NATIONAL TAIWAN UNIVERSITY
DEPARTMENT OF BIOMECHATRONICS ENGINEERING

Optimal Control

EE 5051

Final Project

Instructor: Feng-Li Lian
Student: Hao-Yu Chan B08611046

NOVEMBER 2023

Contents

1 Review: The Characteristics of Model-Following Systems as Synthesized by Optimal Control

The paper is the result of research performed at the Flight Research Department of Cornell Aeronautical Laboratory and performed in conjunction with the author's doctoral thesis for Yale University. This paper describes two model-following techniques and their application to practical systems. The first technique, which incorporates the model only in the performance index, was developed by Dr. R. E. Kalman[6]. The second technique, which incorporates the model-in-the-system has been used by Dr. Kalman, the Flight Research Department at CAL[7], and F. T. Smith[8].

1.1 Model-In-The-Performance-Index

The model-in-the-performance-index method uses the feedback gains generated by the optimal-control theory to alter the open-loop plant. The plant may be expressed in standard form as

$$\dot{\bar{x}} = F\bar{x} + G\bar{u}, \tag{1.1}$$

The performance index is defined as

$$2V = \int_0^T \{(\dot{\bar{y}} - L\bar{y})'Q(\dot{\bar{y}} - L\bar{y}) + \bar{u}'R\bar{u}\}dt, \tag{1.2}$$

where L is the model, $\bar{y}$ is the output vector and is defined as

$$\bar{y} = H\bar{x}. \tag{1.3}$$

The form of the optimal control for this performance index is

$$\bar{u}^0 = -(G'H'QHG + R)^{-1}\{G'\bar{\lambda} + G'H'Q(HF - LH)\}\bar{x}. \tag{1.4}$$

The "hat" matrices are the same form as the canonical equations of the regulator problem. The Riccati equation for the "hat" matrices has the form

$$\dot{P}(t) - P(t)\hat{G}\hat{R}^{-1}\hat{G}'P(t) + P(t)\hat{F} + \hat{F}'P(t) + \hat{H}'\hat{Q}\hat{H} = [0], \tag{1.5}$$

$$\hat{R} = G'H'QHG + R, \tag{1.6}$$

$$\hat{F} = F - G\hat{R}^{-1}G'H'Q(HF - LH), \tag{1.7}$$

$$\hat{G} = G, \tag{1.8}$$

$$\hat{H} = HF - LH, \tag{1.9}$$

and

$$\hat{Q} = Q - QHG\hat{R}^{-1}G'H'Q. \tag{1.10}$$

Since it is desirable to have constant feedback gains in airplane flight control systems, only the steady-state value of the Riccati equation will be used in this work. The control can be found from

$$u^0 = -\{(\hat{R}^{-1}G'P + \hat{R}^{-1}G'H'Q(HF - LH)\}\bar{x}. \tag{1.11}$$

1.2 Model-In-The-Performance-Index Applied to Airplane Lateral Equations

This particular synthesis technique was applied to practical systems, involving the airplane lateral equations of motion. The following lateral equations were used:

$$\begin{bmatrix} \dot{\phi} \\ \ddot{\phi} \\ \dot{\beta} \\ \dot{\gamma} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & L_p & L_\beta & L_r \\ g/V & 0 & Y_\beta & -1 \\ N_{\dot{\beta}}\frac{g}{V} & N_p & N_\beta + N_{\dot{\beta}}Y_\beta & N_\gamma - N_{\dot{\beta}} \end{bmatrix} \begin{bmatrix} \phi \\ \dot{\phi} \\ \beta \\ \gamma \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ L_{\delta_\gamma} & L_{\delta_a} \\ Y_{\delta_\gamma} & 0 \\ N_{\delta_\gamma} + N_{\dot{\beta}}Y_{\delta_\gamma} & N_{\delta_a} \end{bmatrix} \begin{bmatrix} \delta_\gamma \\ \delta_a \end{bmatrix}$$

where

$$\begin{aligned} \phi &= \text{bank angle} \\ \beta &= \text{sideslip angle} \\ \gamma &= \text{yaw angle} \\ \delta_\gamma &= \text{rudder deflection} \\ \delta_a &= \text{aileron deflection} \end{aligned}$$

These equations have been transformed to decouple the inertia terms.

In this example, the B-26 airplane mechanized by the Flight Research Department of the Cornell Aeronautical Laboratory as a variable stability airplane was chosen. The matrices for the B-26 were

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -2.93 & -4.75 & .78 \\ .086 & 0 & -.11 & -1 \\ 0 & -.042 & 2.59 & -.39 \end{bmatrix}, \quad G = \begin{bmatrix} 0 & 0 \\ 0 & -3.91 \\ .035 & 0 \\ -2.53 & .31 \end{bmatrix}$$

The F matrix here has a typo, because the eigenvalue of the F from the paper won't be the same as the eigenvalue given at Table I.
The model was

$$L = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -1 & -73.14 & 3.18 \\ .086 & 0 & -.11 & -1 \\ .0086 & .086 & 8.95 & -.49 \end{bmatrix}$$

This model would result approximately in the following response parameters:

$$\tau_R^{-1} = \text{inverse of roll mode time constant} = 1 \text{ sec}^{-1}$$
$$\tau_S^{-1} = \text{inverse of spiral mode time constant} = 0 \text{ sec}^{-1}$$
$$\zeta_d = \text{Dutch roll damping ratio} = 0.1$$
$$\omega_{n_d} = \text{Dutch roll natural frequency} = 3 \text{ rad/sec}$$

I found these parameters were related to the eigenvalues of the model L, by using **eig(F)** in Matlab, the eigenvalues were -0.2882±2.9423i, -1.0233, -0.0003, so I assumed that inverse of roll mode time constant τ_R^{-1} and inverse of spiral mode time constant τ_S^{-1} are related to the real eigenvalues, dutch roll damping ratio ζ_d and dutch roll natural frequency ω_{n_d} are related to the complex eigenvalue. Learnt from system dynamics, we know that a damped system can be written in the form

$$\omega^2 + 2\zeta\omega_0\omega + \omega_0^2 = 0$$

So, assume the conjugated eigenvalues of the damped system were $a + ib$ and $a - ib$, the relationship between the conjugated eigenvalues and ω_0, ζ will be

$$\begin{cases} (a+ib) * (a-ib) = \omega_0^2 \\ (a+ib) + (a-ib) = 2a = 2 * \zeta * \omega_0 \end{cases} \Longrightarrow \begin{cases} \omega_0 = \sqrt{a^2+b^2} \\ \zeta = \dfrac{|a|}{\sqrt{a^2+b^2}} \end{cases}$$

The value of the R matrix for this example, as well as the examples that follow. The definition of Q matrix is

$$Q(a,b,c,d) = \begin{bmatrix} 10^a & 0 & 0 & 0 \\ 0 & 10^b & 0 & 0 \\ 0 & 0 & 10^c & 0 \\ 0 & 0 & 0 & 10^d \end{bmatrix}$$

And $Q(0)$ is equal to 0 in this example. The response to an initial sideslip angle (β) and an initial bank angle rate ($\dot{\phi}$) are shown in the next section.

1.3 B-26 Model-In-The-Performance-Index Responses

Two responses were conducted in this report: (i) initial sideslip angle response, (ii) initial bank angle rate response. In each responses, 5 models were tested: OPEN-LOOP, MODEL, Q(0, 6, 0, 6), Q(1, 1, 1, 1), Q(3, 3, 3, 3).
I have written the following code in Matlab to reproduce the response.

```
% Plant from the paper
% [phi' phi'' beta' gamma']
F=[0 1 0 0;
    0 -2.93 -4.75 .78;
    .086 0 -.11 -1;
    0 -.042 2.59 -.39];

G=[0 0;
    0 -3.91;
    .035 0;
    -2.53 .31];

% Following model
L=[0 1 0 0;
    0 -1 -73.14 3.18;
    .086 0 -.11 -1;
    .0086 .086 8.95 -.49];

% Cost functions
Q1=diag([0 1e6 0 1e6]);
Q2=diag([10 10 10 10]);
Q3=diag([1e3 1e3 1e3 1e3]);

% Other matrix for the riccati equations
R=eye(2);
N=zeros(4, 2);
```

```
27 H=eye(4);
28 D=zeros(4, 2);
29
30 % Model in the performance index (Fhat=the third column in the first
   row of table 1)
31 % Q1
32 Rhat1=G'*H'*Q1*H*G+R;
33 Fhat1=F-G*inv(Rhat1)*G'*H'*Q1*(H*F-L*H);
34 Ghat1=G;
35 Hhat1=H*F-L*H;
36 Qhat1=Q1-Q1*H*G*inv(Rhat1)*G'*H'*Q1;
37 % Q2
38 Rhat2=G'*H'*Q2*H*G+R;
39 Fhat2=F-G*inv(Rhat2)*G'*H'*Q2*(H*F-L*H);
40 Ghat2=G;
41 Hhat2=H*F-L*H;
42 Qhat2=Q2-Q2*H*G*inv(Rhat2)*G'*H'*Q2;
43 % Q3
44 Rhat3=G'*H'*Q3*H*G+R;
45 Fhat3=F-G*inv(Rhat3)*G'*H'*Q3*(H*F-L*H);
46 Ghat3=G;
47 Hhat3=H*F-L*H;
48 Qhat3=Q3-Q3*H*G*inv(Rhat3)*G'*H'*Q3;
49
50 % =============Simulation=============
51 % initial sideslip angle (beta=1) [0 0 1 0]
52 % initial bank angle rate (phi'=1) [0 1 0 0]
53 X1=[0 0 1 0];
54
55 OpenModel=ss(F, G, H, D);
56 ssModel=ss(L, G, H, D);
57 Q1Model=ss(Fhat1, G, H, D);
58 Q2Model=ss(Fhat2, G, H, D);
59 Q3Model=ss(Fhat3, G, H, D);
60
```

```
t=0:0.01:5;
u1 = 0*ones(size(t));
u2 = 0*ones(size(t));

yOpen=lsim(OpenModel, [u1; u2], t, X1);
yModel = lsim(ssModel, [u1; u2], t, X1);
yQ1=lsim(Q1Model, [u1; u2], t, X1);
yQ2=lsim(Q2Model, [u1; u2], t, X1);
yQ3=lsim(Q3Model, [u1; u2], t, X1);

% for phi plot the first of the lsim, phi' plot the second of the
lsim, beta plot the third of the lsim, gamma plot the fourth of the
lsim
for a=1:4
figure(a)
plot(t, [yOpen(:, a)])
hold on
plot(t, [yModel(:, a)], LineStyle="--")
plot(t, [yQ1(:, a)], LineStyle=":")
plot(t, [yQ2(:, a)], LineStyle="-.")
plot(t, [yQ3(:, a)], LineStyle="-.")

legend(["Open", "Model", "Q(0, 6, 0, 6)", "Q(1, 1, 1, 1)", "Q(3, 3, 3,
3)"])
if a==1
    ylim([0 1])
    ylabel('$\phi$', Interpreter='latex')
elseif a==2
    ylim([-.2 1])
    ylabel('$\dot{\phi}$', Interpreter='latex')
elseif a==3
    ylim([-.002 .007])
    ylabel('$\beta$', Interpreter='latex')
elseif a==4
    ylim([-.01 .10])
```

```
93      ylabel('$\gamma$', Interpreter='latex')
94  end
95  grid minor
```

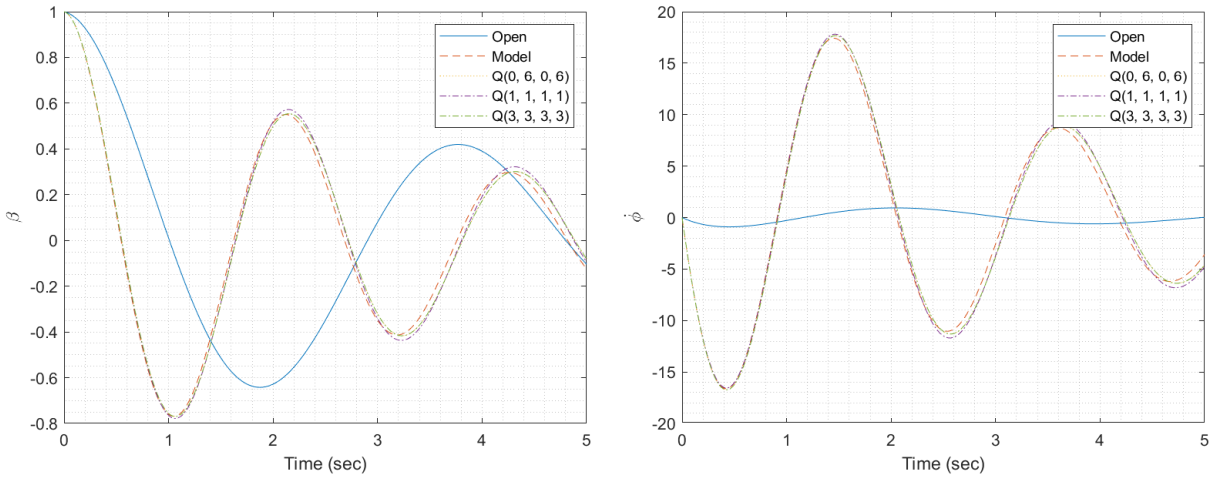

Fig. 1 B-26 model-in-the-performance-index responses to initial sideslip angle (β)

There is a difference in the result from the paper and from my simulation, the greatest difference is when $Q(1, 1, 1, 1)$, in the paper the line damps less than other cost functions, and stables a lot faster.

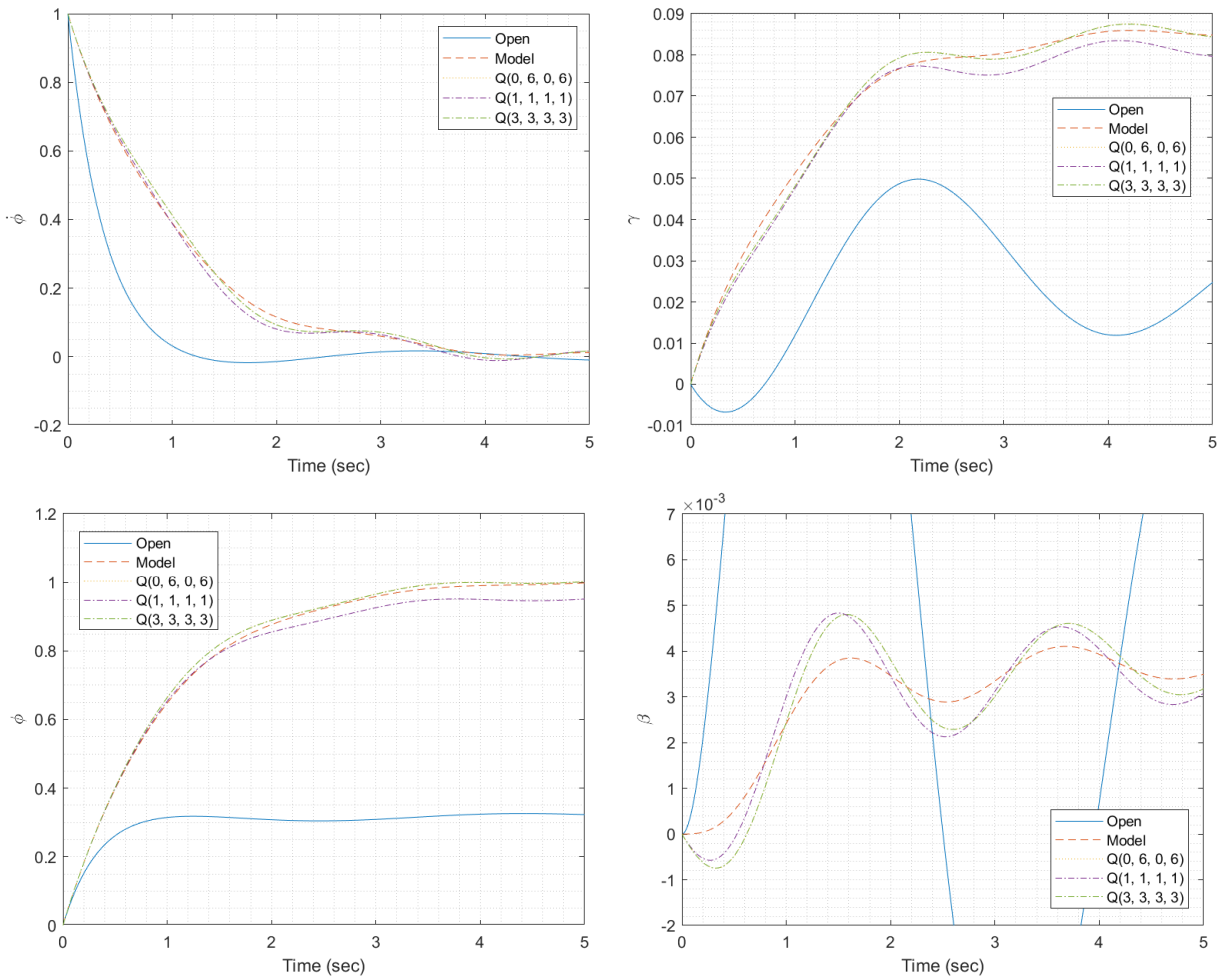

Fig. 2 B-26 model-in-the-performance-index responses to initial bank angle rate ($\dot{\phi}$)

In this part of simulation, the line from $Q(0,6,0,6)$ and $Q(3,3,3,3)$ seems to overlap a lot, although from the paper, they also seems to be close, but not as close as the result in my simulation. Also $Q(1,1,1,1)$ seems a little different from the result in the paper.

Open-Loop System				Model			
0	1	0	0	0	1	0	0
0	−2.93	−4.75	.78	0	−1	−73.14	3.18
.086	0	−.11	−1	.086	0	−.11	−1
0	−.042	2.59	−.39	.0086	.086	8.95	−.49
$Q(0,6,0,6)$				$Q(1,1,1,1)$			
0	1	0	0	0	1	0	0
0	−1	−73.14	3.18	0	−1.0129	−72.6968	3.1643
.086	−.0039	−.123	−1.0012	.0859	−.0038	−.123	−1.0012
.0086	.086	8.95	−.49	.0085	.0816	8.9349	−.49

Table I. Model-In-The-Performance-Index for B-26

	τ_R^{-1}	τ_S^{-1}	ζ_d	ω_{n_d}
Open-Loop System	-2.9817	.0017	.1346	1.6718
Model	-1.0233	-.0003	.0975	2.9564
$Q(0,6,0,6)$	-1.0506	0	.0965	2.9130
$Q(1,1,1,1)$	-1.0949	-.0022	.0909	2.9170

Table II. Parameters for different models

	Feedback Control Matrices			
$Q(0,6,0,6)$	.0034	.1111	.3707	.0357
	0	.4936	-17.49	.6138
$Q(1,1,1,1)$	.0033	.1089	.3786	.0346
	0	.4903	-17.38	.6098
$Q(3,3,3,3)$	.0034	.1110	.3707	.0357
	0	.4936	-17.49	.6138

Table III. Feedback Control Matrices for different Q

After reproducing the results of the paper, I was still curious about how different Q affects the control. So I simulated the control using $Q(6,0,0,0)$, $Q(0,6,0,0)$, $Q(0,0,6,0)$, $Q(0,0,0,6)$ and compare them with the model and open-loop system with initial sideslip angle.

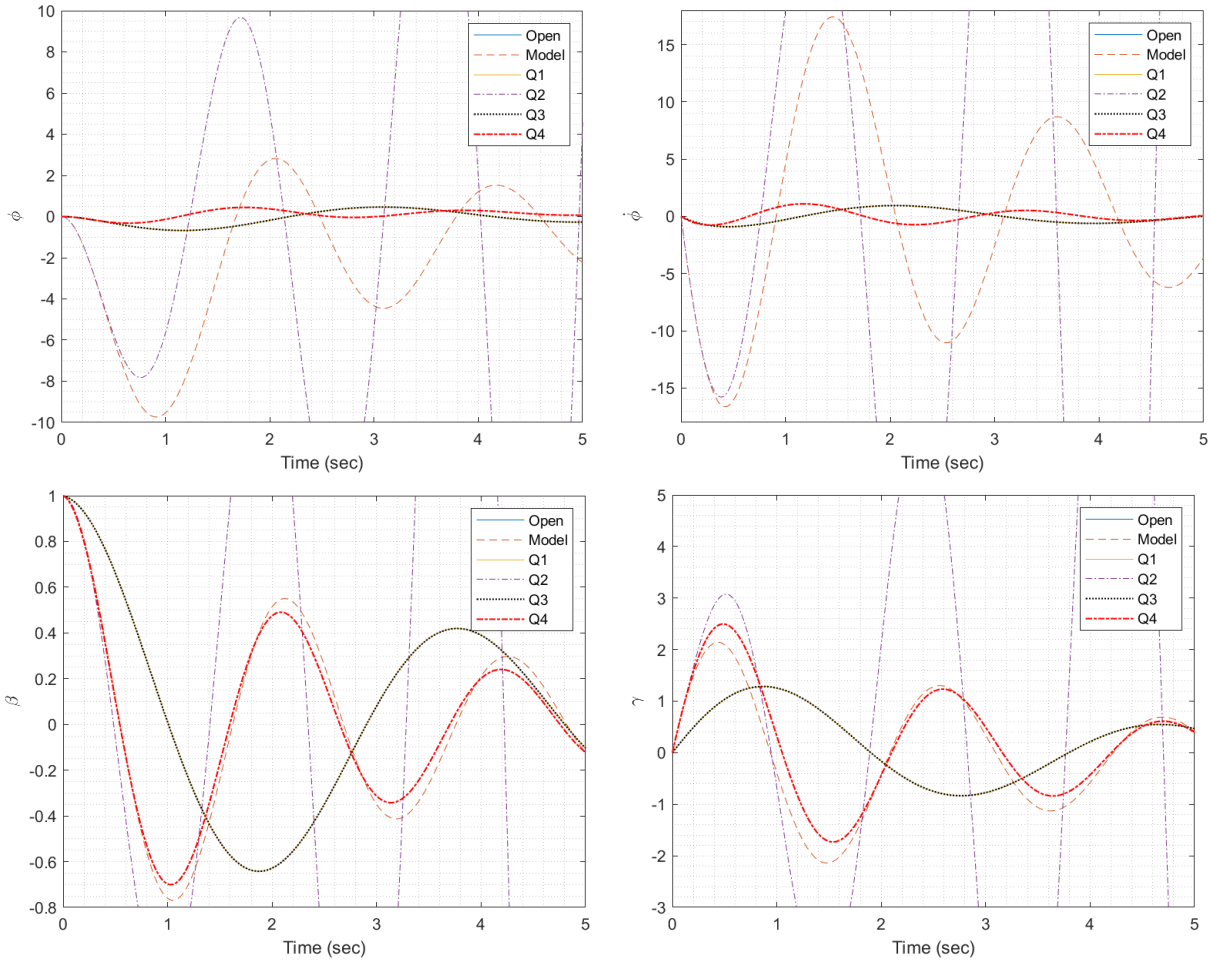

Fig. 3 B-26 model-in-the-performance-index response to initial sideslip angle (β) with Q only on one element

From the simulation results we can see that $Q1$ ($Q(6,0,0,0)$) and $Q3$ ($Q(0,0,6,0)$) overlaps with the open-loop model, so I think there is not much to do with the control for the first and third element of Q. We can also see that $Q4$ ($Q(0,0,0,6)$) is the most similar to the Model in β and γ, but doesn't perform well in ϕ and $\dot{\phi}$. Last, we can see that $Q2$ ($Q(0,6,0,0)$) diverges in every response, but it may not be a bad thing, because in ϕ and $\dot{\phi}$ by combining $Q2$ and $Q4$, the response might be greater than using $Q4$ only. The author explained this result is due to the first and third rows of the model and the system matrices were identical, since the control K_m depends on the difference between the system and the model $(F-L)$, the product $\hat{G}'H'Q(HF-LH)$ is independent on q_{11}, q_{33} and depends on q_{22}, q_{44} instead.

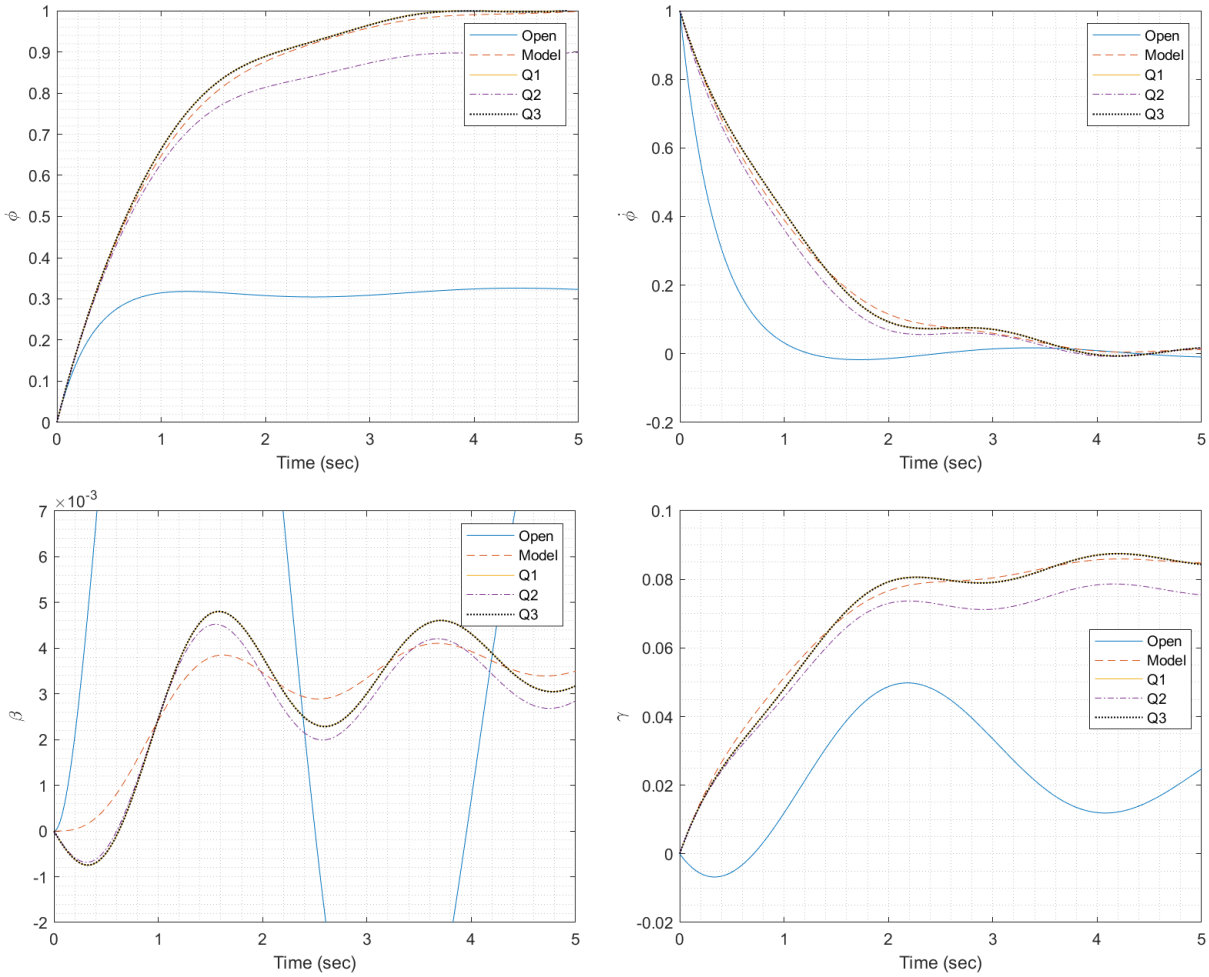

Fig. 4 The most similar result to the paper with $Q2 = Q(0, 1, 0, 2)$
(P.S. The 1 in $Q2$ means 1, instead of 10 here)

After the simulation above, I have tried to make the response more similar to $Q(1, 1, 1, 1)$ in the paper, tested with multiple different parameters, I found out that $Q(0, 1, 0, 2)$ is the most similar to the result in the paper, but the 1 in q_{22} here is 1 instead of 10 by the definition of Q. In the paper, the author mentioned that $Q(0, 6, 0, 6)$ were perfectly reasonable for the system, which also fits the conclusion from my simulation.

1.4 Model-in-The-System

In the previous method, feedback were placed around the plant to mimic the bahavior of the model, but no effort is made to make the system response similar to that of the model for the same input. In optimal-control theory, constant feedback gains are generated for each **plant** state variable to improve the plant dynamics, also constant feedforward gains are produced for each of the **model** state variable to provide a lead network. If

enough weight is placed on the error term in the performance index, the feedback system around the plant will be fast enough to follow the output of the model.
The equations of the plant are

$$\dot{\bar{x}}_p = F_p\bar{x}_p + G_p\bar{u}_p \tag{1.12}$$

The model dynamics are defined as

$$\dot{\bar{x}}_m = L\bar{x}_m \tag{1.13}$$

The state equation is now defined as a combination of the plant and the model equations

$$\begin{bmatrix}\dot{\bar{x}}_m \\ \dot{\bar{x}}_p\end{bmatrix} = \begin{bmatrix}L & [0] \\ [0] & F_p\end{bmatrix}\begin{bmatrix}\bar{x}_m \\ \bar{x}_p\end{bmatrix} + \begin{bmatrix}[0] \\ G_p\end{bmatrix}\begin{bmatrix}\bar{u}_p\end{bmatrix}$$

1.5 Model-in-The-System Technique Applied to Airplane Longitudinal Equations

	$Q(1,1,1,1)$				$Q(0,6,0,6)$			
Closed-Loop Matrix $F-GK$	0	1	0	0	0	1	0	0
	−12.33	−13.60	−2.064	1.040	−6.3	−3905	−.1	188.6
	.0941	.0078	−.1915	−.8856	.1	1.7	−.1	34
	.3895	.2419	8.270	−8.677	.3	187.6	0	−2542
Closed-Loop Roots	−1.019	−1.122	−7.660	−12.66	−3931	−2517	0	0.1
Feedback Control Matrices K	−.2324	−.2221	2.329	−3.267	−.0624	−48.11	−.8611	−998.7
	−3.153	−2.728	−.6869	−.0666	−1.603	−998	1.198	48.03

Table IV. Model-In-The-System Results

The paper mentioned the control results from $Q(1,1,1,1)$ proved to be a reasonable design, most of the values in the feedback control matrices weren't too big, but if we change q_{22} we can lower -2.728 in K to -1.917, and the new feedback control matrix will become

$$K = \begin{bmatrix}-.3070 & -.1939 & 2.316 & -3.262 \\ -3.143 & -1.917 & .6080 & .1085\end{bmatrix}$$

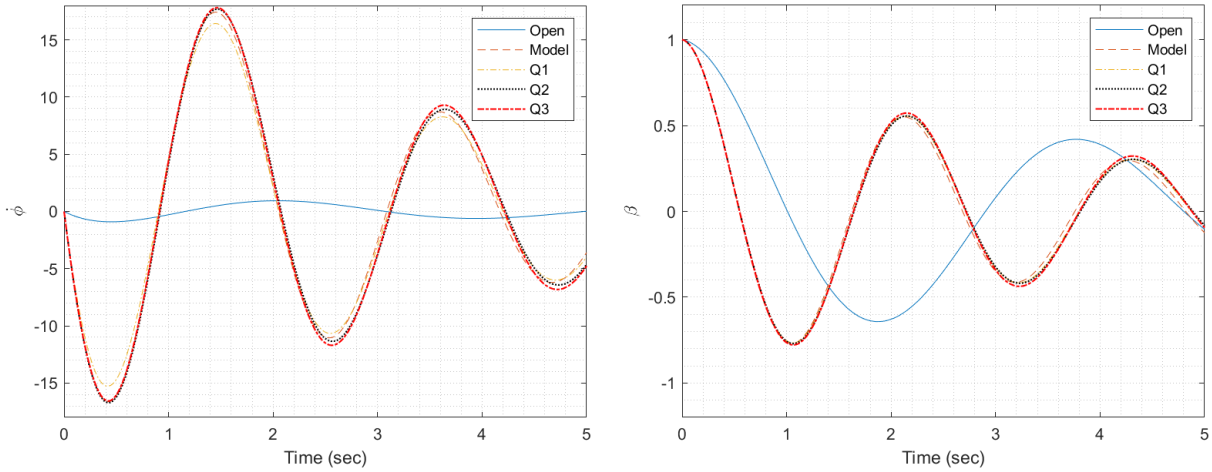

Fig. 5 B-26 model-in-the-system responses to an initial sideslip angle

P.S. $Q1 = Q(0,1,0,2)$, $Q2 = Q(2,2,2,2)$, $Q3 = Q(1,1,1,1)$, $Q(0,1,0,2)$ looks more similar to $Q(1,1,1,1)$ in the paper here too.

2 Review: The Use of a Quadratic Performance Index to Design Multivariable Control Systems

At 1966 very little optimal control theory has been applied on multivariable control systems. Lack of applications is due to the difficulty of defining a performance index. In realistic, we obtain numerical methods that yield solutions to a particular problem. However, it is desired to obtain a closed-form solution for the control, therefore simple performance indexes must be used which do not specify many design requirements. Thus, we have to compromise between the realistic criterion and mathematically tractable.

In this paper, the quadratic performance index is considered as a generalized criterion for designing linear multivariable systems. The advantages of using this performance are 1) the results will be a **closed-form** solution, therefore, the properties of the control and the optimal system can be determined; 2) under reasonable restrictions [5] it always produces a stable system; 3) once the numerical elements of the performance index are specified, the optimal feedback gains can be determined by a straightforward computer solution; and 4) this index results in a class of multivariable systems that satisfy a number of well-known criteria.

In this paper the weighting matrix elements can be selected according to well-known criteria on the transient response of the optimal system.

2.1 The Quadratic Performance Index and the Optimal Control Function

The plant is defined as

$$\dot{\bar{x}} = \bar{x} + G\bar{u}, \tag{2.1}$$

where $\bar{x}$ is the nth-order state vector, $\bar{u}$ is the control vector, F is the plant matrix, and G is the input matrix. The output is defined by

$$\bar{y} = H\bar{x}, \tag{2.2}$$

where $\bar{y}$ is the output vector and H is the output matrix. The generalized quadratic performance index considered in this paper is

$$V = \int_0^T (\bar{x}'H'QH\bar{x} + \bar{u}'\bar{u})dt, \tag{2.3}$$

where Q is a diagonal non-negative definite matrix.

The necessary conditions for the optimization of V are the solution of the canonical equations

$$\begin{bmatrix} \dot{\bar{x}} \\ \dot{\bar{\lambda}} \end{bmatrix} = \begin{bmatrix} F & -GG' \\ -HQ'H & -F' \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{\lambda} \end{bmatrix} = F_c \begin{bmatrix} \bar{x} \\ \bar{\lambda} \end{bmatrix}, \tag{2.4}$$

where $\bar{\lambda}$ is an nth-order vector referred to as the adjoint, costate, or Lagrange multiplier vector. The form of the optimal control function is

$$\bar{u}^0 = -G'P(t)\bar{x}, \tag{2.5}$$

where $P(t)$ is a time-varying, symmetric matrix which is found by

$$P(t) = [\Phi_{21}(t) + \Phi_{22}(t)P_0][\Phi_{11}(t) + \Phi_{12}(t)P_0]^{-1}. \tag{2.6}$$

$\Phi(t)$ is the transition matrix of the canonical system matrix F_c, and the Φ_{ij} elements are $n \times n$ submatrices of $\Phi(t)$. To simplify the analysis, only the steady-state value of $P(t)$, $P(\infty)$ will be used. With this, equation 2.5 and equation 2.1 can be rewritten as

$$\bar{u}^0 = -K\bar{x}, \quad \dot{\bar{x}} = [F - GK]\bar{x}.$$

Since $P(\infty)$ does not depend on P_0, $P(\infty)$ can be computed by equation 2.6 with $P_0 = 0$,

$$P(t) = \Phi_{21}(t)[\Phi_{11}(t)]^{-1}$$

2.2 The Design of an Optimal System on the Basis of its Transient Response

It is desired to express the characteristic equation of the optimal system as an explicit function of the elements of Q. The characteristic equation of the optimal system is defined by

$$|sI - F + GK| = 0 = \prod_{i=0}^{n}(s - \alpha_i), \tag{2.7}$$

It is known that the $2n$ eigenvalues of F_c in (2.4) consist of the eigenvalues of the optimal system $[F - GK]$ and their mirror image about the imaginary axis of the $s-$plane. Thus the characteristic equation for F_c can be written as

$$|sI - F_c| = (-1)^n \Delta(s)\Delta(-s) = \prod_{i=1}^{n}(s - \alpha_i)(s + \alpha_i). \tag{2.8}$$

It is assumed that the system described by F, G and H is completely **controllable** and **observable**. This implies that the optimal system is **stable**. The characteristic equation of the canonical system matrix $|sI - F_c| = 0$ is obtained by first premultiplying $[sI - F_c]$ by the transformation [4]

$$T = \begin{bmatrix} (sI - F)^{-1} & 0 \\ (H'QH)(sI - F)^{-1} & I \end{bmatrix}. \tag{2.9}$$

This results in

$$T \times [sI - F_c] = \begin{bmatrix} (sI - F)^{-1} & 0 \\ (H'QH)(sI - F)^{-1} & I \end{bmatrix} \begin{bmatrix} sI - F & -GG' \\ -H'QH & sI + F' \end{bmatrix}$$

$$= \begin{bmatrix} I & -(sI - F)^{-1}(GG') \\ 0 & -H'QH(sI - F)^{-1}GG' + (sI + F') \end{bmatrix}. \tag{2.10}$$

Equation (2.8) can also be expressed in the form

$$|sI - F_c| = (-1)^n \Delta(s)\Delta(-s) = \bar{\Delta}(s) det[(sI + F') - H'QH\frac{\theta(s)}{\bar{\Delta}(s)}GG'], \tag{2.11}$$

where

$$\bar{\Delta}(s) = |sI - F|, \tag{2.12}$$

and $\theta(s)$ is the adjoint matrix of $(sI - F)$.

2.3 Numerical Example-The Design of an Optimal System with an Unstable Plant

Consider the system defined by the following matrices:

$$F = \begin{bmatrix} 1 & 2 & -3 \\ -4 & 5 & -.6 \\ 7 & 8 & -.9 \end{bmatrix}, \quad G = \begin{bmatrix} .1 & 0 & 2 \\ 0 & 10 & 0 \\ 0 & 0 & 100 \end{bmatrix},$$

$$H = I, \quad Q = \begin{bmatrix} q_{11} & 0 & 0 \\ 0 & q_{22} & 0 \\ 0 & 0 & q_{33} \end{bmatrix}. \tag{2.13}$$

The eigenvalues of F in (2.13) are

$$\alpha = 5.384, \quad -.1239 \pm j5.892, \tag{2.14}$$

Since one of the eigenvalue equals 5.384>0, so the system is initially unstable. In order to obtain a feedback control matrix K such that the system is stable and has a transient response that is well-damped.

To determine numerical elements of Q that will result in a satisfactory design, the determinant $|sI - F_c| = 8.987 * 10^4 * q_{11}s^2 - 1.59 * 10^6 * q_{22} - 1.363 * 10^6 * q_{33} - 8.892 * 10^6 * q_{11} * q_{22} - 3.388 * 10^6 * q_{11} * q_{33} - 7.412 * 10^5 * q_{22} * q_{33} - 1.955 * 10^6 * q_{11} - 4.01 * q_{11} * s^4 + 605.1 * q_{22} * s^2 - 100 * q_{22} * s^4 + 1.166 * 10^5 * q_{33} * s^2 - 10^4 * q_{33} * s^4 + 401 * q_{11} * q_{22} * s^2 + 100 * q_{11} * q_{33} * s^2 + 10^6 * q_{22} * q_{33} * s^2 - 10^4 * q_{11} * q_{22} * q_{33} - 778.6 * s^2 + 40.79 * s^4 + s^6 - 3.448 * 10^4$

Root-locus diagrams of the determinant above for variations in each of the three q_{ii} elements with the other elements set equal to zero are shown below.

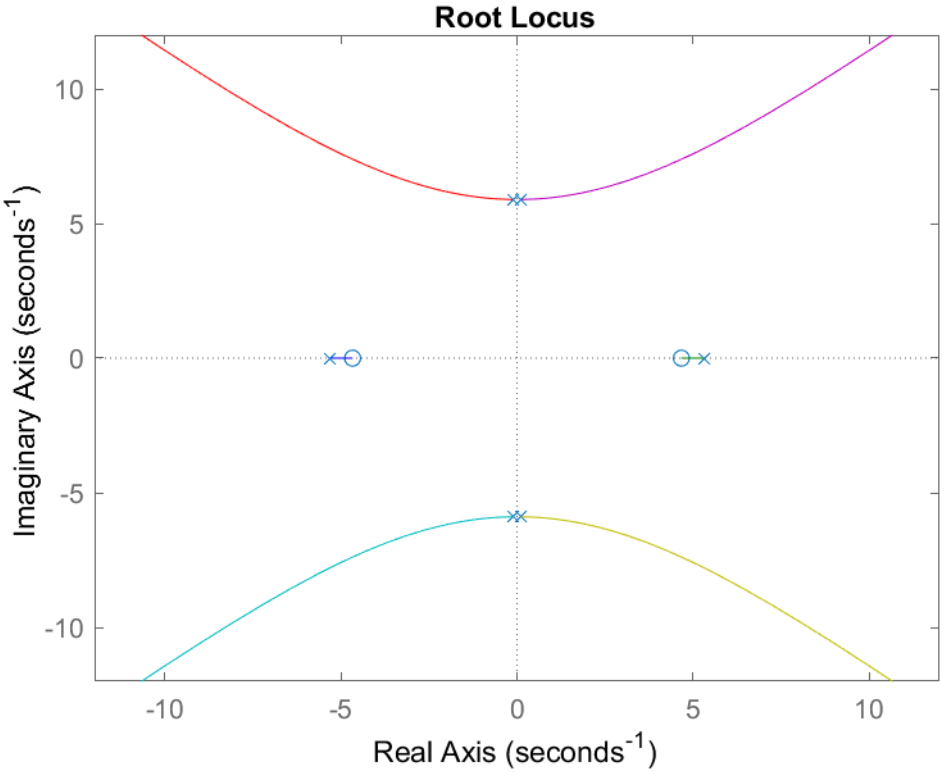

Fig. 6 Root loci of $|sI - F + GK| = 0$ for changes in q_{11} with $q_{22} = q_{33} = 0$. (Zeros=150, -150 haven't been plotted)

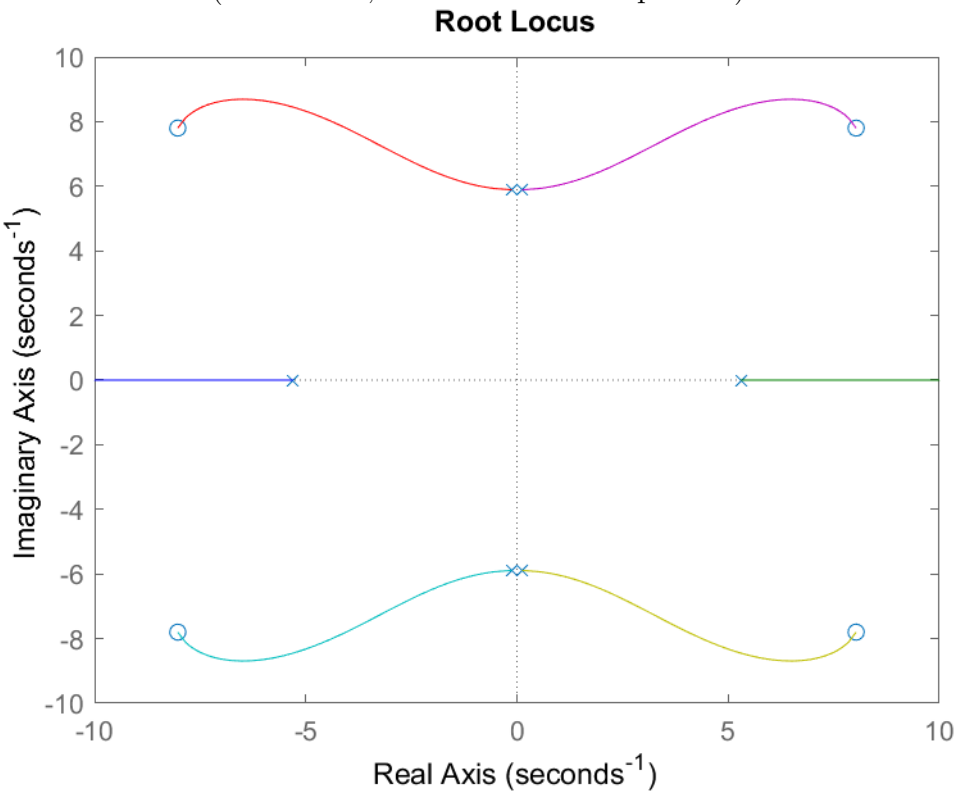

Fig. 7 Root loci of $|sI - F + GK| = 0$ for changes in q_{22} with $q_{11} = q_{33} = 0$.

(the rlocus plot is a little different from the paper)

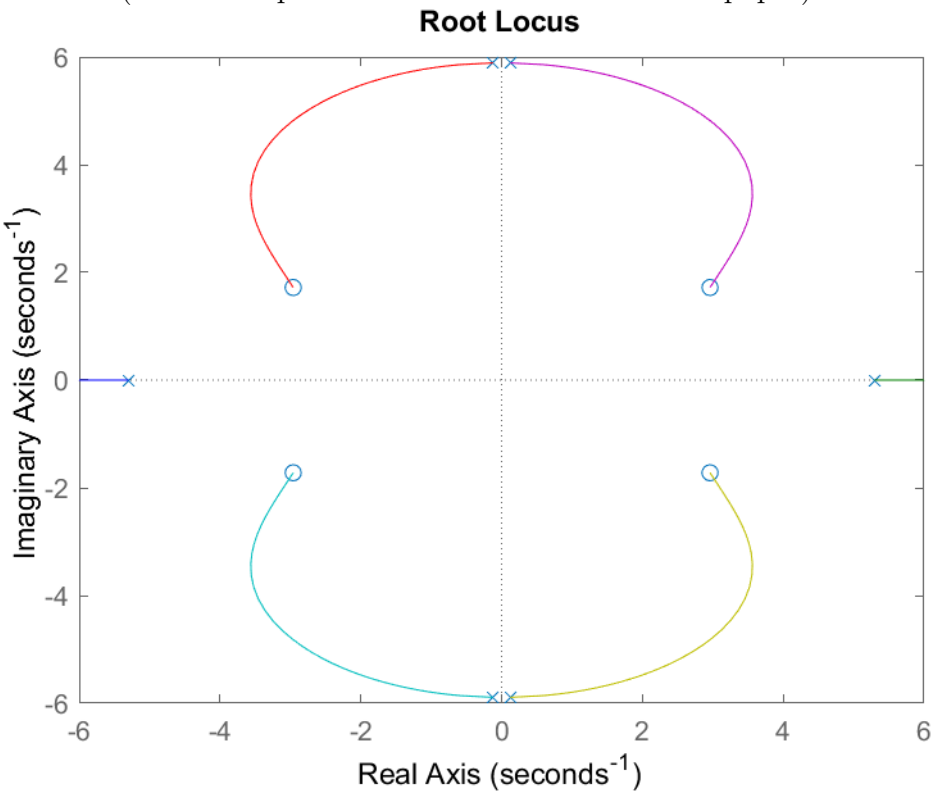

Fig. 8 Root loci of $|sI - F + GK| = 0$ for changes in q_{33} with $q_{11} = q_{22} = 0$.

After plotting root loci, the paper chose $Q = diag(0, 0, .125)$, then the control matrix K has been computed through equation 2.5-2.6. The result is

$$K = \begin{bmatrix} .008 & -.0042 & -.0006 \\ -.4188 & .7867 & .0305 \\ -.4796 & .2217 & .3893 \end{bmatrix},$$

and the eigenvalues are

$$\alpha = -34.61, \quad -3.063 \pm 1.922i.$$

A transient response for $[F - GK]$ has been computed for a step input to $u_1(t)$, in the paper x_2 is positive, but the result I got was negative. If I flipped x_2 they will be the same. We can also observe that the response of the system is not as damped as we expected, this is because the eigenvalues of the optimal system do not completely define its transient response and we can have a satisfied response, by changing q_{ii} of the Q matrix.

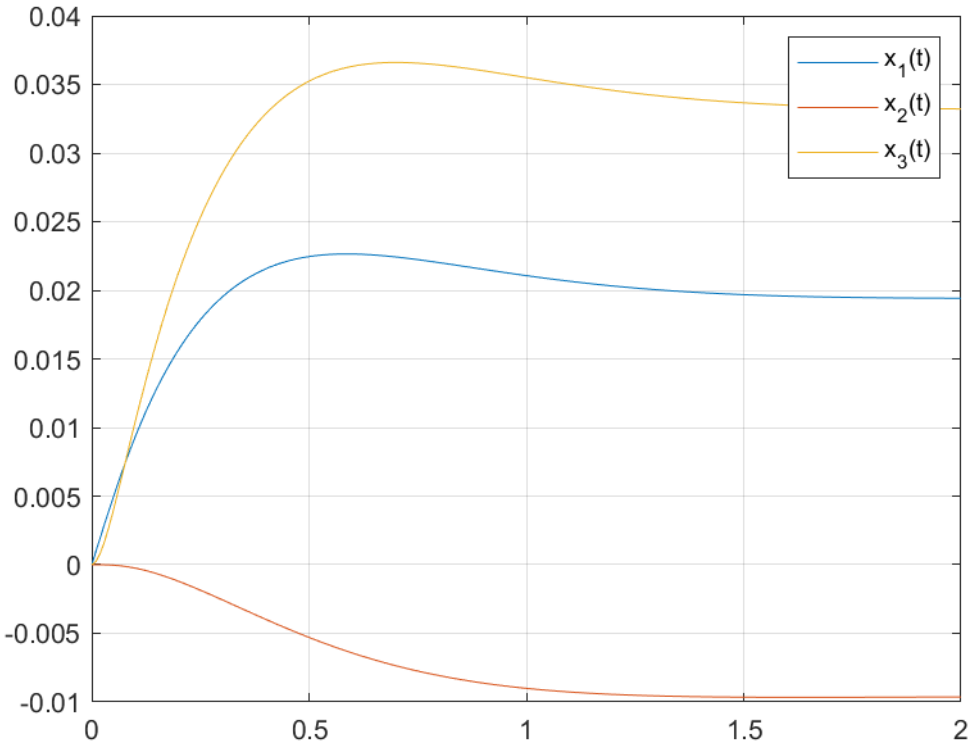

Fig. 9 Transient responses of $x_1(t)$, $x_2(t)$, and $x_3(t)$ to a step function on u_1.

2.4 The Complete Design of an Optimal System when the Design Criterion is a Linear Model

In this section, the design of a model-following system for a B-26 airplane is considered, and the model will be the same as the model in section 1.2. With Q defined as ρI, the expansion of $|sI - F_c|$

$$\begin{aligned}=&s^8 - (21.79\rho + 3.5)s^6 + (97.88\rho^2 + 16.77\rho - 40.08)s^4 \\ &- (197.26\rho^2 + 210.26\rho + 69.44)s^2 + (100.09\rho^2 + 228.57\rho)\end{aligned}$$

For large values of ρ, $|sI - F_c|$ can be simplify to

$$\approx s^8 - 21.79\rho s^6 + 97.88\rho^2 s^4 - 197.26\rho^2 s^2 + 100.09\rho^2$$

As ρ increases, the magnitude of s also increases, then the equation above can be further reduced to

$$\approx s^4(s^4 - 21.79\rho s^2 + 97.88\rho^2)$$

and has 4 non-zero roots

$$s \approx \pm 2.52\rho^{1/2}, \quad \pm 3.93\rho^{1/2}.$$

The eigenvalues of model L in the paper is also different from the result I got from Matlab.

$$\alpha_i = \quad -1.0233, \quad -.0003, \quad -.288 \pm j2.94$$

Q	K_{fb}				Eigenvalues of $F_p - G_p K_{fb}$
5, 5, 5, 5	$-.253$	$-.186$	1.58	-2.34	-.99, -1.3, -5.13, -9.14
	-2.21	-1.83	.7	.01	
5, 5, 5, 20	$-.201$	$-.185$	1.41	-4.42	-.663, -.908, -9.09, -11.2
	-2.23	-1.83	.614	.264	

Table V. Control Matrices and Eigenvalues for the B-26 Control System

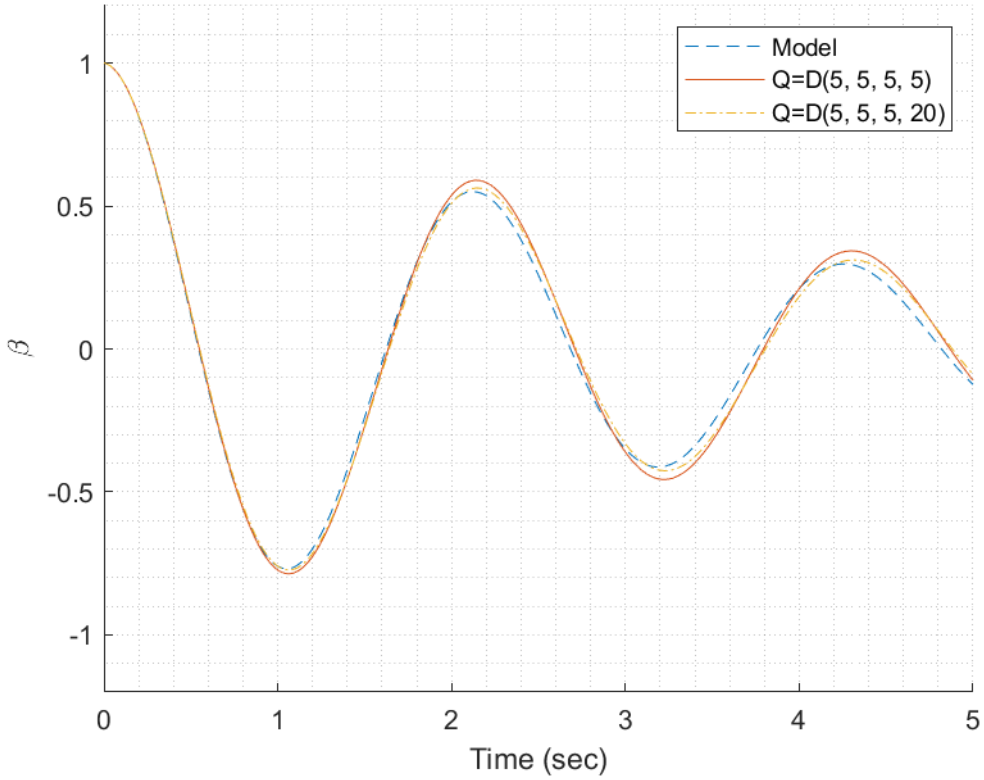

Figure 10. B-26 Transient Response of $\beta(t)$ with $\beta(0)$=1.

The result is pretty similar, but when $Q = D(5, 5, 5, 5)$, the curve didn't reach as low as it is in the paper. And I used the same code I wrote for the first paper to draw this figure.

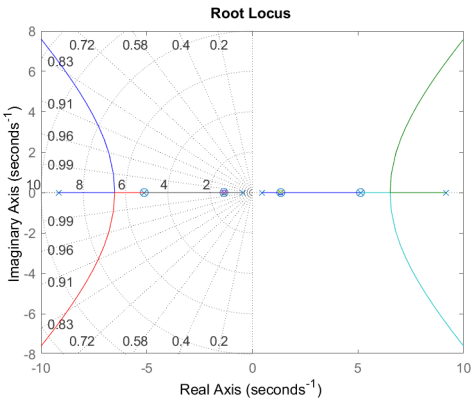

Figure 11. Root loci of $|sI - F + GK| = 0$ for changes in q_{11} with q_{22}, q_{33}, and $q_{44} = 5$.

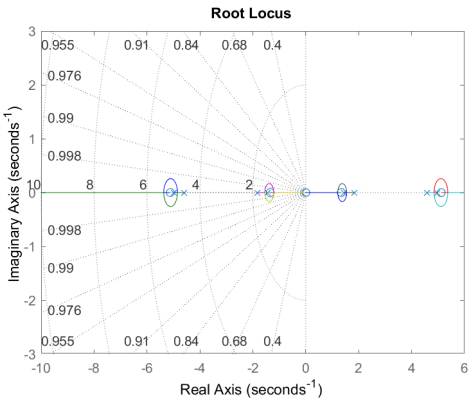

Figure 12. Root loci of $|sI - F + GK| = 0$ for changes in q_{22} with q_{11}, q_{33}, and $q_{44} = 5$.

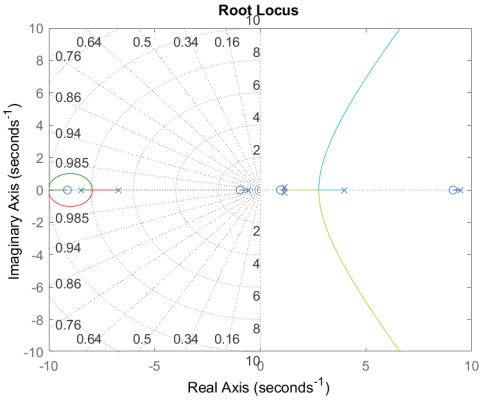

Figure 13. Root loci of $|sI - F + GK| = 0$ for changes in q_{33} with q_{11}, q_{22}, and $q_{44} = 5$.

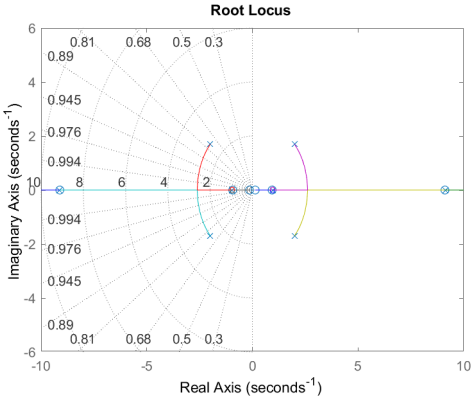

Figure 14. Root loci of $|sI - F + GK| = 0$ for changes in q_{44} with q_{11}, q_{22}, and $q_{33} = 5$.

Compared with the result in the paper, figure 11 and figure 14 are similar, but figure 12 and figure 13 differs a lot. In figure 12, there is a pole at about $s = -2.2 + j2$, but there isn't one in mine. And in figure 13, the root locus plot is totally flipped over. I think this is due to the different eigenvalues of the model L, so maybe the model L used by the author is wrong.

3 Review: Linear Quadratic Control Using Model-Free Reinforcement Learning

In this paper LQR is used without knowing the model of the system, and the system model is estimated by RL method. One possible RL method is dynamic programming, and valued by minimizing the the Bellman error [2]. Another class contains policy search algorithms where the policy is learned by directly optimizing the performance index.

3.1 LQ Problem

In this paper, they used another notation of the system, instead of F and G, A and B are used, and they also added process noise w_k and measurement noise v_k drawn from a Gaussian distribution $N(\mathbf{0}, W_w)$. The system can be written as:

$$x_{k+1} = Ax_k + Bu_k + w_k \tag{3.1}$$

$$y_k = x_k + v_k \tag{3.2}$$

The model (A, B) is unknown but stabilizable. The model order n is assumed to be known and the measurements y_k, u_k can be used to form the next input, while x_k is not measurable. And the cost is defined as:

$$r(y_k, u_k) = r_k = y_k^T R_y y_k + u_k^T R_u u_k \tag{3.3}$$

where it is also in quadratic term and $R_y \geq 0$ and $R_u > 0$ are the output and the control weighting matrices, respectively. It has the same definition as what we learned, but exchanging Q to R_y. In the paper, they designed stationary policies of the form $u_k = Ky_k$ to control the system in (3.1) and (3.2). Let $L := A + BK$. The policy gain K is stabilizing if $\rho(L) < 1$, and the *average cost associated with the policy* $u_k = Ky_k$ is defined by

$$\lambda(K) = \lim_{\tau \longrightarrow \infty} \frac{1}{\tau} \mathbf{E}\Big[\sum_{t=1}^{\tau} r(y_t, Ky_t) \Big] \tag{3.4}$$

and we define the *value function* associated with the given policy

$$V(y_k, K) = \mathbf{E}\Big[\sum_{t=k}^{+\infty} (r(y_t, Ky_t) - \lambda(K)) | y_k \Big] \tag{3.5}$$

And there exists three problems, *Problem 1:* Design the gain K^* such that the policy $K^* y_k$ minimizes the *value function. Problem 2 (LQR):* Design the controller $u_k = Kx_k$ to minimize the cost of the dynamical system. *Problem 3 (LQG):* Design the controller $u_k(y_1, ..., y_k)$ to minimize $\lambda(K)$

Consider *Problem 1*, from equation 3.1 we know that w_k, v_k and v_{k+1} has no effect on x_k, so we know that $\mathbf{E}[x_k^T \Xi w_k | y_k] = \mathbf{E}[x_k^T \Xi v_k | y_k] = \mathbf{E}[x_k^T \Xi v_{k+1} | y_k] = 0$, and $\mathbf{E}[x_k^T \Xi x_k^T | y_k] = \mathbf{E}[(y_k - v_k)^T \Xi (y_k - v_k)] = y_k^T \Xi y_k + Tr(\Xi W_v)$. Then by using the knowledge from class instead of substituting $x^T P^* x$ as V^*, here we substitute V^* as $y_k^T P^* y_k$ to

be the optimal control, then equation (3.5) can be rewritten as:

$$y_k^T P^* y_k = y_k^T R_y y_k + y_k^T K^{*T} R_u K^* y_k - \lambda(K^*) + \mathbf{E}[y_{k+1}^T P^* y_{k+1} | y_k]. \tag{3.6}$$

The term $\mathbf{E}[y_{k+1}^T P^* y_{k+1}|y_k]$ can be turned into (I don't think there should be v_k in it)

$$\begin{aligned}
\mathbf{E}[y_{k+1}^T P^* y_{k+1}|y_k] &= \mathbf{E}[(x_{k+1} + v_{k+1})^T P^* (x_{k+1} + v_{k+1})|y_k] \\
&= \mathbf{E}[(Ax_k + Bu_k + w_k + v_{k+1})^T P^* (Ax_k + Bu_k + w_k + v_{k+1})|y_k] \\
&= \mathbf{E}[x_k^T (A + BK^*)^T P^* x_k (A + BK^*)|y_k] + 2\mathbf{E}[x_k^T (A + BK^*)^T P^* v_k | y_k] \\
&\quad + 2\mathbf{E}[x_k^T (A + BK^*)^T P^* (w_k + v_{k+1})|y_k] \\
&\quad + \mathbf{E}[v_k^T K^{*T} B^T P^* B K^* v_k | y_k] + 2\mathbf{E}[v_k^T K^{*T} B^T P^* (w_k + v_{k+1})|y_k] \\
&\quad + \mathbf{E}[(w_k + v_{k+1})^T P^* (w_k + v_{k+1})|y_k] \\
&= y_k^T (A + BK^*)^T P^* (A + BK^*) y_k + Tr(K^{*T} B^T P^* B K^* W_v) \\
&\quad + Tr(P^* W_w) - Tr((A + BK^*)^T P^* (A + BK^*) W_v) + Tr(P^* W_v)
\end{aligned}$$

Then by substituting the result above into equation (3.6) and matching the terms $y_k^T(...)y_k$

$$(A + BK^*)^T P^* (A + BK^*) - P^* + R_y + K^{*T} R_u K^* = 0 \tag{3.7}$$

We can do partial derivative to the equation above with respect to K^*, we'll get

$$\begin{aligned}
\frac{\partial\{\}}{\partial K^*} &= 2R_u K^* + 2B^T P^* (A + BK^*) = 0 \\
\implies K^* &= -(R_u + B^T P^* B)^{-1} B^T P^* A
\end{aligned} \tag{3.8}$$

Then substitute K^* back into equation (3.7), we'll get the ARE

$$\begin{aligned}
&A^T P^* A + K^{*T} B^T P^* B K^* + 2K^{*T} B^T P^* A - P^* + R_y + K^{*T} R_u K^* \\
&= A^T P^* A - P^* + K^{*T} (R_u + B^T P^* B) K^* + 2K^{*T} B^T P^* A + R_y \\
&\implies A^T P^* A - P^* - A^T P^* B (B^T P^* B + R_u)^{-1} B^T P^* A + R_y = 0
\end{aligned}$$

3.1.1 Iterative Model-Based Approach for the LQ Problem

By using the solution solved from the ARE, if the dynamics of the system is known, we can find the optimal gain K_{LQR} by multiple iterations, the algorithm will be like:

Algorithm 1: Model-Based LQ[3]

1: **Initialize:** Select a stabilizing policy gain K^1, set $i = 1$

2: **for** $i = 1, ...,$**N do**

3: Find P^i from *the model-based Bellman equation*
$(A + BK^i)^T P^i (A + BK^i) - P^i + R_y + K^{iT} R_u K^i = 0.$

4: Update the policy gain
$K^{i+1} = -(R_u + B^T P^i B)^{-1} B^T P^i A.$

5: **end for**

In each iteration, P is calculated using equation (3.8), then P is used to update the policy gain K by using equation (3.8). And it have been proved in[3] that after multiple iterations, K^i will converge to K_{LQR}.

3.1.2 Proposed Model-Free Algorithms

For all proposed methods, they all initial the algorithm using a stabilizing policy gain K^1, then the policy is evaluated to find the value function and average cost function. After that, the value function and average cost function is used to perform policy improvement to improve the policy gain K^i.

1) Average Off-Policy Learning

The algorithm iterates N times over the policy evaluation and the policy improvement steps. Then the policy gain K^i is evaluated by estimating its associated value function (equation (3.5)) and average cost (equation (3.4)). To estimate P^i and $\lambda^i = \lambda(K^i)$, $K^i y_k$ is executed and collected τ samples of the output y_k, then the samples were used to estimate λ^i

$$\bar{\lambda}^i = \frac{1}{\tau} \sum_{t=1}^{\tau} r_t. \tag{3.9}$$

The average-cost estimator of P^i is given by[9]

$$vecs(\hat{P}^i) = \left(\sum_{t=1}^{\tau} \Phi_t (\Phi_t - \Phi_{t+1}) \right)^{-1} \left(\sum_{t=1}^{\tau} \Phi_t (r_t - \bar{\lambda}^i) \right), \tag{3.10}$$

where vecs(P)=$[p_{11}, ..., p_{1n}, p_{22}, ..., p_{2n}, ..., p_{nn}]^T$, $r_t = r(y_t, K^i y_t)$
Φ_k = vecv$(y_k) = [v_1^2, 2v_1v_2, ..., 2v_1v_n, v_2^2, .., 2v_2v_n, ..., v_n^2]^T$.

Then we have to improve the policy gain, the idea is to have a behavioral policy u_k while we are heading toward the target (optimal) policy $K^i y_k$. In this paper the behavioral policy is defined as $u_k = K^i y_k + \eta_k$ where η_k is sampled from $N(\mathbf{0}, W_\eta)$ at each time k and W_η is the covariance of probing.

Algorithm 2: Average Off-Policy Learning.

1: **Initialize:** Select a stabilizing policy gain K^1, set $i = 1$

2: **for** $i = 1, ...,$**N do**

3: Execute $K^i y_k$ for τ rounds and estimate $\hat{P}^i$, $\bar{\lambda}^i$ from equation (3.9) and (3.10)

4: **for** $t = 1, ..., \tau'$ **do**
Execute K_y for τ'' rounds and observe y.
Sample $\eta \sim N(\mathbf{0}, W_\eta)$ and set $u = Ky + \eta$.
Take u and observe y_+.
Add (y, u, y_+) to Z.
end for

5: Update the policy gain K^{i+1} using Z

6: **end for**

2) Average Q-Learning

The second proposed method is average Q-learning, extended from the model-free linear quadratic (MFLQ) control[1]. Different from average off-policy learning, the policy is evaluated by the Q function. In this method, the policy gain K^i is evaluated by finding the quadratic kernel G^i of the Q function and the average cost λ^i using equation (3.9) too. The quadratic kernel G^i is estimated using τ' sample points using

$$\text{vecs}(\hat{G}^i) = \left(\sum_{t=1}^{\tau'} \Psi_t \Psi_t^T \right)^{-1} \left(\sum_{t=1}^{\tau'} \Psi_t c_t' \right), \tag{3.11}$$

where $\Psi_t = \text{vecv}(z_t)$, $z_t = [y_k^T, u_k^T]^T$, $c_k' = r(y_k, u_k) - \bar{\lambda}^i + y_{k+1}^T \hat{P}^i y_{k+1}$.

Then the improved policy is chosen to be greedy with respect to the average of all previously estimated Q functions

$$K^{i+1} = \arg\min_a \frac{1}{i} \sum_{j=1}^{i} \hat{Q}^j(y_k, a) = \sum_{j=1}^{i} -(\hat{G}_{22}^j)^{-1} \hat{G}_{12}^{jT}. \tag{3.12}$$

Algorithm 3: Average Q-Learning.

1: **Initialize:** Select a stabilizing policy gain K^1, set $i = 1$

2: **for** $i = 1, ...,$**N do**

3: Execute $K^i y_k$ for τ rounds and estimate $\hat{P}^i$, $\bar{\lambda}^i$ from equation (3.9) and (3.10)

4: **for** $t = 1, ..., \tau'$ **do**
Execute K_y for τ'' rounds and observe y.
Sample $\eta \sim N(\mathbf{0}, W_\eta)$ and set $u = Ky + \eta$.
Take u and observe y_+.
Add (y, u, y_+) to Z.
end for

5: Estimate $\hat{G}^i$ using Z

6: Update the policy gain K^{i+1} using equation (3.12)

7: **end for**

3.2 Simulation Results

The system is set by using the parameters below

$$x_{k+1} = \begin{bmatrix} 1.01 & 0.01 & 0 \\ 0.01 & 1.01 & 0.01 \\ 0 & 0.01 & 1.01 \end{bmatrix} x_k + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} u_k + w_k$$

$$y_k = x_k + v_k$$

and the quadratic running cost is

$$r(y_k, u_k) = 0.001 y_k^T y_k + u_k^T u_k.$$

In this paper, the author simulated the system and made a table about the percentage of stability in all iterations in 100 simulations, before looking into the table, I guessed that model-building approach will be the most stable, and the model-building approach failed non of the simulations. And the model-building approach has the least relative average cost error too. Sorry that I had no time to simulate the methods proposed in this paper.

References

[1] Yasin Abbasi-Yadkori, Nevena Lazic, and Csaba Szepesvári. Model-free linear quadratic control via reduction to expert prediction. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3108–3117. PMLR, 2019.

[2] Dimitri Bertsekas. *Reinforcement learning and optimal control*. Athena Scientific, 2019.

[3] Gary Hewer. An iterative technique for the computation of the steady state gains for the discrete optimal regulator. *IEEE Transactions on Automatic Control*, 16(4):382–384, 1971.

[4] AS Householder. Matrix calculus. e. bodewig. north-holland, amsterdam; interscience, new york, 1956. 334 pp. $7.50. *Science*, 126(3270):410–410, 1957.

[5] R. E. Kalman. When is a linear control system optimal? 1964.

[6] R. E. Kalman, T. S. Englar, and R. S. Bucy. Fundamental study of adaptive control systems. Technical report, MARTIN MARIETTA CORP BALTIMORE MD RESEARCH INST FOR ADVANCED STUDIES, 1962.

[7] E. G. Rynaski, P. A. Reynolds, and W. H. Shed. Design of linear flight control systems using optimal control theory. Technical report, CORNELL AERONAUTICAL LAB INC BUFFALO NY, 1964.

[8] F. T. SMITH. An introduction to the application of dynamic programming to linear control systems. Technical report, RAND CORP SANTA MONICA CALIF, 1963.

[9] Huizhen Yu and Dimitri P Bertsekas. Convergence results for some temporal difference methods based on least squares. *IEEE Transactions on Automatic Control*, 54(7):1515–1531, 2009.