Adaptive Control and Reinforcement Learning (16-899C, S09) Lecture #13 (2/26/09)

Separation Theorem

Lecturer: Atkeson *Scribe: Laura Lindzey*

1 Estimation

1.1 Stock Market Prediction Using KF

How do we combine the estimates of two "experts"?

expert	predicted Δ	Bias	Var
1	-250	0	1
2	-200	0	10
actual	-80		

Linearized system and measurement model :
$\hat{x} = Ax + Bu + w$
$z = Cx + v$

Process Update :
$x_{k+1}^{-} = Ax_k^{+} + Bu_k$
$p_k^{+} = \frac{Vp_k^{-}}{p_k^{-}+V}$
$p_{k+1}^{-} = A^T p_k^{+} A + W$

Measurement dynamics :
$p^{+} = p^{-} - kcp^{-}A^T$

Variable Definitions :
x^{-} before measurement
x^{+} after measurement
z measurement received
$p^{-} = var(x^{-})$
$p^{+} = var(x^{+})$
$V = var(z - x)$
$W = var(w)$, where w is system noise
k is the kalman gain: $k = Ap^{+}c^T(cp^{+}c^T + V)^{-1}$

1.2 Estimation without a model?

The above equations work If you have the model. Just estimate parameters from data, then plug in. However, we're interested in the alternative: how to do this WITHOUT using a model?

Could use a policy (e.g. vote Republican, listen to CNN, listen to expert 2)

2 Price of Variability

A more theoretical issue: there's a price to ignorance - what does uncertainty/variance cost you?

Given a random variable x and a utility function v(x) that tells how much x is worth, we want to calculate $E[V(x)]$. As shown in figure **??** we have the following cases:

- a $\rightarrow$ constant term: $E[V(x)] = V_0$
- b $\rightarrow$ linear: $E[V(x)] = V_x^T x$
- c $\rightarrow$ quadratic: $E[V(x)] = tr[V_{xx}\Sigma)]$ (assuming 0-mean)

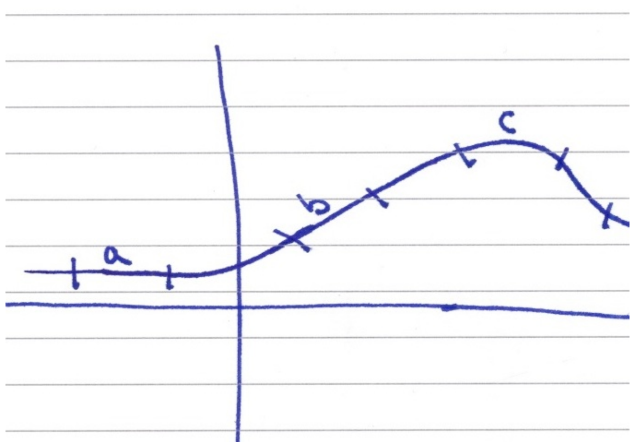

Figure 1:

Notes:
For constant or linear, sampling noise doesn't matter. If quadratic, it kills, because the quadratic term will always add to the cost.

There's a cost for uncertain belief states (value functions are quadratic, not linear)

3 Separation Theorem

Certainty Equivalence is a FACT for linear systems with gaussian noise

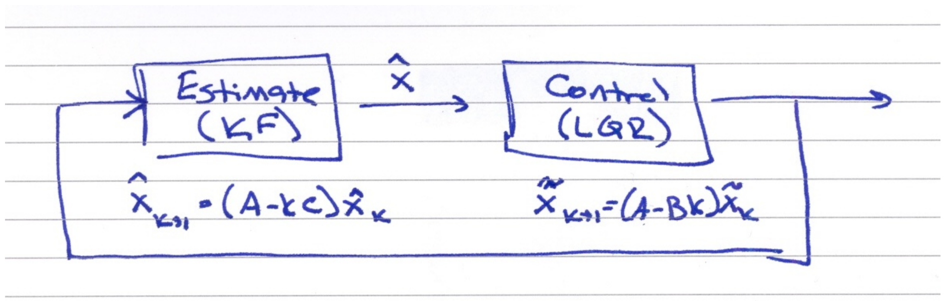

Figure 2:

Certainty Equivalence: the concept that we can design the KF and LQR entirely independently, and this is OK. More formally, action has no effect on p^+, p^- and neither does the measurement.

Actual measurements don't matter and controls don't matter with respect to variance estimate and kalman gain. Thus, we can calculate over all time what the kalman gains will be.

Given a system with 10 elements in the state, with only 3 sensors .... this is a POMDP!

If the observability matrix is of rank m, then at some point in the sequnce we will have measured everything over history.

Observability matrix: $\begin{bmatrix} cC \\ CA \\ CA^2 \\ .... \\ CA^{n-1} \end{bmatrix}$

LQR and the dynamics go backwards through time

KF goes forward in time

if we go far enough, everything converges to const/asymptotic value

3.1 KF/LQR Duality

Note the similarities in the equations for the KF/LQR (Measurement noise acts as u, process noise var is related to Q)

KF	**LQR**
A^T	A
C^T	B
W	Q
V	R

KF
$x^-_{k+1} = [A + KC]x^-_k - Kz_k$
$K = AP^-C^T[CP^-C^T + V]^{-1}$
$P^-_{k+1} = A(P^- - P^-C^T[CP^-C^T + V]^{-1}CP^-)A^T + W$ (Riccati equation)
$x^+_k = x^-_k + P^-C^T(CP^-C^T + V)^{-1}(y_k - Cx^-_k)$

LQR
$u_k = -Kx_k$
$K = (B^T SB + R)^{-1} B^T SA$
$S_{k+1} = A^T(S - SB[B^T SB + R]^{-1} B^T S)A + Q$, which is the algebraic Riccati equation again

3.2 What goes wrong (in CE)?

- may not be gaussian
- may not be linear
- may not have zero-mean noise
- may have wrong model
- may have unmodelled dynamics

the last 2 can *really* screw things up ...

3.3 poles and zeros

...are ways to talk about the modes of a system

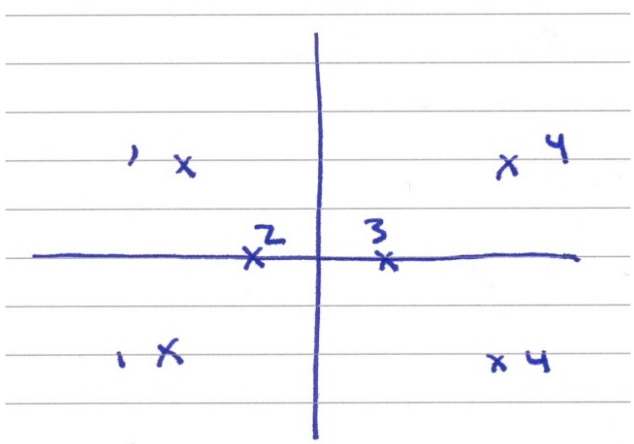

Figure 3:

As seen in figure **??**, there are 4 qualitatively different pole placements: (note that imaginary poles always come in pairs)

- 1: pair of imaginary poles (real part negative): there will be "jiggle"
- 2: pole on neg real axis: gives a mode that only decays

- 3: pole on real axis but pos: blows up
- 4: pair of imaginary poles (real part positive): jiggle AND blow up

The trick is to get jiggle at the same frequency - a good estimator puts poles in a circle around the origin

Problems:

- poles of KF and LQR wind up on top of each other
- if estimator has error, it will oscillate (causing controller to oscillate in same way).

We want 2 groups of poles - will have to pick high/low frequency for the controller/estimator, as shown in figure **??**. Essentially, this is chucking optimality in favor of stability.

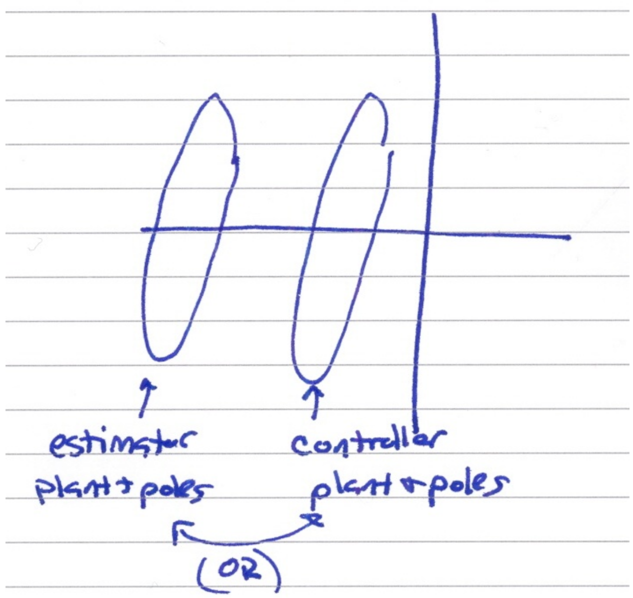

Figure 4:

This design process is called "Loop Transfer Recovery":

1) Place LQR poles

2) set $W = \alpha BB^T$, $V = CC^T$, where you increase α to place KF poles

- It is possible (due to plant dynamics) for this to not work

- this process cancels poles with zeros

In frequency domain, plant is: $\rightarrow \frac{N}{D} \rightarrow$

poles are zeros of D, zeros are zeros of N

(at this point, lecturer trailed off into mumbo-jumbo, and admitted such)

When controller has to move away from goal to get there in the end, this is a signature of non-minimum-phase system (meaning it has a zero in the right half plane). To cancel this zero, have to put an unstable pole out there ... which (naturally) causes instability if not placed exactly

3.4 exploration

There is NO exploration in Dual Control/Certainty Equivalence ....

.... but what if you need to explore?

Suppose you have a simple process: $F = ma = m\ddot{x}$ State is $\begin{bmatrix} cx \\ \dot{x} \\ \hat{m} \end{bmatrix}$; this becomes nonlinear!

1) point of exploration is $\hat{m} \rightarrow m$ (minimize some norm of p or some weighted sum of $x^T QX + u^T Ru$)

2) $E(\hat{x}^T Q\hat{x} + u^T Ru)$

3) $min(V(x_0))$ ....(lecturer seemed uncertain about this - is it different?)

4 3-minute rant

- ML often talks about grid worlds, where each cell has uncertainty

- often, the goal of exploration is to minimize some norm of p (which requires going everywhere and trying everything!)

- this may not be the right thing to do

- need to distinguish between information goals ($min|p|$) and utility goals (making $$, etc.)