + Q(x,t) + (b+Bu)^T \partial_x J(t,x) + \frac{1}{2}Tr(\nu(t,x,u)\partial_x^2 J(t,x))$.
Optimizing over u yields: $u'(x,t) = -R^{-1}B\partial_x J(t,x)$. Plugging this optimal control back in to the stochastic HJB, defining $J(x,t) = -\lambda \log \psi(x,t)$ and using the constraint discussed yields an equation linear in ψ: $\partial_t \psi = \left(\frac{V}{\lambda} - b^T \partial_x - \frac{1}{2}Tr(\nu(t,x,u)\partial_x^2)\right)\psi$.
This can be solved backwards in time starting with $\psi(x,T) = exp(-\phi(x)/\lambda)$ (since $J(x,T) = \phi(x_T)$).

Using diffusions to solve a large class of control problems

Without proof (though it is not complicated: see [3]) it turns out that $\partial_t \psi = \left(\frac{V}{\lambda} - b^T \partial_x - \frac{1}{2} Tr(\nu(t,x,u)\partial_x^2)\right) \psi$ can be solved by solving the diffusion process:
$\partial_t \rho = \left(-\frac{V}{\lambda} - \partial_x(b\rho) + \frac{1}{2}\sum_{ij} \nu_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \rho\right)$. We discover we can use solutions ρ of the preceding to find $J(x,t)$ using the form
$J(x,t) = -\lambda log \int dy \rho(y, T|x,t) exp(-\phi(y)/\lambda)$.
This is great: we can thus evolve ρ forward in time to solve for $J(x,t)$. In particular we can construct $J(x,t)$ by running Monte Carlo simulations consistent with the dynamics of ρ (initialized at x, t) and then taking their weighted sum at time T.

Using diffusions to solve a large class of control problems II

$$\partial_t \rho = \left(-\frac{V}{\lambda} - \partial_x(b\rho) + \frac{1}{2}\sum_{ij}\nu_{ij}\frac{\partial^2}{\partial x_i \partial x_j}\rho\right).$$

$$J(x,t) = -\lambda log \int dy \rho(y,T|x,t) exp(-\phi(y)/\lambda).$$

This is great: we can thus evolve ρ forward in time to solve for $J(x,t)$. In particular we can construct $J(x,t)$ by running Monte Carlo simulations consistent with the dynamics of ρ (initialized at x, t) and then taking their weighted sum at time T. We thus initialize a particle i at (x,t), record its location y_i at T and weight it by $\exp(-\phi(y_i)/\lambda)$. If a particle is absorbed (by the $V(x,t)$ field – I'll explain) we weight it to zero.

$$J(x,t) \simeq -\lambda \log \frac{1}{N}\sum_{i\in unabsorbed}^{N} \exp(-\phi(y_i)/\lambda).$$

$$\psi(x,t) = \frac{1}{N}\sum_{i\in unabsorbed}^{N} \exp(-\phi(y_i)/\lambda)$$

Using diffusions to solve a large class of control problems II

$\partial_t \rho = \left(-\frac{V}{\lambda} - \partial_x(b\rho) + \frac{1}{2}\sum_{ij} \nu_{ij} \frac{\partial^2}{\partial x_i \partial x_j}\rho\right)$.
$J(x,t) = -\lambda log \int dy \rho(y, T|x,t) exp(-\phi(y)/\lambda)$.
$J(x,t) \simeq -\lambda \log \frac{1}{N}\sum_{i \in unabsorbed}^{N} \exp(-\phi(y_i)/\lambda)$.
$\psi(x,t) \simeq \frac{1}{N}\sum_{i \in unabsorbed}^{N} \exp(-\phi(y_i)/\lambda)$
We can then deduce our optimal controls from our optimal cost-to-go:
$u'(x,t) = -R^{-1}B\partial_x J(t,x)$
In the case where the optimal control is unique we can approximate $u(x,t)$ directly through the form:
$u(x,t_j) \simeq \frac{1}{\psi(x,t)}\sum_{i \in unabsorbed}^{N} \exp(-\phi(y_i)/\lambda)\xi_j$ where we are considering the discrete time of our simulation and ξ_j is the perturbation at time j. This is particularly cute since it tells us that we can interpret our noise perturbations which are successful at steering the particle as controls. We are now set for the practical.

Remark

I've just taken us through what is sometimes called Path-Integral Control [3]. It relied on this constraint (inversely) connecting my the costs of control in a direction to whether the noise was strong in that direction (scaled by a parameter which can be interpreted as a temperature).

It turns out that this approach can be considered as a special case of a more general approach to control that looks at the Kullback-Leibler divergence between the probabilistic effect of a control $p(x_{t+1}|x_t, u_t)$ compared to uncontrolled dynamics $p(x_{t+1}|x_t, u_t = 0)$ [12] (take a look it's a great read). If the distribution associated with my control is close to that of my passive dynamics I consider this low cost. I thus pick up a term in my cost function which is the Kullback-Leibler divergence between $p(x_{t+1}|x_t, u_t)$ and $p(x_{t+1}|x_t, u_t = 0)$. There is an energetic interpretation to this since the amount of work one can extract if $p(x_{t+1}|x_t, u_t)$ relaxes to $p(x_{t+1}|x_t, u_t = 0)$ is specified by the Kullback-Leibler divergence.

Summarizing the week

I'll give a summary of the ideas that we've investigated this week.

Bibliography

[1] E. Todorov, Optimal control theory, In Bayesian Brain: Probabilistic Approaches to Neural Coding, Doya K at al (eds), chap 12, 269, MIT Press 2006. Free online.

[2] H.J. Kappen, Optimal conrol theory and the linear Bellman Equation. Book chapter. Free online.

[3] H.J. Kappen, Path integrals and symmetry breaking for optimal control theory. Journal of statistical mechanics: theory and Experiment, 11011, 2005.

[4] E. Libby, T.J. Perkins, and P.S. Swain Noisy information processing through transcriptional regulation, PNAS 104, 7151, 2007.

[5] T.J Kobayashi and A. Kamimura,Dynamics of intracellular information decoding, Phys. Biol. 8, 055007, 2011.

[6] K.J. Åström and R.J. Murray, Feedback systems, Princeton University Press. 2008. Free online.

[7] J. Bechhoefer, Feedback for physicists: A tutorial essay on control, Reviews of Modern Physics, 77, 783, 2005. Free to you online.

[8] R.S. Sutton and A.G. Barto, Reinforcement learning, MIT Press, 1998. Free online.

[9] J. Zabczyk, Classical control theory, 2001. Lecture notes online.

[10] J. Zabczyk, Mathematical control theory, Birkauser, 1995. Free to IC students.

[11] E. Sontag, Mathematical control theory, Springer, 1998. Free online.

[12] E. Todorov, Efficient computation of optimal actions, PNAS, 11478, 83, 2009.