Stochastic Optimal Feedback Control Theory

UCL Max Planck Journal Club

Spencer Wilson

Sainsbury Wellcome Centre
spencer.wilson@ucl.ac.uk

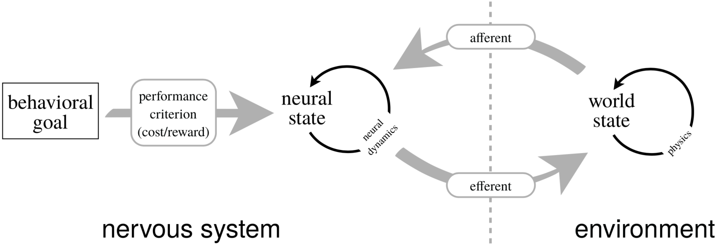
January 18, 2020

Outline

The Motor Optimality Hypothesis I

Biological behavior maximizes performance with respect to internally measured goals.

Tassa, 2011



The Motor Optimality Hypothesis II

- How do we coordinate a large number of degrees of freedom to accomplish task goals? How does the variability in individual degrees of freedom combine to form an optimal mapping from states to controls throughout a task?
- Pre-Todorov, the prevailing strategy in the motor control literature was to use a serial model of planning+execution with open-loop trajectory-optimized control.
- Stochastic Optimal Feedback Control is an online, model-based control scheme to derive task-specific optimal controllers rather than task-specific state trajectories. This framework assumes the dynamics of the controlled system are noisy, and that we observe the system through noisy observations.

Open-Loop Trajectory Optimization

In the trajectory optimization framework, a criterion is based on trajectory smoothness (e.g. *jerk*):

$$C = \frac{1}{2}\int_0^T \left(\left(\frac{d^3x}{dt^3}\right)^2 + \left(\frac{d^3y}{dt^3}\right)^2 \right) dt$$

x and y are extracorporal Cartesian coordinates of the end effector. Using setting particulars, $x(t)$ and $y(t)$ of the trajectory are found via Euler-Poisson, Pontryagin, Lagrange... Note that trajectory error could be used in a feedback controller. (Flash & Hogan, 1985)
Note:

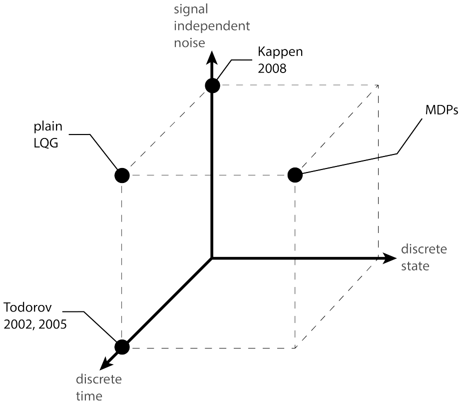
- Recovers average trajectory cost over trials
- Does not predict perturbation response
- Does not predict redundancy characteristics

Model Space

Optimal control modeling is a combination of dynamical systems and optimization.

Variables

- Discretization (Time & Space)
- Noise Model (Additive Gaussian)
- Dynamical Model (Linear)
- Cost Function (Quadratic)
- Control Constraints (Bounded)
- Horizon (Finite, Discount, Infinite)



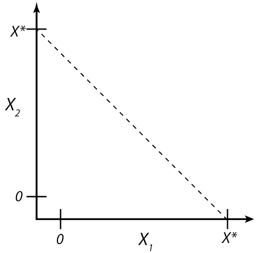
Goals

- Motivate the optimal feedback control framework
- Define **redundancy** and **synergy**
- Derive the feedback control law for the LQG case
- Discuss the signal-dependent noise extension for the LQG
- Discuss examples from Todorov, Jordan 2002
- Discuss pros and cons of the framework in general

One-Step Optimal Control

Dynamics for a single time step (start, finish):

$$x_i^{final} = ax_i + u_i(1 + \sigma\xi_i)$$
$$\xi_i \sim \mathcal{N}(0, 1)$$
$$i \in \{1, 2\}$$



The goal is to minimize the expected cost:

$$u_i^* = \underset{u}{\operatorname{argmin}} \mathbb{E}[(x_1^{final} + x_2^{final} - X^*)^2 + r(u_1^2 + u_2^2)]$$

Plugging in for the final states using the dynamics:

$$u_i^* = \underset{u}{\operatorname{argmin}} \mathbb{E}[((a(x_1 + x_2) - X^*) + (u_1 + u_2) + \sigma(u_1\xi_1 + u_2\xi_2))^2 + r(u_1^2 + u_2^2)]$$

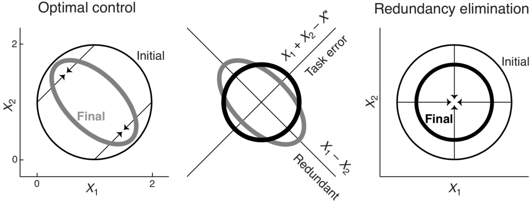
One-Step Optimal Control

Multiplying out, applying the expectation, and using $\mathbb{E}[\xi_i] = 0$ and $\mathbb{E}[\xi_i^2] = 1$:

$$u_i^* = \underset{u}{\operatorname{argmin}} \left[(a(x_1 + x_2) - X^* + (u_1 + u_2))^2 + (r + \sigma^2)(u_1^2 + u_2^2) \right]$$

Assuming that σ, r are small and knowing that $(u_1 + u_2)^2 > (u_1^2 + u_2^2)$: "the [unbiased] optimal controls u_i^* minimize $(r + \sigma^2)(u_1^2 + u_2^2)$ subject to $u_1 + u_2 = -Err \equiv -(a(x_1 + x_2) - X^*)$ where Err is the expected error when $u_1 = u_2 = 0$." Compare optimal controls to "desired-state controls":

$$u_i = -Err_{ds} = X^*/2 - ax_i$$



Outline

Setup

Dynamics

$$x_{t+1} = A_t x_t + B_t u_t$$

Cost to minimize

$$J_N(u) = \sum_{t=0}^{N-1} ||D_t x_t + F_t u_t||^2 + x_N^T Q x_N$$

Optimal Cost-to-go (Bellman Equation)

$$V_t(x) = \min_u \left[||D_t x + F_t u||^2 + V_{t+1}(A_t x + B_t u) \right]$$
$$V_N(x) = x^T Q x$$

Minimum Cost

$$V_0(x_0) = J_N(u^*)$$

Goal

- We want to find an optimal **control law** (an optimal policy) $u_t^* = u_t^*(x_t)$ under the scalar performance index $J(u)$ that maps states to controls. Note that we could run the dynamics forward, compute costs, and use gradient descent to find an open loop control trajectory u^* but this would run into the curse of dimensionality and not provide use with a feedback control solution in the face of perturbations.
- We will show that under linear dynamics with quadratic costs, our control law is linear and our value function is quadratic (ansatz method)

$$V_t(x) = x^T S_t x$$
$$u_t^*(x) = -L_t x$$

- Notes
  - We will assume finite horizon in all problems
  - We will assume time invariance ($A(t) \to A$)

Derivation I

At $t = N$, $S(N) = Q$. We will prove that for $t < N$, $V_{t+1}(x) = x^T S_{t+1} x$.

$$V_t(x) = \min_u \left[||Dx + Fu||^2 + V_{t+1}(Ax + Bu) \right]$$

$$V_t(x) = \min_u [x^T D^T Dx + u^T F^T Fu + 2x^T D^T Fu$$
$$+ (x^T A^T + u^T B^T) S_{t+1}(Ax + Bu)]$$

$$V_t(x) = \min_u [x^T D^T Dx + u^T F^T Fu + 2x^T D^T Fu$$
$$+ x^T A^T S_{t+1} Ax + u^T B^T S_{t+1} Bu + 2x^T A^T S_{t+1} Bu]$$

We want to have S_j in terms of S_{j+1} without u.

$$V_t(x) = \min_u [x^T (D^T D + A^T S_{t+1} A) x$$
$$+ u^T (F^T F + B^T S_{t+1} B) u$$
$$+ 2x^T (D^T F + A^T S_{t+1} B) u]$$

Derivation II

Let's rename some terms

$$\begin{aligned} V_t(x) &= x^T W x + u^T Y u + 2x^T Z u \\ W &= D^T D + A^T S_{t+1} A \\ Y &= F^T F + B^T S_{t+1} B \\ Z &= D^T F + A^T S_{t+1} B \end{aligned}$$

(Lemma) Let R be symmetric positive definite, then

$$\begin{aligned} (u+a)^T R(u+a) &= u^T R u + a^T R a + 2a^T R u \\ u^T R u + 2a^T R u &= (u+a)^T R(u+a) - a^T R a \end{aligned}$$

Derivation III

To minimize the second expression w.r.t. u, we set $u = -a$ with minimum $-a^T Ra$. Identifying $R = Y$ and $Ra = Z^T x$, then

$$a = R^{-1} Z^T x = Y^{-1} Z^T x$$

Now we can drop the *min* and solve for $V_t(x)$ in terms of S and constants.

$$\begin{aligned}
V_t(x) &= x^T W x - 2x^T Z a + a^T Y a \\
&= x^T W x - 2x^T Z Y^{-1} Z^T x + x^T Z Y^{-1} Y Y^{-1} Z^T x \\
&= x^T W x - x^T Z Y^{-1} Z^T x \\
&= x^T (W - Z Y^{-1} Z^T) x \\
&= x^T S_t x \quad \square
\end{aligned}$$

Derivation IV

Thus, we find a recursive solution for S, the Discrete-time Algebraic Ricatti Equation. S_t is found using a backwards pass from $S_N = Q$.

$$\begin{aligned} S_t &= W - ZY^{-1}Z^T \\ &= D^T D + A^T S_{t+1} A \\ &\quad - (D^T F + A^T S_{t+1} B)(F^T F + B^T S_{t+1} B)^{-1}(F^T D + B^T S_{t+1} A) \end{aligned}$$

As expected, solving the Bellman Equation gives us the control law "for free":

$$\begin{aligned} u_t^*(x) &= -L_t x \\ &= -a = -R^{-1} Z^T x \\ &= -(F^T F + B^T S_{t+1} B)^{-1}(F^T D + B^T S_{t+1} A)x \end{aligned}$$

Outline

Setup I

Take the deterministic case and add independent random vectors:

$$x_{t+1} = Ax_t + Bu_t + W\xi_t$$
$$\xi_t \sim \mathcal{N}(0, \mathbb{I})$$
$$x_0 \sim \mathcal{N}(m_0, P_0)$$

Our total expected cost to minimize is

$$J_N(u) = \mathbb{E}[\sum_{t=0}^{N-1} ||Dx_t + Fu_t||^2 + x_N^T Q x_N]$$

Setup II

The cost-to-go is now an expectation which forms the Stochastic Bellman Equation:

$$V_t(x) = \min_u \left[||Dx + Fu||^2 + \mathbb{E}_\xi [V_{t+1}(Ax + Bu + W\xi)] \right]$$
$$V_N(x) = x^T Q x$$

We append a constant to our ansatz:

$$V_t(x) = x^T S_t x + \alpha_t$$
$$S_N = Q$$

Derivation I

To prove the ansatz as before, we take the expectation of V_{t+1}. Let $f(x,u) = Ax + Bu$.

$$\begin{aligned}
\mathbb{E}[V_{t+1}(f(x,u) + W\xi) &= \mathbb{E}[(f(x,u) + W\xi_t)^T S_{t+1}(f(x,u) + W\xi)] + \alpha_{t+1} \\
&= f(x,u)^T S_{t+1} f(x,u) + \cancel{2\mathbb{E}[f(x,u)^T S_{t+1} W\xi]}^{\mathbb{E}(\xi)=0} \\
&+ \mathbb{E}[\xi^T W^T S_{t+1} W\xi] + \alpha_{t+1} \\
&= f(x,u)^T S_{t+1} f(x,u) + \mathbb{E}[tr(\xi\xi^T W^T S_{t+1} W)] + \alpha_{t+1} \\
&= f(x,u)^T S_{t+1} f(x,u) + tr(W^T S_{t+1} W) + \alpha_{t+1}
\end{aligned}$$

Since $\alpha_N = 0$, the α_t term accumulates the error over time due to noise. Thus

$$\alpha_t = \sum_{j=t}^{N-1} tr(W_j^T S_{j+1} W_j)$$

Derivation II

Returning to the Bellman Equation, the objective is identical to that of the deterministic case

$$V_t(x) = \min_u [||Dx + Fu||^2 + f(x,u)^T S_{t+1} f(x,u)] + \sum_{j=t}^{N-1} tr(C_j^T S_{j+1} C_j)$$

Thus, the optimal control law for the stochastic case is identical to the deterministic control law

$$u^*(x) = -L_t x$$

However, the cost of this policy is greater as a result of error accumulated due to noise.

Outline

Setup

In this case, we cannot observe the exact state of the system, we only have a noisy measurement.

$$
\begin{aligned}
x_{t+1} &= Ax_t + Bu_t + W\xi_t \\
y_t &= Hx_t + G\xi_t \\
\xi_t &\sim \mathcal{N}(0, \mathbb{I}) \\
x_0 &\sim \mathcal{N}(m_0, P_0)
\end{aligned}
$$

The control law depends on the observation history rather than the state $u^* = u^*(y)$. Assume we have derived a Kalman Filter to predict system state

$$
\begin{aligned}
\hat{x}_{t+1|t} &= A\hat{x}_{t+1|t} + Bu_t + K_t v_t \\
v_t &= y_t - H\hat{x}_{t|t-1}
\end{aligned}
$$

Result

The total cost becomes:

$$J_N(u) = \mathbb{E}\left[\sum_{t=0}^{N-1} ||D\hat{x}_{t|t-1} + Fu_t||^2 + \hat{x}_{N|N-1}^T Q\hat{x}_{N|N-1}\right] \\ + tr(P_t D^T D) + tr(P_N Q)$$

We're solving the fully observable case for dynamics with v_t:

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t-1} + Bu_t + K_t v_t$$

Thus, the optimal control law does not depend on the noise covariance, as before:

$$u_t^* = -L_t\hat{x}_{t|t-1}$$

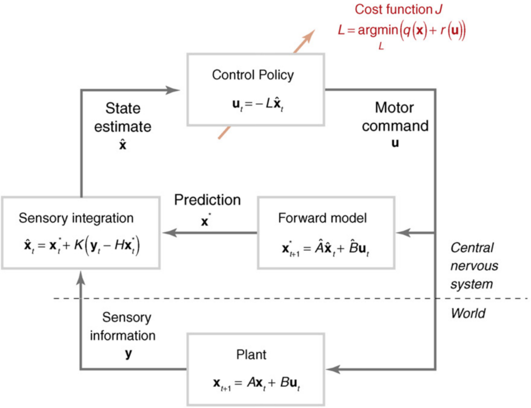
However, the cost is even higher for this policy as we have accumulated estimation noise as well as state noise.

Separability

In the full LQG, we're solving two problems:

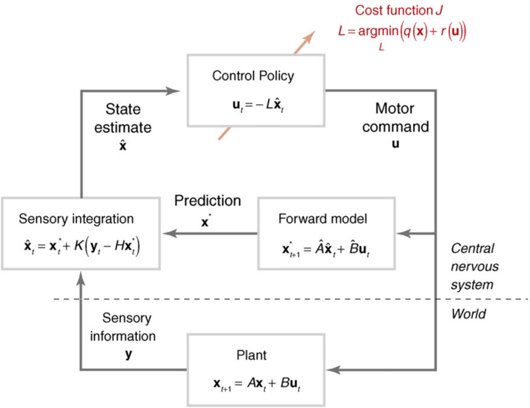
- Control (Linear Quadratic Regulator)
- Estimation (Linear Quadratic Estimator)

This is called **separability** and is the case regardless of the cost function.



Certainty Equivalence

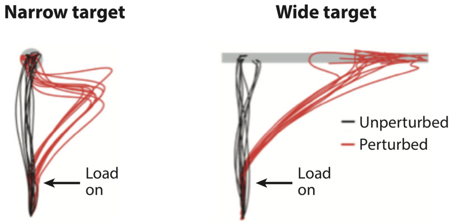
In the case of quadratic costs, we have what is called **certainty equivalence**. The controller acts as if the estimate $\hat{x}_{t|t-1}$ were the true state x_t with certainty. This is the optimal strategy in the case of quadratic cost.



Redundancy

> *Redundancy is not a 'problem'; on the contrary, it is part of the solution to the problem of performing tasks well.*
>
> *Todorov 2002*

Uncontrolled Manifold – The null space of the control law, where variability in task-irrelevant dimensions is shuttled. This is especially important for modeling perturbation response, and noise can be thought of as intrinsic perturbation to the task goal. This is "structured variability".

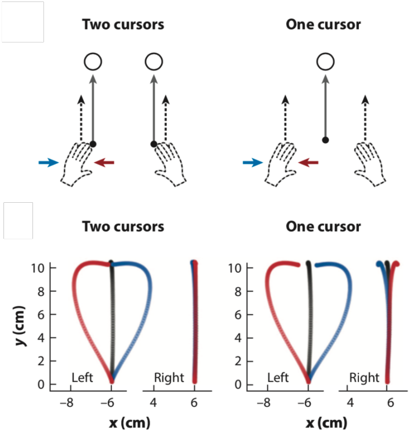


Synergies

"OCT can predict how muscles work in a synergistic manner without using the concept of synergies as an explanatory concept"

Diedrichson et al. 2009

Synergy – A control subspace. The rank of the linear control law L_t. Optimality can be achieved by recruiting more than one actuator, whereas the task goal can be accomplished using only one actuator. The one-step example is a simple synergy.



LQG+SDN

Quantitatively, the relationship between motor noise and control magnitude is surprisingly simple. Such noise has been found to be multiplicative.

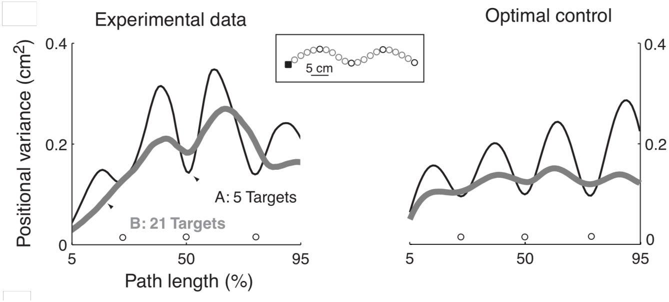
Todorov, 2005

$$
\begin{aligned}
&\textit{Dynamics} && \mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + \boldsymbol{\xi}_t + \sum_{i=1}^{c} \varepsilon_t^i C_i \mathbf{u}_t \\
&\textit{Feedback} && \mathbf{y}_t = H\mathbf{x}_t + \boldsymbol{\omega}_t + \sum_{i=1}^{d} \epsilon_t^i D_i \mathbf{x}_t \\
&\textit{Cost per step} && \mathbf{x}_t^T Q_t \mathbf{x}_t + \mathbf{u}_t^T R \mathbf{u}_t
\end{aligned}
$$

The control law is shown to be linear, but analytical separability is lost.

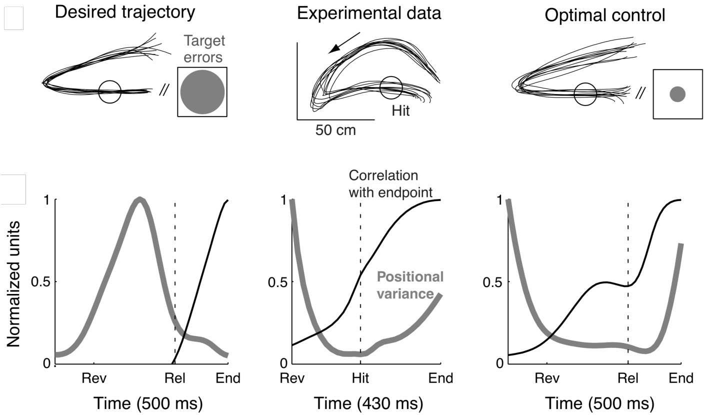
Via Point Task

All analysis within subject. Condition B is the average trajectory of condition A. Desired trajectory would predict no difference between A and B. "Variability is reduced where accuracy is most needed and is allowed to increase elsewhere."



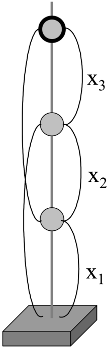
Ping Pong Task

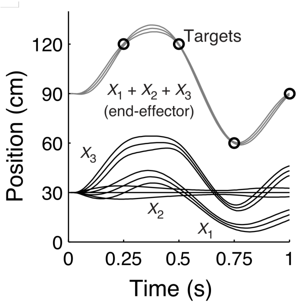
The desired trajectory controller is an optimal controller for a 10 via-point based on the average trajectory of the optimal control with endpoint error only. Positional variance is $det(\Sigma_t(x, y, z))$.

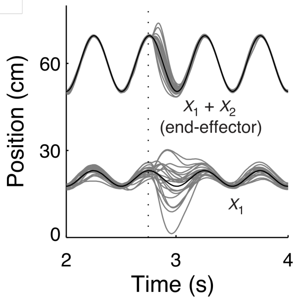


"Arm" Model Tasks I

Mechanical redundancy for 3 DoFs and response to perturbation for 2 DoFs. End effector must pass through 4 targets. Fluctuations in X_2 compensates fluctuations in X_1 and X_3.

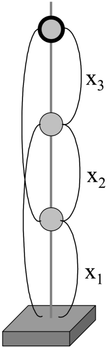


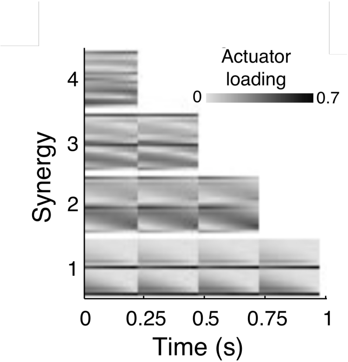




"Arm" Model Tasks II

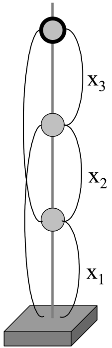
Synergies as nonzero eigenvectors of the control law. 10-dimensional problem (19 actuators).

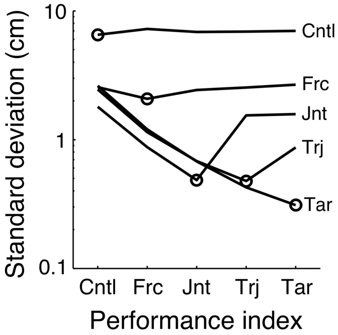




"Arm" Model Tasks III

Hierarchy for performance criteria at different levels of description: Control, Force, Joint angles, Endpoint trajectory, Target error.





Discussion

The motor system is not just one big feedback loop; rather, it is highly distributed and provides multiple pathways through which feedback can influence behaviour.

Scott, 2004

- Results
  - LQG+SDN recovers several important aspects of human movement
- Drawbacks
  - Model is a balance between detail and tractability
  - Cost function is a modeling choice
- Open Questions
  - How is control and estimation implemented in the brain?
  - How do we acquire models for control?
  - What cost functions do we optimize for?
  - How do we discover costs?
  - How does variability change throughout learning?

References I

- Jörn Diedrichsen, Reza Shadmehr, and Richard B. Ivry, *The coordination of movement: Optimal feedback control and beyond*, Trends in Cognitive Sciences **14** (2010), no. 1, 31–39.
- T Flash and N Hogan, *The coordination of arm movements: An experimentally confirmed mathematical model*, The Journal of Neuroscience **5** (1985), no. 7, 1688–1703 (en).
- Christopher M. Harris and Daniel M. Wolpert, *Signal-dependent noise determines motor planning*, Nature **394** (1998), no. 6695, 780–784 (en).
- Stephen H. Scott, *Optimal feedback control and the neural basis of volitional motor control*, Nature Reviews Neuroscience **5** (2004), no. 7, 532–545 (en).

References II

Emanuel Todorov and Michael I. Jordan, *Optimal feedback control as a theory of motor coordination*, Nature Neuroscience **5** (2002), no. 11, 1226–1235 (en).

Emanuel Todorov, *Stochastic Optimal Control and Estimation Methods Adapted to the Noise Characteristics of the Sensorimotor System*, Neural Computation **17** (2005), no. 5, 1084–1108 (en).

Outline

The random variable v_t is an observation prediction error. From filter theory we know that the distribution of the Kalman estimate $\mathbb{E}[\hat{x}_k|y_{k-1}] = \mathcal{N}(\hat{x}_{t|t-1}, P_t)$

Our goal is to write the cost using $\hat{x}_{t|t-1}$ rather than x_t. We use expectations conditioned on past observations and write x_t as

$$x_t = \hat{x}_{t|t-1} + \tilde{x}_{t|t-1}$$

where $\tilde{x}_{t|t-1}$ is a random variable representing estimation error, and $\hat{x}_{t|t-1}$ is a function of y_{t-1}. The error $\tilde{x}_{t|t-1} \sim \mathcal{N}(0, P_t)$ where P is the Kalman covariance. Our total expected cost to minimize is

$$J_N(u) = \mathbb{E}[\sum_{t=0}^{N-1} \mathbb{E}[||Dx_t + Fu_t||^2|y_{t-1}] + \mathbb{E}[x_N^T Q x_N|y_{t-1}]]$$

Let $D\hat{x}_t + Fu_t = \Lambda(y_{t-1})$ for convenience and expand x_t

$$
\begin{aligned}
J_N(u) &= \mathbb{E}[\sum_{t=0}^{N-1} \mathbb{E}[(\Lambda(y_{t-1}) + D\tilde{x}_{t|t-1})^T(\Lambda(y_{t-1}) + D\tilde{x}_{t|t-1})|y_{t-1}] \\
&\quad + \hat{x}_{N|N-1}^T Q \hat{x}_{N|N-1} + \mathbb{E}[\tilde{x}_{N|N-1}^T Q \tilde{x}_{N|N-1}|y_{t-1}]] \\
&= \mathbb{E}[\sum_{t=0}^{N-1} \mathbb{E}[\Lambda^T\Lambda + \cancel{2\Lambda^T D\tilde{x}_{t|t-1}}^{\mathbb{E}[\tilde{x}_{t|t-1}|y_{t-1}]=0} + \tilde{x}_{t|t-1}^T D^T D\tilde{x}_{t|t-1}|y^{t-1}] \\
&\quad + \hat{x}_{N|N-1}^T Q \hat{x}_{N|N-1} + tr(\mathbb{E}[\tilde{x}_{N|N-1}\tilde{x}_{N|N-1}^T|y^{t-1}]Q)] \\
&= \mathbb{E}[\sum_{t=0}^{N-1} \Lambda^T\Lambda + tr(\mathbb{E}[\tilde{x}_{t|t-1}\tilde{x}_{t|t-1}^T|y^{t-1}]D^T D) \\
&\quad + \hat{x}_{N|N-1}^T Q \hat{x}_{N|N-1} + tr(P_N Q)]
\end{aligned}
$$

Outline

Why Quadratic Cost? I

Why not something more akin to the sum, which would be a proxy for power/energy?

$$x_{t+1} = Ax_t + Bu_t + \xi$$

$$\begin{aligned}
\mathbb{V}(x) &= \mathbb{E}(x^2) - \mathbb{E}(x)^2 \\
\mathbb{V}(x) &= \mathbb{E}(x^2) - (Ax + Bu + \cancel{\mathbb{E}[\xi]})^2 \\
\mathbb{V}(x) &= \mathbb{E}(x^2) - (Ax + Bu)^2
\end{aligned}$$

Why Quadratic Cost? II

$$
\begin{aligned}
\mathbb{E}[x^2] &= \mathbb{E}[x^T x] \\
&= \mathbb{E}[(x^T A^T + u^t B^T + \xi)(Ax + Bu + \xi)] \\
&= \mathbb{E}[x^T A^T A x + u^T B^T A x + x^T A^T B u + u^T B^T B u + \xi^2] \\
&= \mathbb{E}[\xi^2] + (Ax + Bu)^2 \\
&= \mathbb{V}[\xi] + (Ax + Bu)^2
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{V}(x) &= \mathbb{V}[\xi] + (Ax + Bu)^2 - (Ax + Bu)^2 \\
\mathbb{V}(x) &= \mathbb{V}[\xi]
\end{aligned}
$$

Why Quadratic Cost? III

Let $\xi = cu\phi$ where $\phi \sim \mathcal{N}(0, 1)$:

$$\begin{aligned}
\mathbb{V}(\xi) &= c^2u^2\mathbb{E}[\phi^2] - \cancel{c^2u^2\mathbb{E}[\phi]^2} \\
\mathbb{V}(\xi) &= c^2u^2\mathbb{V}[\phi] \\
\mathbb{V}(\xi) &= c^2u^2
\end{aligned}$$

Thus

$$\mathbb{V}(x) = c^2u^2.$$

Minimizing control is (somewhat) equivalent to minimizing movement variability.

Outline

9 Appendix
- Full LQG Derivation
- Is Quadratic Cost Reasonable?
- Redundancy Example
- Deterministic CT LQG
- Maximum Principle Solution
- Stochastic CT LQG

Redundancy the Math Way I

Deviation Δx from the average trajectory $\bar{x}$ is redundant if

$$\Delta v^*(\bar{x} + \Delta x) - \Delta v^*(\bar{x}) = 0$$

Stochastic Dynamics:

$$dx = a(x,t)dt + B(x,t)udt + \sum_{i=1}^{k} C_i(x,t)ud\xi_i$$

State change due only to optimal control:

$$\dot{x}_u = B(\bar{x} + \Delta x, t)u^*(\bar{x} + \Delta x, t)$$

Correction can be defined as the projection:

$$corr(\Delta x) \equiv \langle \Delta x, -\dot{x}_u \rangle$$

Redundancy the Math Way II

Taylor expanding $v^*(\Delta x)$ and $corr(\Delta x)$ and using this fact of optimal control:

$$u^* = -Z(x,t)^{-1} B(x,t)^T v_x^*(x,t)$$
$$Z(x,t) \equiv 2R(x,t) + \sum_{i=1}^{k} C_i(x,t)^T v_{xx}^*(x,t) C_i(x,t),$$

where R is the control cost coefficient matrix $u^T R u$, we find

$$\Delta v^*(\Delta x, t) \approx \langle \Delta x, v_x^*(x,t) + v_{xx}^*(x,t)\Delta x \rangle$$
$$corr(\Delta x) \approx \langle \Delta x, v_x^*(x,t) + v_{xx}^*(x,t)\Delta x \rangle_{-BZ^{-1}B^T}$$

Thus, corrections to deviations only occur in the direction of value function increase.

Outline

Deterministic CT LQG I

Assumptions

- Linearity in the Dynamics
- Quadratic costs
- Continuous Time and Space
- Noise Free

Dynamics

$$f(x, u) = \dot{x} = Ax + Bu$$

Cost

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

Policy

$$\pi(x) = u$$

Deterministic CT LQG II

Ansatz for the optimal Cost-to-go

$$v^* = x^T S x$$

Cost-to-go

$$\begin{aligned} v^\pi(x,t) &= l_T(x(T)) + \int_t^T l(x(s)), \pi(x(s)) ds \\ &= l_T + \int_t^{t+dt} l_s ds + \int_{t+dt}^T l_s ds \\ &\approx L_t + l(x,\pi(x))dt - L_t + [l_T + \int_{t+dt}^T l_s ds] \\ &= l(x,\pi(x))dt + v^\pi(x(t+dt), t+dt) \end{aligned}$$

Deterministic CT LQG III

Expanding the new cost-to-go to first order and plugging in

$$x(t+dt) \approx x(t) + f(x(t+dt), u)dt$$
$$v^\pi(x(t+dt), t+dt) \approx v^\pi(x,t) + v_t^\pi(x,t)dt + f(x,u)^T v_x^\pi(x,t)dt$$

since $\frac{d}{dt} = \frac{\partial}{\partial t} + \frac{\partial x}{\partial t}\frac{\partial}{\partial x}$. We find

$$v^\pi(x,t) \approx l(x,\pi(x))dt + v^\pi(x,t) + v_t^\pi(x,t)dt + f(x,u)^T v_x^\pi(x,t)dt.$$

Rearranging yields

$$-v_t^\pi(x,t) = l(x,\pi(x)) + f(x,\pi(x))^T v_x^\pi(x,t) \tag{1}$$

Which is an HJB Equation. This is analogous to the discrete case where we would derive a Bellman Equation and use the Principle of Optimality to

Deterministic CT LQG IV

solve the optimal path and controls recursively. We can solve the PDE using the boundary condition $(l(x(T)))$ but .
Letting controls be optimal, we have

$$-v_t^*(x,t) = \min_u [l(x,u) + f(x,u)^T v_x^*(x,t)]$$

and

$$\pi^*(x) = \operatorname*{argmin}_u [l(x,u) + f(x,u)^T v_x^*(x,t)]$$

Deterministic CT LQG V

[Without proof] In the LQ case, taking the minimum w.r.t. u is NEC and SUF:

$$\begin{aligned} 0 &= \frac{d}{du^*}[l(x,u^*) + f(x,u^*)^T v_x^*(x,t)] \\ &= \frac{d}{du^*}(\frac{1}{2}x^T Qx + \frac{1}{2}u^{*T} Ru^*) + \frac{d}{du^*}(Ax + Bu^*)^T v_x^*(x,t) \\ &= Ru^* + B^T v_x^*(x,t) \end{aligned}$$

Plugging in the ansatz $v^*(x,t) = \frac{1}{2}x^T Sx$:

$$\pi^*(x) = -R^{-1}B^T v_x^*(x,t)$$
$$\pi^*(x) = -R^{-1}B^T Sx$$

Deterministic CT LQG VI

Plugging this into the HJB equation we find

$$\begin{aligned} -v_t^* &= l(x, u^*) + f(x, u^*)v_x^* \\ &= \frac{1}{2}[x^t Qx + x^T SBR^{-1}RR^{-1}B^T Sx] + (Ax - BR^{-1}B^T Sx)^T Sx) \\ &= \frac{1}{2}x^T(Q + A^T S + SA - SBR^{-1}B^T S)x \end{aligned}$$

In the infinite horizon case the time derivative of the Cost-to-go is 0 and we have an algebraic Ricatti equation

$$0 = Q + A^T S + SA - SBR^{-1}B^T S$$

Note that this is a different equation than the discrete-time version! This equation can be solved using the particular dynamics of the problem by constraining the eigenvalues.

Outline

Maximum Principle Solution I

Start with the HJB equation and assume u is the optimal control $\pi(x)$

$$0 = v_t(x,t) + l(x,\pi(x)) + f(x,\pi(x))^T v_x(x,t)$$

Differentiate w.r.t. x and remember $\frac{d}{dx} = \frac{\partial}{\partial x} + \frac{\partial u}{\partial x}\frac{\partial}{\partial u}$

$$\begin{aligned} 0 = {} & v_{x,t}(x,t) + l_x(x,\pi(x)) + \pi_x^T(x) l_u(x,\pi(x)) \\ & + f_x(x,\pi(x))^T v_x(x,t) + \pi_x^T(x) f_u(x,\pi(x))^T v_x(x,t) \\ & + f(x,\pi(x))^T v_{xx}(x,t) \end{aligned}$$

Maximum Principle Solution II

Gathering terms

$$
\begin{aligned}
0 = &\, v_{x,t}(x,t) + l_x(x,\pi(x)) \\
&+ f_x(x,\pi(x))^T v_x(x,t) \\
&+ f(x,\pi(x))^T v_{xx}(x,t) \\
&+ \pi_x^T(x)\cancel{(l_u(x,\pi(x)) + f_u(x,\pi(x))^T v_x(x,t))}
\end{aligned}
$$

We can cancel the last term since it is simply the u derivative which we know is 0 under the optimal controls:

$$
\begin{aligned}
0 &= \frac{d}{du^*}[l(x,u^*) + f(x,u^*)^T v_x^*(x,t)] \\
&= l_u(x,u^*) + f_u(x,u^*)^T v_x^*(x,t)
\end{aligned}
$$

Maximum Principle Solution III

Rearranging

$$-v_{x,t}(x,t) + f(x,\pi(x))^T v_{xx}(x,t) = l_x(x,\pi(x)) + f_x(x,\pi(x))^T v_x(x,t)$$

The left-hand side is the total derivative $\frac{d}{dt} = \frac{\partial}{\partial t} + \sum_i \frac{\partial x_i}{\partial t}\frac{\partial}{\partial x_i}$ of v_x. The new Hamiltonian is thus

$$\dot{v}_x(x,t) = l_x(x,\pi(x)) + f_x(x,\pi(x))^T v_x(x,t)$$

Letting $\lambda = v_x(t)$ and writing the HJB equation's minimization as H:

$$H(x,u,\lambda) = l(x,u) + f(x,u)^T \lambda$$

Maximum Principle Solution IV

we now have minimizing conditions and a boundary value problem

$$\begin{aligned}
\dot{x} &= H_\lambda(x, u, \lambda) \\
-\dot{\lambda} &= H_x(x, u, \lambda) \\
u &= \operatorname*{argmin}_s H(x, s, \lambda) \\
x(0) &= x_0 \\
\lambda(T) &= l_x(x(T))
\end{aligned}$$

Outline

9 Appendix
- Full LQG Derivation
- Is Quadratic Cost Reasonable?
- Redundancy Example
- Deterministic CT LQG
- Maximum Principle Solution
- Stochastic CT LQG

Stochastic CT LQG I

Assumptions

- Linearity in the Dynamics
- Quadratic costs
- Continuous Time and Space
- Additive Gaussian Noise

Stochastic Dynamics

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

where ξ is a Wiener process with covariance $\Sigma(x, u, t)$. Rather than minimizing the controlled trajectory, we will minimize the expectation under the process. The Cost-to-go is now

$$v^{\pi}(x, t) = l(x, \pi(x)) + \langle v^{\pi}(x(t + dt), t + dt)\rangle_{x_t}$$

Stochastic CT LQG II

Taking the expectation of the second-order approximation

$$v^{\pi}(x,t) \approx l(x,\pi(x))dt + v^{\pi}(x,t) + v_t^{\pi}(x,t)dt + v_x^{\pi}(x,t)\langle dx\rangle + \frac{1}{2}v_{xx}^{\pi}(x,t)\langle dx^2\rangle$$

We can use some rules from stochastic calculus:

$$\begin{aligned}
dt^2 &= 0 \\
dt d\xi &= 0 \\
\langle d\xi^2\rangle &= \Sigma(x,u,t)dt \\
\langle dx\rangle &= f(x,u,t)dt + 0 \\
\langle dx^2\rangle &= f(x,u,t)^2 dt^2 + 2f(x,u,t)dt\langle d\xi\rangle + \langle d\xi^2\rangle \\
&= \langle d\xi^2\rangle \\
&= \Sigma(x,u,t)dt
\end{aligned}$$

Stochastic CT LQG III

To find the stochastic HJB equation

$$-v_t^\pi(x,t) = l(x,\pi(x)) + f(x,\pi(x),t)^T v_x^\pi(x,t) + \frac{1}{2}\Sigma(x,\pi(x),t)v_{xx}^\pi(x,t)$$

Letting Σ go to 0 recovers the deterministic case. This is a boundary problem as before, with terminal condition $v^\pi(x,t_f) = \phi(x)$.