EN530.603 Applied Optimal Control
Lecture 11: Stochastic Control
November 24, 2014

Lecturer: Marin Kobilarov

Consider a control system with dynamics given by

$$\dot{x}(t) = f(x(t), u(t), w(t), t),$$

where $w(t)$ is a random variable encoding uncertainty in the model. Since the state x is a random variable it is common to consider the minimization of an expected cost defined according to

$$J(u(\cdot)) = \mathbb{E}\left[\phi(x(t_f), t_f) + \int_{t_0}^{t_f} \mathcal{L}(x(\tau), u(\tau), \tau) d\tau\right]$$

where the expectation $\mathbb{E}[\cdot]$ is taken with respect to the evolution of the whole state trajectory $x(t)$. We will make the standard assumption that the noise w is uncorrelated in time and is such that

$$E[w(t)] = 0, \qquad E[w(t)w(\tau)^T] = W(t)\delta(t-\tau),$$

for a given positive semidefinite symmetric covariance $W(t)$.

1 Stochastic Control with Perfect Measurements

We first consider the case when the state x can be fully observed in the future and develop optimal control methods which only account for the uncertainty in system dynamics. First, consider the nonlinear system with additive noise given by

$$\dot{x}(t) = f(x(t), u(t)) + L(t)w(t), \tag{1}$$

where $L(t)$ is a given matrix.

Unlike the deterministic setting, when the state is a random variable one cannot directly use the variational optimality conditions since the adjoint equation $\dot{\lambda} = -\partial_x H$ cannot describe all possible random evolutions of $x(t)$. We must resort to dynamic programming.

We next consider dynamic programming for stochastic systems through the derivation of the stochastic Hamilton-Jacobi-Bellman equation. The *value function* $V(x,t)$ computed over the time interval $[t, t_f]$ is defined by

$$V(x(t), t) = \min_{u(t), t\in[t,t_f]} \mathbb{E}\left[\phi(x(t_f), t_f) + \int_{t}^{t_f} \mathcal{L}(x(\tau), u(\tau), \tau) d\tau\right]$$

As noted earlier, Bellman's principle of optimality states that if a trajectory over $[t_0, t_f]$ is optimal then it is also optimal on any subinterval $[t, t+\Delta t] \subset [t_0, t_f]$.

This can be expressed more formally through the recursive relationship

$$V(x(t),t) = \min_{u(t),t\in[t,t+\Delta t]} \mathbb{E}\left[\int_t^{t+\Delta t} \mathcal{L}\left(x(\tau),u(\tau),\tau\right)\mathrm{d}\tau + V(x(t+\Delta t),t+\Delta t)\right], \tag{2}$$

where the optimization is over the continuous control signal $u(t)$ over the interval $[t, t+\Delta t]$.

We will now show that optimal trajectories must satisfy the stochastic *Hamilton-Jacobi-Bellman equations* (HJB) given by:

$$-\partial_t V(x,t) = \min_{u(t)} \mathbb{E}\left\{\mathcal{L}(x,u,t) + \nabla_x V(x,t)^T f(x,u,w,t) + \frac{1}{2}\mathrm{tr}\left[\nabla_x^2 V(x,t) L(t) W(t) L(t)^T\right]\right\}, \tag{3}$$

$$= \min_{u(t)} \left\{\mathcal{L}(x,u,t) + \nabla_x V(x,t)^T f(x,u) + \frac{1}{2}\mathrm{tr}\left[\nabla_x^2 V(x,t) L(t) W(t) L(t)^T\right]\right\} \tag{4}$$

where $\mathrm{tr}[\cdot]$ is the matrix trace. To derive the principle we need to following relationship giving us an expression for the expected covariance over a finitely small time-interval Δt.

Discrete-time covariances. Recall that over a small interval $[t, t+\Delta t]$ we previously obtained that

$$E[L(t)w(t)w(\tau)^T L(\tau)^T] \approx \frac{1}{\Delta t} L(t) W(t) L(t)^T \delta_{jk}, \tag{5}$$

where the approximation sign means that the relationship holds as $\Delta t \to 0$. Here, δ_{jk} is the Kronecker delta (i.e. $\delta_{jk} = 1$ only when $j = k$, otherwise $\delta_{jk} = 0$). The indices j and k correspond to noise terms at times $t = t_0 + k\Delta t$ and $\tau = t_0 + j\Delta t$, respectively.

HJB Proof. To prove HJB, expand $V(x,t)$ to first order in Δt according to

$$\begin{aligned} V(x+\Delta x, t+\Delta t) &= V(x,t) + \partial_t V(x,t)\Delta t + \nabla_x V(x,t)^T \Delta x + \frac{1}{2}\Delta x^T \nabla_x^2 V(x,t)\Delta x + o(\Delta t) \\ &= V + \partial_t V \Delta t + \nabla_x V^T (f + Lw)\Delta t + \frac{1}{2}(f+Lw)^T \nabla_x^2 V (f+Lw)\Delta t^2 + o(\Delta t), \end{aligned}$$

where we suppressed the arguments for brevity. Note that in the above it was necessary expand V to second order in Δx because it is still not clear exactly what the order of w is. In fact, heuristically w will be of order $1/\sqrt{\Delta t}$ since its variance is of order $1/\Delta t$. That is why the terms above containing two multiples of w will be of order Δt rather than Δt^2. This is in fact the key difference between the chain rule in differentiation of stochastic and deterministic systems. Substituting the above in (2) and taking $\Delta t \to 0$ we have

$$V = \min_{u(t)} \mathbb{E}\left\{\mathcal{L}\Delta t + V + \partial_t V \Delta t + \nabla_x V^T [f + Lw]\Delta t + \frac{1}{2}(f+Lw)^T \nabla_x^2 V (f+Lw)\Delta t^2\right\}. \tag{6}$$

By assumption, we have that the expectations of all functions of $x(t)$ and $u(t)$ are themselves (that is $\mathcal{L}$, V, $\nabla_x V^T f$) because x can be measured exactly. We have that

$$\begin{aligned}
&\mathbb{E}\left[(f+Lw)^T\nabla_x^2 V(f+Lw)\right] \\
&= f^T\nabla_x^2 V f + 2f^T\nabla_x^2 V L\mathbb{E}[w] + \mathbb{E}\left[(Lw)^T\nabla_x^2 V(Lw)\right] \\
&= f^T\nabla_x^2 V f + \mathbb{E}\left[\mathrm{tr}(\nabla_x^2 V L w w^T L^T)\right] \\
&= f^T\nabla_x^2 V f + \mathrm{tr}(\nabla_x^2 V L\mathbb{E}\left[ww^T\right]L^T) \\
&= f^T\nabla_x^2 V f + \frac{1}{\Delta t}\mathrm{tr}(\nabla_x^2 V L W L^T),
\end{aligned}$$

where the last equality follows from (5) applied at $\tau = t$. The above is then substituted into (6) to obtain

$$0 = \min_{u(t)}\left\{\mathcal{L} + \partial_t V + \nabla_x V^T f + \frac{1}{2}\mathrm{tr}(\nabla_x^2 V L W L^T)\right\}\Delta t + o(\Delta t).$$

Since this must hold for any Δt then we obtain the HJB equation.

1.1 Continuous Linear-quadratic systems

Consider the linear system

$$\dot{x} = Fx + Gu + Lw$$

with $x(0) = x_0$ and given t_0 and t_f. The cost function is

$$J = \frac{1}{2}\mathbb{E}\left\{x^T(t_f)S_f x(t_f) + \int_{t_0}^{t_f}\begin{bmatrix} x(t)^T & u(t)^T\end{bmatrix}\begin{bmatrix} Q(t) & M(t)\\ M(t)^T & R(t)\end{bmatrix}\begin{bmatrix} x(t)\\ u(t)\end{bmatrix}dt\right\}$$

To solve the stochastic HJB equation we can consider a value function of the form

$$V(t) = \frac{1}{2}x^T(t)S(t)x(t) + v(t), \tag{7}$$

where $\frac{1}{2}x^T(t)S(t)x(t)$ is the *certainty-equivalent value function*, i.e. it is the same as in the deterministic setting. The term $v(t)$ is the *stochastic value function increment* defined by

$$v(t) = \frac{1}{2}\int_t^{t_f}\mathrm{tr}[S(\tau)LWL^T]d\tau$$

The stochastic HJB equation is then:

$$-\partial_t V = \min_u \frac{1}{2}\left\{E[x^TQx + 2x^TMu + u^TRu + x^TS(Fx+Gu)] + \mathrm{tr}(SLWL^T)\right\} \tag{8}$$

$$= \min_u \frac{1}{2}\left\{x^TQx + 2x^TMu + u^TRu + x^TS(Fx+Gu) + \mathrm{tr}(SLWL^T)\right\} \tag{9}$$

with terminal condition

$$V(t_f) = \frac{1}{2}x(t_f)^T S_f x(t_f).$$

Following our previous developments we can show that the minimizing control law is

$$u = -R^{-1}[G^TS + M^T]x \tag{10}$$

This can be substituted in (8) to get

$$-\partial_t V = \frac{1}{2}x^T \left\{(F - GR^{-1}M^T)^T S + S(F - GR^{-1}M^T) + Q - SGR^{-1}G^T S - MR^{-1}M^T\right\} x + \frac{1}{2}\mathrm{tr}(SLWL^T)$$

Using the assumed form for V given in (7) we have

$$\partial_t V = \frac{1}{2}x^T \dot{S} x + \dot{v},$$

and hence the matrix S must satisfy

$$-\dot{S} = (F - GR^{-1}M^T)^T S + S(F - GR^{-1}M^T) + Q - SGR^{-1}G^T S - MR^{-1}M^T$$

while the stochastic increment must satisfy

$$-\dot{v} = \frac{1}{2}\mathrm{tr}(SLWL^T)$$

in order for the HJB equation to hold.

1.2 Discrete-time linear-quadratic systems

Consider the discrete-time linear system

$$x_{k+1} = F_k x_k + G_k u_k + L_k w_k$$

with $x(0) = x_0$ and given t_0 and t_f. The cost function is

$$J_0 = \frac{1}{2}\mathbb{E}\left\{x_N^T S_N x_N + \sum_{k=0}^{N-1} \begin{bmatrix} x_k^T & u_k^T \end{bmatrix} \begin{bmatrix} Q & M \\ M^T & R \end{bmatrix} \begin{bmatrix} x_k \\ u_k \end{bmatrix}\right\}$$

The general discrete stochastic HJB is

$$V_k(x_k) = \min_{u_k} \mathbb{E}\left[\mathcal{L}_k(x_k, u_k) + V_{k+1}(x_{k+1})\right],$$

which for our problem is

$$V_k = \min_{u_k} \frac{1}{2}\mathbb{E}\left[x_k^T Q x_k + 2x_k^T M u_k + u_k^T R u_k + x_{k+1}^T S_{k+1} x_{k+1} + 2v_{k+1}\right], \tag{11}$$

To solve it assume a value function of the form

$$V_k = \frac{1}{2}x_k S_k x_k + v_k,$$

where $\frac{1}{2}x_k^T S_k x_k$ is the *certainty-equivalent* discrete value function, and v_k is the *stochastic value function increment* defined by

$$v_k = \frac{1}{2}\mathrm{tr}[S_{k+1}LW_kL^T] + v_{k+1}, \quad v_N = 0.$$

After substituting the above and the dynamics in (11) we have

$$V_k = \min_{u_k} \frac{1}{2}\mathbb{E}\ \left[x_k^T Q x_k + 2x_k^T M u_k + u_k^T R u_k + (Fx_k + Gu_k)^T S_{k+1}(Fx_k + Gu_k) + \mathrm{tr}(S_{k+1} L W_k L^T) + 2v_{k+1}\right], \tag{12}$$

and the minimizing control is found as

$$u = -(R + G^T S_{k+1} G)^{-1}(M^T + G^T S_{k+1} F)x_k$$

which is substituted back in (12) to obtain

$$V_k = \frac{1}{2}x_k^T[(F^T S_{k+1} F + Q) - (M^T + G^T S_{k+1} F)^T (R + G^T S_{k+1} G)^{-1}(M^T + G^T S_{k+1} F)]x_k + \frac{1}{2}\mathrm{tr}[S_{k+1} L W_k L^T] + v_{k+1},$$

and hence the HJB can be solved through the discrete Riccati equation

$$S_k = (F^T S_{k+1} F + Q) - (M^T + G^T S_{k+1} F)^T (R + G^T S_{k+1} G)^{-1}(M^T + G^T S_{k+1} F).$$

1.3 Neighboring-optimal control

For nonlinear systems there is no closed-form solution. A common approach is ignore the uncertainty and compute a deterministic solution using any of the numerical methods studied. The system dynamics can then be linearized around this solution and the second-order terms in the cost function are considered. The uncertainty is approximated using Gaussian functions. The resulting linear-quadratic-Gaussian problem is then solved using the preceding methods, i.e. using LQR.

The procedure is as follows:

1. Solve the deterministic optimal control problem for

   $$\dot{x}(t) = f(x(t), u(t), w_0(t), t),$$

   where $w_0(t)$ is a *given* nominal value for the noise, by minimizing the deterministic cost

   $$J(u(\cdot)) = \phi(x(t_f), t_f) + \int_{t_0}^{t_f} \mathcal{L}(x(\tau), u(\tau), \tau) d\tau.$$

   Let the solution be denoted by $u_0^*(t)$and the corresponding optimal trajectory by $x_0^*(t)$. Note, that by a given w_0 we mean that it could be just set to its expected value (i.e. 0) or it could be sampled using its covariance.

2. Construct a second-order variation of the cost function

   $$\Delta^2 J = \frac{1}{2}\Delta x^T(t_f)\phi_{xx}(t_f)\Delta x + \frac{1}{2}\int_{t_0}^{t_f} \begin{bmatrix} \Delta x_k^T & \Delta u_k^T \end{bmatrix} \begin{bmatrix} Q & M \\ M^T & R \end{bmatrix} \begin{bmatrix} \Delta x_k \\ \Delta u_k \end{bmatrix} dt, \tag{13}$$

   where

   $$\Delta x(t) = x(t) - x_0^*(t), \qquad \Delta u(t) = u(t) - u_0^*(t), \qquad \Delta w(t) = w(t) - w_0(t)$$

   and

   $$Q(t) = \mathcal{L}_{xx}(x_0^*(t), u_0^*(t)), \quad R(t) = \mathcal{L}_{uu}(x_0^*(t), u_0^*(t)), \quad M(t) = \mathcal{L}_{xu}(x_0^*(t), u_0^*(t)),$$

and specify the dynamic constraints

$$\dot{\Delta x} = F(t)\Delta x + G(t)\Delta u + L(t)\Delta w, \tag{14}$$

where

$$F(t) = f_x(x_0^*(t), u_0^*(t), w_0(t), t), \quad G(t) = f_u(x_0^*(t), u_0^*(t), w_0(t), t), \quad L(t) = f_w(x_0^*(t), u_0^*(t), w_0(t), t),$$

and minimize (13) subject to (14) (i.e. solve an LQR problem) to obtain the *perturbation feedback control* term $\Delta u^*(t)$ given by

$$\Delta u^*(t) = -C(t)\Delta x^*(t),$$

where $C(t) = R(t)^{-1}(G(t)^T S(t) + M(t)^T)$ is the gain matrix of the LQR problem (see (10)).

3. The actual control law combines the nominal control and the perturbation control, i.e.

$$u^*(t) = u_0^*(t) + \Delta u^*(t),$$

and since $\Delta x^*(t)$ in practice is computed using the measured state $x_M(t)$ the control is used according to

$$u^*(t) = u_0^*(t) - C(t)[x_M(t) - x_0^*(t)].$$

Analogous procedure applies in the discrete dynamics case for which that optimal control would take the form

$$u^*(t_k) = u_o^*(t_k) - C_k[x_M(t_k) - x_0^*(t_k)],$$

where the discrete-time gain matrix is

$$C_k = (R_k + G_k^T S_{k+1} G_k)^{-1}(M_k^T + G_k^T S_{k+1} F_k).$$

Note that when the controller is used in real-time on the real system Δw in (14) is provided by the system. But the controller can also be used for design/testing purposes using a simulated model, in which case Δw would be generated using a random sampling process.

The procedure can be performed multiple times for different sampled trajectories $w_0(t)$ so that the state space of interest is seeded with possible future evolutions of the dynamics around which we have constructed neighboring-optimal feedback controls.

Link to backward-forward sweep methods. Recall that the sweep methods such as stage-wise Newton or differential dynamic programming (DDP) resulted in a similar control law combining a nominal and perturbation parts, i.e.

$$u_i = k_i + K_i \Delta x_i.$$

Interestingly, sweep methods directly extends to neighboring-optimal stochastic control since the optimized perturbation part can be regarded as the optimal feedback controller analogous to the term $\Delta u^*(t)$. In this sense, DDP for instance simultaneously solves the deterministic optimization of a nominal trajectory as well as the neighboring optimal controller. This is not surprising, since DDP formulation was based on a quadratic model expansion which includes both first-order terms (related to necessary conditions for optimality) and second-order terms (related to sufficient conditions but capturing the $\Delta^2 J$ terms used above, which provides the feedback control terms as a byproduct of the numerical optimization.)

2 Stochastic Control with Uncertain Measurements

The most general case is when both the dynamics and measurements contain uncertainty. Linear-Gaussian systems satisfy the *certainty-equivalence principle*, i.e. the optimal control is the deterministic optimal feedback control $u = K\hat{x}$ assuming zero noise (i.e. LQR) where the feedback state $\hat{x}$ is estimated using optimal linear estimator (i.e. using the Kalman filter).

In the nonlinear case there are no closed form solutions. The common approach is to compute a locally optimal trajectories ignoring uncertainty, i.e. using any of the numerical optimization techniques. The system dynamics and measurement models can then be linearized around this reference trajectory and linear-Gaussian methods are applied. A systematic application of this approach entails generating multiple optimal trajectories i.e. by sampling from the dynamics, which approximately cover the region in state space where the system is expected to operate.

One strategy is to formulate the problem as the optimization of three competing terms, i.e.

$$J = J_D + J_C + J_P,$$

where J_D is the cost of the *deterministic* control which ignores uncertainty, J_P denotes *cautious* feedback control which essentially accounts for the increase in uncertainty from noise in the dynamics, and J_C denotes *probing* feedback which captures the reduction in uncertainty from processing measurements. Recall that (as in the EKF) the uncertainty grows after each dynamic update so that J_C attempts to minimize this growth, while J_P encourages the collection of measurements which improve the state estimate. More specifically, assuming that the uncertainty is propagated (approximately) using an EKF these costs are given by

$$J_C(k) = \frac{1}{2}\text{tr}[S_{k+1}P_{k+1|k}] + \frac{1}{2}\sum_{j=k+1}^{N-1}\text{tr}[S_{j+1}LW_jL^T],$$

where $P_{k+1|k}$ is the predicted covariance before a measurement is processed, and by

$$J_P(k) = \frac{1}{2}\sum_{j=k+1}^{N-1}\text{tr}[\mathcal{C}_jP_{j|j}],$$

where $P_{j|j}$ is the corrected covariance after a measurement update, and $\mathcal{C}_j$ is weighing matrix which relates the effect the controls u to estimated states $\hat{x}$. For example, if DDP was used to solve the deterministic control J_D then we could set $\mathcal{C}_k = Q_{ux}^TQ_{uu}^{-1}Q_{ux}$. Note that these choices are heuristic and still the required solution remains computationally challenging.

2.1 Neighboring-optimal control

In case when only small deviations from a deterministic solutions are expected then one can apply a neighboring-optimal approach to the full stochastic control problem. The key difference from the case with perfect measurements would be that the actual feedback state would have to be estimated using an estimator that is optimal in the vicinity of the reference trajectory such as the EKF.

In particular, the controller would take the form

$$u_k = u_k^* - C_k[\hat{x}_k - x_k^*],$$

where the estimated state $\hat{x}_k$ is computed using the EKF based on a linearization around x_k^*.

References