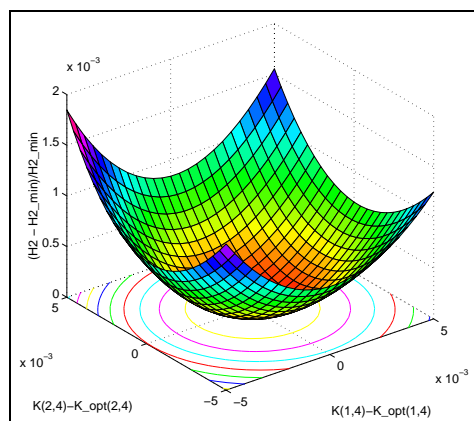


П. Хр. Петков

# МНОГОМЕРНИ СИСТЕМИ ЗА УПРАВЛЕНИЕ



София  
1997



# Съдържание

<b>ПРЕДГОВОР</b>	<b>5</b>
<b>1 ЛИНЕЙНИ СИСТЕМИ ЗА УПРАВЛЕНИЕ</b>	<b>7</b>
1.1 Описание на линейните системи за управление . . . . .	7
1.2 Управляемост и наблюдаемост . . . . .	9
1.3 Канонична декомпозиция на Калман . . . . .	12
1.4 Канонични форми . . . . .	15
1.5 Операции над системи . . . . .	22
1.6 Реализация на предавателни матрици . . . . .	23
1.7 Уравнения на Ляпунов . . . . .	25
1.8 Уравновесени реализации . . . . .	27
1.9 Скрити моди и съкращаване на полюси и нули . . . . .	31
1.10 Полюси и нули на многомерни системи . . . . .	33
<b>2 УСТОЙЧИВОСТ И КАЧЕСТВО</b>	<b>39</b>
2.1 Норми на сигнали и системи . . . . .	39
2.2 Многомерни системи с обратна връзка . . . . .	48
2.3 Свойства на обратната връзка . . . . .	54
2.4 Формиране честотните характеристики на контура . . . . .	58
2.5 Претеглено $\mathcal{H}_2$ и $\mathcal{H}_\infty$ качество . . . . .	62
<b>3 НЕОПРЕДЕЛЕНОСТ И РОБАСТНОСТ</b>	<b>67</b>
3.1 Представяне на неопределеността . . . . .	67
3.2 Робастност на устойчивостта . . . . .	74
3.3 Робастност на качеството . . . . .	79
3.4 $\mu$ -синтез и $DK$ -итерации . . . . .	83
<b>4 СИНТЕЗ ПРИ ЗАДАДЕНИ ПОЛЮСИ</b>	<b>87</b>
4.1 Задачи на синтеза . . . . .	87
4.2 Постановка на синтеза при зададени полюси . . . . .	90
4.3 Методи за синтез при зададени полюси . . . . .	94
4.4 Синтез на оптимална собствена структура . . . . .	101
4.5 Синтез на наблюдатели на състоянието . . . . .	108
4.6 Реализация на закони за управление с наблюдатели . . . . .	114
4.7 Синтез на дискретни наблюдатели . . . . .	119

<b>5</b>	<b><math>\mathcal{H}_2</math> ОПТИМАЛНО УПРАВЛЕНИЕ</b>	<b>121</b>
5.1	Линейно–квадратична оптимизация . . . . .	121
5.2	Робастни свойства . . . . .	125
5.3	Стандартна задача за $\mathcal{H}_2$ оптимално управление . . . . .	129
5.4	LQG: специално $\mathcal{H}_2$ управляващо устройство . . . . .	134
<b>6</b>	<b><math>\mathcal{H}_\infty</math> ОПТИМАЛНО УПРАВЛЕНИЕ</b>	<b>137</b>
6.1	Определение за $\mathcal{H}_\infty$ оптимално управление . . . . .	137
6.2	Задачи за $\mathcal{H}_\infty$ оптимално управление . . . . .	140
6.3	Решение на стандартната задача . . . . .	146
6.4	Пример за синтез при смесена чувствителност . . . . .	152
<b>7</b>	<b>ЧИСЛЕНИ МЕТОДИ</b>	<b>159</b>
7.1	Грешки от закръгляване . . . . .	159
7.2	Обусловеност и числена устойчивост . . . . .	163
7.3	Решаване на системи линейни уравнения . . . . .	169
7.4	Елементарни ортогонални преобразувания . . . . .	173
7.5	QR–декомпозиция . . . . .	176
7.6	Изчисляване на собствени стойности . . . . .	178
7.7	Изчисляване на обобщени собствени стойности . . . . .	186
7.8	Решаване на линейни матрични уравнения . . . . .	187
7.9	Привеждане в ортогонална канонична форма . . . . .	194
7.10	Синтез при зададена собствена структура . . . . .	196
7.11	Решаване на матричното уравнение на Рикати . . . . .	199
	<b>ПРИЛОЖЕНИЕ. МАТРИЧЕН АНАЛИЗ</b>	<b>207</b>
	П.1 Множества . . . . .	207
	П.2 Вектори и матрици . . . . .	208
	П.3 Линейни подпространства . . . . .	209
	П.4 Собствени стойности и собствени вектори . . . . .	210
	П.5 Декомпозиция по сингулярни числа . . . . .	214
	П.6 Векторни и матрични норми . . . . .	221
	П.7 Линейни дробни преобразувания . . . . .	228
	<b>БИБЛИОГРАФИЯ</b>	<b>231</b>

# ПРЕДГОВОР

След 1980 г. теорията на многомерните системи за управление е в процес на съществено развитие, при което се наблюдава възвръщане на интереса към честотните методи за анализ и синтез, използвани в класическата теория на управлението. Причина за този интерес от една страна е възможността за ясна физическа интерпретация на изискванията към качеството на системата в честотната област, а от друга страна – ограничените възможности на апарата на пространство на състоянията по отношение намирането на закони за управление, които да осигурят устойчивост и желано качество в присъствието на параметрична и структурирана неопределеност (иначе казано робастност) на затворената система. Съществен момент тук е, че независимо от по-високото ниво на използвания математически апарат, новите методи за синтез се характеризират с по-широка приложимост, тъй като те отчитат много по-пълно реалните условия, в които работи системата за управление. Разбира се, "завръщането" на честотните методи се извършва на едно ново концептуално и математическо ниво, при което формулировката на задачата за синтез се осъществява в честотната област, но нейното числено решаване по същество става изцяло в пространство на състоянията, където има разработен богат инструментариум от числени методи и програмно осигуряване. По такъв начин е налице едно успешно съчетаване на методи от двете области, което дава възможност за решаване на нови задачи.

В настоящият учебник се прави опит да се представи на достъпно ниво съвременната теория на многомерните системи за управление, която се разглежда като органично съчетание на временни и честотни методи за анализ и синтез.

В гл.1 са изложени накратко някои основни резултати в теорията на линейните системи за управление, отнасящи се до устойчивостта, управляемостта, наблюдаемостта, каноничните форми и полюсите и нулите на многомерни системи за управление.

В гл. 2 се въвеждат нормите на сигнали и системи и се показва, че  $\mathcal{H}_2$ - и  $\mathcal{H}_\infty$ -нормите възникват по естествен начин като мерки на възможно най-лошото качество за редица класове от входни сигнали. Въвежда се стандартната структурна схема на многомерна система и се дефинират изискванията към нейното качество.

В гл. 3 се описват различни начини за въвеждане на неопределеността в модела на многомерната система, въз основа на които се намират условия за робастна устойчивост и робастно качество на затворената система.

В гл. 4 са изложени методи за синтез при зададени полюси и синтез на

наблюдатели на състоянието на многомерни системи. Разглежда се и по-общата задача за синтез на оптимална собствена структура.

В гл. 5 се разглежда линейно-квадратичната оптимизация на многомерни системи, представена като задача за минимизация на  $\mathcal{H}_2$ -нормата на затворената система.

В гл. 6 са представени съвременни методи за  $\mathcal{H}_\infty$ -оптимизация на многомерни системи, които дават възможност да се постигне както добро подтискане на смущаващите въздействия, така и робастност на затворената система в присъствието на неопределеност в модела.

В гл. 7 се разглеждат числени методи за решаване на основните задачи, възникващи при анализа и синтеза на многомерни системи за управление.

В Приложението са дадени необходимите сведения от матричния анализ и са въведени всички тематически означения, които се използват в текста.

Учебникът е предназначен за студентите от специализация "Системи и управление" към специалността "Автоматика и системотехника" в Техническия университет-София, но може да бъде полезен и за специалисти, занимаващи се с проблемите на теория на управлението. Предполага се, че читателят е запознат с теорията на линейните системи за автоматично управление в обема на учебната дисциплина "Линейни системи за управление".

Набирането на текста е осъществено на  $\text{\LaTeX} 2_\epsilon$  с помощта на програмното осигуряване  $\text{\emTeX}$ , предназначено за операционната система MS-DOS. Всички примери са изчислени с помощта на програмната система MATLAB<sup>1</sup>, използвайки Control System Toolbox и Robust Control Toolbox. Фигурите са създадени с програмите MATLAB и  $\text{\TeX}cad$ .

Авторът изказва своята благодарност на рецензента проф. д.т.н. Илия Томов, който прочете внимателно ръкописа и направи редица препоръки, допринесли за подобряване на изложението.

---

<sup>1</sup>MATLAB е запазена марка на MathWorks, Inc.

# Глава 1

## ЛИНЕЙНИ СИСТЕМИ ЗА УПРАВЛЕНИЕ

В тази глава се прави преглед на някои основни резултати в теорията на линейните системи за управление, като се акцентира върху особеностите на многомерните системи. За пълнота на изложението най-напред се разглеждат накратко управляемостта, наблюдаемостта, стабилизируемостта и детектируемостта, както и декомпозицията на Калман и решенията на уравненията на Ляпунов. Въвеждат се канонични форми на многомерните системи и се изучават съединения на системи и реализации, като в частност се разглежда уравнивсената реализация на системата. Накрая се въвеждат концепциите за полюси и нули на многомерни системи. С цел опростяване на изложението повечето от резултатите са дадени без доказателства.

### 1.1 Описание на линейните системи за управление

Нека една крайномерна линейна стационарна система за управление е описана със следните линейни диференциални уравнения с постоянни коефициенти

$$\dot{x} = Ax + Bu, \quad x(t_0) = x_0 \quad (1.1)$$

$$y = Cx + Du \quad (1.2)$$

където  $x(t) \in \mathbb{R}^n$  се нарича *състояние* на системата,  $x(t_0)$  се нарича *начално условие* на системата,  $u(t) \in \mathbb{R}^m$  се нарича *вход* на системата и  $y(t) \in \mathbb{R}^r$  е *изходът* на системата.  $A, B, C, D$  са реални матрици с постоянни елементи, които имат съответните размерности. Система за управление с един вход ( $m = 1$ ) и един изход ( $r = 1$ ) се нарича SISO система (от английското Single Input Single Output System – система с един вход и един изход), в противен случай тя се нарича MIMO система (от английското Multi Input Multi Output System – система с много входове и много изходи, т.е. *многомерна система*). Съответната

предавателна матрица от  $u$  към  $y$  се определя от

$$Y(s) = G(s)U(s),$$

където  $U(s)$  и  $Y(s)$  са Лапласовите преобразувания на  $u(t)$  и  $y(t)$  при нулево начално условие ( $x(0) = 0$ ). Следователно, имаме че

$$G(s) = C(sI - A)^{-1}B + D.$$

За улесняване на изчисленията, включващи предавателни матрици, се използва означението

$$\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] \stackrel{\text{def}}{=} C(sI - A)^{-1}B + D.$$

Уравненията на системата (1.1) и (1.2) могат да се запишат в компактният вид

$$\begin{bmatrix} \dot{x} \\ y \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Обърнете внимание, че

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

е реална блочна матрица, а не предавателна матрица.

За дадено начално условие  $x(t_0)$  и вход  $u(t)$  динамичната реакция (преходния процес)  $x(t)$  и  $y(t)$  се определят по формулите

$$\begin{aligned} x(t) &= e^{A(t-t_0)}x(t_0) + \int_{t_0}^t e^{A(t-\tau)}Bu(\tau)d\tau \\ y(t) &= Cx(t) + Du(t). \end{aligned}$$

В случай, че  $u(t) = 0$ ,  $t \geq t_0$ , имаме че за всяко  $t_1 \geq t_0$  и  $t \geq t_0$  е изпълнено

$$x(t) = e^{A(t-t_1)}x(t_1).$$

Следователно, матричната функция  $\Phi(t, t_1) = e^{A(t-t_1)}$  действа като преобразуване от едно състояние в друго, поради което  $\Phi(t, t_1)$  се нарича *преходна матрица на състоянието*. Тъй като състоянието на линейна система в даден момент може да се получи от състоянието в друг момент чрез переходната матрица, може да се приеме без загуба на общност, че  $t_0 = 0$ .

Импулсната переходна матрица на линейна система се дефинира като

$$g(t) = \mathcal{L}^{-1}\{G(s)\} = Ce^{At}B1(t) + D\delta(t)$$

където  $\delta(t)$  е единичният импулс и  $1(t)$  е единичното въздействие, определено от

$$1(t) \stackrel{\text{def}}{=} \begin{cases} 1, & t \geq 0; \\ 0, & t < 0. \end{cases}$$

Входно-изходната връзка (при нулево начално състояние,  $x_0 = 0$ ) се описва с интеграла на свиването

$$y(t) = (g * u)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} g(t - \tau)u(\tau)d\tau = \int_{-\infty}^t g(t - \tau)u(\tau)d\tau.$$



## 1.2 Управление и наблюдаемост

**Определение 1.1** Линейната система, описана с уравнението (1.1), или двойката  $(A, B)$ , се нарича *управляема*, ако за всяко начално състояние  $x(0) = x_0$ ,  $t_1 > 0$  и крайно състояние  $x_1$  съществува (по-части непрекъснат) вход  $u(\cdot)$ , такъв че решението на (1.1) удовлетворява  $x(t_1) = x_1$ . В противен случай системата или двойката  $(A, B)$  се нарича *неуправляема*.

Управляемостта на система може да се провери чрез някои алгебрични или геометрични критерии.

**Теорема 1.1** Следните твърдения са еквивалентни:

(a)  $(A, B)$  е управляема.

(б) Матрицата

$$W_c(t) = \int_0^t e^{A\tau} B B^H e^{A^H \tau} d\tau$$

е положително определена за всяко  $t > 0$ .

(в) Матрицата на управляемостта

$$\mathcal{C} = [B \ AB \ A^2 B \ \dots \ A^{n-1} B]$$

има пълен редови ранг, т.е.

$$\text{rank } \mathcal{C} \stackrel{\text{def}}{=} \sum_{i=1}^n \text{rank}(A^{i-1} B) = n.$$

(г) Матрицата  $[A - \lambda I, B]$  има пълен редови ранг за всички  $\lambda \in \mathbb{C}$ .

(д) Нека  $\lambda$  и  $x$  е коя да е собствена стойност и съответния ляв собствен вектор на  $A$ , т.е.  $x^H A = x^H \lambda$ , тогава  $x^H B \neq 0$ .

(е) Собствените стойности на  $A + BF$  могат да се разположат произволно (при ограничението, че комплексните собствени стойности са в спрегнати двойки) посредством подходящ избор на  $F$ .

**Определение 1.2** Линейната система без външно въздействие  $\dot{x} = Ax$  се нарича *устойчива*, ако всички собствени стойности на  $A$  са в отворената лява комплексна полуравнина, т.е.  $\text{Re} \lambda(A) < 0$ . Матрица  $A$  с такова свойство се нарича *устойчива* или *Хурвицова*.

**Определение 1.3** Линейната система (1.1), или двойката  $(A, B)$ , се нарича *стабилизируема*, ако съществува обратна връзка по състоянието  $u = Fx$ , такава че системата е устойчива, т.е.  $A + BF$  е устойчива.

Следната теорема е следствие от теорема 1.1.

**Теорема 1.2** Следните твърдения са еквивалентни:

- (a)  $(A, B)$  е стабилизируема.
- (б) Матрицата  $[A - \lambda I, B]$  има пълен редови ранг за всички  $\operatorname{Re} \lambda \geq 0$ .
- (в) За всяко  $\lambda$  и  $x$ , такива че  $x^H A = x^H \lambda$  и  $\operatorname{Re} \lambda \geq 0$ ,  $x^H B \neq 0$ .
- (г) Съществува матрица  $F$ , таква че  $A + BF$  е устойчива матрица.

Дуалните понятия на управляемост и стабилизируемост за системата, описана с уравненията (1.1) и (1.2) са *наблюдаемост* и *детектируемост*.

**Определение 1.4** Линейната система, описана с уравненията (1.1) и (1.2) или с двойката  $(C, A)$ , се нарича *наблюдаема*, ако за всяко  $t_1 \geq 0$  началното състояние  $x(0) = x_0$  може да се определи от поведението на входа  $u(t)$  и изхода  $y(t)$  в интервала  $[0, t_1]$ . В противен случай системата или двойката  $(C, A)$  е *ненаблюдаема*.

**Теорема 1.3** Следните твърдения са еквивалентни:

- (a)  $(C, A)$  е наблюдаема.
- (б) Матрицата

$$W_o(t) = \int_0^t e^{A^H \tau} C^H C e^{A \tau} d\tau$$

е положително определена за всяко  $t > 0$ .

- (в) Матрицата на наблюдаемостта

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}$$

има пълен колонен ранг, т.е.

$$\bigcap_{i=1}^n \operatorname{Ker}(CA^{i-1}) = 0.$$

- (г) Матрицата  $\begin{bmatrix} A - \lambda I \\ C \end{bmatrix}$  има пълен колонен ранг за всички  $\lambda \in \mathbb{C}$ .

(д) Нека  $\lambda$  и  $y$  е коя да е собствена стойност и съответния десен собствен вектор на  $A$ , т.е.  $Ay = \lambda y$ , тогава  $Cy \neq 0$ .

(е) Собствените стойности на  $A+LC$  могат да се разположат произволно (при ограничението, че комплексните собствени стойности са в спрегнати двойки) посредством подходящ избор на  $L$ .

(ж) Двойката  $(A^H, C^H)$  е управляема.

**Определение 1.5** Системата, или двойката  $(C, A)$ , е *детектируема* ако  $A+LC$  е устойчива за някакво  $L$ .

**Теорема 1.4** Следните твърдения са еквивалентни:

(а)  $(C, A)$  е детектируема.

(б) Матрицата  $\begin{bmatrix} A - \lambda I \\ C \end{bmatrix}$  има пълен колонен ранг за всички  $\operatorname{Re} \lambda \geq 0$ .

(в) За всяко  $\lambda$  и  $y$ , такива че  $Ay = \lambda y$  и  $\operatorname{Re} \lambda \geq 0$ ,  $Cy \neq 0$ .

(г) Съществува матрица  $L$ , такава че  $A+LC$  е устойчива.

(ж) Двойката  $(A^H, C^H)$  е стабилизируема.

Условията (г) и (д) на Теорема 1.1 и Теорема 1.3 и условията (б) и (в) на Теорема 1.2 и Теорема 1.4 често се наричат тест на Попов-Белевиц-Хаутус (ПБХ – тест). В тази връзка се използват следните определения за модална управляемост и наблюдаемост.

**Определение 1.6** Нека  $\lambda$  е собствена стойност на  $A$  или, еквивалентно, мода на системата. Тогава модата  $\lambda$  се нарича *управляема* (наблюдаема), ако  $x^H B \neq 0$  ( $Cx \neq 0$ ) за *всички* леви (десни) собствени вектори на  $A$  съответстващи на  $\lambda$ , т.е.  $x^H A = x^H \lambda$  ( $Ax = \lambda x$ ) и  $0 \neq x \in \mathbb{C}^n$ . В противен случай модата се нарича *неуправляема* (ненаблюдаема).

От това определение следва, че системата е управляема (наблюдаема) ако и само ако всяка мода е управляема (наблюдаема). Подобно на това, системата е стабилизируема (детектируема), ако и само ако всяка неустойчива мода е управляема (наблюдаема).

**Пример 1.1** Да разгледаме следната система от 4-ти ред:

$$\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] = \left[ \begin{array}{cccc|c} \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 & 1 \\ 0 & 0 & \lambda_1 & 0 & \alpha \\ 0 & 0 & 0 & \lambda_2 & 1 \\ \hline 1 & 0 & 0 & \beta & 0 \end{array} \right]$$

с  $\lambda_1 \neq \lambda_2$ . Тогава модата  $\lambda_1$  не е управляема, ако  $\alpha = 0$  и  $\lambda_2$  не е наблюдаема, ако  $\beta = 0$ . Обърнете внимание, че ако  $\lambda_1 = \lambda_2$  системата е неуправляема и не-наблюдаема за всяко  $\alpha$  и  $\beta$ , тъй като в този случай и двата вектора

$$x_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \text{ и } x_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

са леви собствени вектори на  $A$ , съответстващи на  $\lambda_1$ . Следователно, всяка линейна комбинация на  $x_1$  и  $x_2$  е отново собствен вектор на  $A$ , съответстващ на  $\lambda_1$ . В частност, нека  $x = x_1 - \alpha x_2$ , тогава  $x^H B = 0$  и в резултат системата е неуправляема. Подобни аргументи могат да се приложат за проверка на наблюдаемостта. Ако обаче матрицата  $B$  се промени в  $4 \times 2$  матрица с последните два реда линейно независими, то системата е управляема даже ако  $\lambda_1 = \lambda_2$ . Например, системата с

$$B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ \alpha & 1 \\ 1 & 0 \end{bmatrix}$$

е управляема за всяко  $\alpha$ .

### 1.3 Канонична декомпозиция на Калман

Съществуват много координатни системи, в които може да се опише многомерна система за управление. Някои от тези координатни системи се предпочитат, тъй като в тях се получава по-естествено описание на физическата система. В някои случаи, обаче, използването на координати, различни от естественото описание, може да направи анализа и синтеза на системата по-лесен.

Нека  $T \in \mathbb{R}^{n \times n}$  е неособена матрица и да дефинираме

$$\tilde{x} = Tx.$$

Уравненията на системата (1.1) и (1.2) се преобразуват в

$$\begin{aligned} \dot{\tilde{x}} &= TAT^{-1}\tilde{x} + TBu \\ y &= CT^{-1}\tilde{x} + Du \end{aligned}$$

Тези уравнения описват същата линейна система за всяка неособена матрица  $T$  и следователно двете представяния можем да разглеждаме като еквивалентни. Лесно се вижда, че входно/изходната предавателна матрица не се променя при координатното преобразуване, т.е.

$$G(s) = C(sI - A)^{-1}B + D = CT^{-1}(sI - TAT^{-1})^{-1}TB + D.$$

В тази точка се разглеждат структурни декомпозиции използвайки координатни преобразувания при условие, че системата не е напълно управляема или напълно наблюдаема. За целта нека най-напред разгледаме две многомерни системи, свързани с преобразуване на подобие:

$$\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] \mapsto \left[ \begin{array}{c|c} \tilde{A} & \tilde{B} \\ \hline \tilde{C} & \tilde{D} \end{array} \right] = \left[ \begin{array}{c|c} TAT^{-1} & TB \\ \hline CT^{-1} & D \end{array} \right]$$

Матриците на управляемостта и наблюдаемостта на тези системи са свързани с

$$\tilde{\mathcal{C}} = T\mathcal{C}, \quad \tilde{\mathcal{O}} = \mathcal{O}T^{-1}.$$

От тези зависимости, както и от ПБХ-теста, получаваме следния резултат.

**Теорема 1.5** *Управляемостта (или стабилизируемостта) и наблюдаемостта (или детектируемостта) са инвариантни относно преобразувания на подобие.*

Използвайки този факт, може да се докаже следната теорема.

**Теорема 1.6** *Ако матрицата на управляемостта  $\mathcal{C}$  има ранг  $k_1 < n$ , то съществува преобразуване на подобие*

$$\bar{x} = \begin{bmatrix} \bar{x}_c \\ \bar{x}_{\bar{c}} \end{bmatrix} = Tx$$

такова че

$$\begin{bmatrix} \dot{\bar{x}}_c \\ \dot{\bar{x}}_{\bar{c}} \end{bmatrix} = \begin{bmatrix} \bar{A}_c & \bar{A}_{12} \\ 0 & \bar{A}_{\bar{c}} \end{bmatrix} \begin{bmatrix} \bar{x}_c \\ \bar{x}_{\bar{c}} \end{bmatrix} + \begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix} u,$$

$$y = \begin{bmatrix} \bar{C}_c & \bar{C}_{\bar{c}} \end{bmatrix} \begin{bmatrix} \bar{x}_c \\ \bar{x}_{\bar{c}} \end{bmatrix} + Du$$

с  $\bar{A}_c \in \mathbb{C}^{k_1 \times k_1}$  и  $(\bar{A}_c, \bar{B}_c)$  управляема. Освен това

$$G(s) = C(sI - A)^{-1}B + D = \bar{C}_c(sI - \bar{A}_c)^{-1}\bar{B}_c + D.$$

**Следствие 1.7** Ако системата е стабилизируема и матрицата на управляемостта има ранг  $k_1 < n$ , то съществува неособено преобразуване  $T$ , такова че

$$\left[ \begin{array}{c|c} TAT^{-1} & TB \\ \hline CT^{-1} & D \end{array} \right] = \left[ \begin{array}{cc|c} \bar{A}_c & \bar{A}_{12} & \bar{B}_c \\ 0 & \bar{A}_{\bar{c}} & 0 \\ \hline \bar{C}_c & \bar{C}_{\bar{c}} & D \end{array} \right]$$

с  $\bar{A}_c \in \mathbb{C}^{k_1 \times k_1}$ ,  $(\bar{A}_c, \bar{B}_c)$  управляема и  $\bar{A}_{\bar{c}}$  устойчива.

По такъв начин пространството на състоянията  $\bar{x}$  се разделя на две ортогонални подпространства

$$\left\{ \begin{bmatrix} \bar{x}_c \\ 0 \end{bmatrix} \right\} \quad \text{и} \quad \left\{ \begin{bmatrix} 0 \\ \bar{x}_{\bar{c}} \end{bmatrix} \right\}$$

като първото е управляемо от входа, а второто е напълно неуправляемо от входа (т.е. състоянието  $\bar{x}_{\bar{c}}$  не се влияе от управлението  $u$ ).

По силата на дуалността имаме следната декомпозиция, при условие, че системата не е напълно наблюдаема.

**Теорема 1.8** Ако матрицата на наблюдаемостта  $\mathcal{O}$  има ранг  $k_2 < n$ , то съществува преобразуване на подобие  $T$ , такова че

$$\left[ \begin{array}{c|c} TAT^{-1} & TB \\ \hline CT^{-1} & D \end{array} \right] = \left[ \begin{array}{cc|c} \bar{A}_o & 0 & \bar{B}_o \\ A_{21} & \bar{A}_{\bar{o}} & \bar{B}_{\bar{o}} \\ \hline \bar{C}_o & 0 & D \end{array} \right]$$

с  $\bar{A}_o \in \mathbb{C}^{k_2 \times k_2}$ ,  $(\bar{C}_o, \bar{A}_o)$  и наблюдаема.

**Следствие 1.9** Ако системата е детектируема и матрицата на наблюдаемостта има ранг  $k_2 < n$ , то съществува преобразуване на подобие  $T$ , такова че

$$\left[ \begin{array}{c|c} TAT^{-1} & TB \\ \hline CT^{-1} & D \end{array} \right] = \left[ \begin{array}{cc|c} \bar{A}_o & 0 & \bar{B}_o \\ A_{21} & \bar{A}_{\bar{o}} & \bar{B}_{\bar{o}} \\ \hline \bar{C}_o & 0 & D \end{array} \right]$$

с  $\bar{A}_o \in \mathbb{C}^{k_2 \times k_2}$ ,  $(\bar{C}_o, \bar{A}_o)$  наблюдаема и  $\bar{A}_{\bar{o}}$  устойчива.

По подобен начин имаме, че

$$G(s) = C(sI - A)^{-1}B + D = \bar{C}(sI - \bar{A})^{-1}\bar{B} + D.$$

Използвайки тези две теореми, получаваме каноничната декомпозиция на Калман.

**Теорема 1.10** Нека линейна стационарна система е описана с уравненията (1.1) и (1.2). Тогава съществува неособено координатно преобразуване  $\bar{x} =$

$Tx$ , такава че

$$\begin{bmatrix} \dot{\bar{x}}_{co} \\ \dot{\bar{x}}_{c\bar{o}} \\ \dot{\bar{x}}_{\bar{c}o} \\ \dot{\bar{x}}_{\bar{c}\bar{o}} \end{bmatrix} = \begin{bmatrix} \bar{A}_{co} & 0 & \bar{A}_{13} & 0 \\ \bar{A}_{21} & \bar{A}_{c\bar{o}} & \bar{A}_{23} & \bar{A}_{24} \\ 0 & 0 & \bar{A}_{\bar{c}o} & 0 \\ 0 & 0 & \bar{A}_{43} & \bar{A}_{\bar{c}\bar{o}} \end{bmatrix} \begin{bmatrix} \bar{x}_{co} \\ \bar{x}_{c\bar{o}} \\ \bar{x}_{\bar{c}o} \\ \bar{x}_{\bar{c}\bar{o}} \end{bmatrix} + \begin{bmatrix} \bar{B}_{co} \\ \bar{B}_{c\bar{o}} \\ 0 \\ 0 \end{bmatrix} u,$$

$$y = \begin{bmatrix} \bar{C}_{co} & 0 & \bar{C}_{\bar{c}o} & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_{co} \\ \bar{x}_{c\bar{o}} \\ \bar{x}_{\bar{c}o} \\ \bar{x}_{\bar{c}\bar{o}} \end{bmatrix} + Du$$

или еквивалентно

$$\left[ \begin{array}{c|c} TAT^{-1} & TB \\ \hline CT^{-1} & D \end{array} \right] = \left[ \begin{array}{cccc|c} \bar{A}_{co} & 0 & \bar{A}_{13} & 0 & \bar{B}_{co} \\ \bar{A}_{21} & \bar{A}_{c\bar{o}} & \bar{A}_{23} & \bar{A}_{24} & \bar{B}_{c\bar{o}} \\ 0 & 0 & \bar{A}_{\bar{c}o} & 0 & 0 \\ 0 & 0 & \bar{A}_{43} & \bar{A}_{\bar{c}\bar{o}} & 0 \\ \hline \bar{C}_{co} & 0 & \bar{C}_{\bar{c}o} & 0 & D \end{array} \right]$$

където векторът  $\bar{x}_{co}$  е управляем и наблюдаем,  $\bar{x}_{c\bar{o}}$  е управляем, но ненаблюдаем,  $\bar{x}_{\bar{c}o}$  е наблюдаем, но неуправляем и  $\bar{x}_{\bar{c}\bar{o}}$  е неуправляем и ненаблюдаем. Освен това, предавателната матрица от  $u$  към  $y$  се дава от

$$G(s) = \bar{C}_c(sI - \bar{A}_c)^{-1} \bar{B}_c + D.$$

Важно обстоятелство е, че независимо от равенството на предавателната матрица на линейна система

$$\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]$$

на нейната управляема и наблюдаема част

$$\left[ \begin{array}{c|c} \bar{A}_{co} & \bar{B}_{co} \\ \hline \bar{C}_{co} & D \end{array} \right],$$

тяхното вътрешно поведение може да е твърде различно. Иначе казано, въпреки че входно/изходното поведение е същото, динамичната реакция при ненулеви начални условия е различна.

## 1.4 Канонични форми

Управляема и наблюдаема система може да се приведе чрез неособени преобразувания в еквивалентна система, матриците на която имат специална структура. Тази структура се характеризира с голям брой фиксирани елементи (напр. нули и единици) в определени позиции и се нарича *канонична форма* на системата. Използвайки различни неособени преобразувания, една и съща система може

да се приведе в различни канонични форми, в които нетривиалните елементи са разположени по различен начин.

Нека е дадена линейната система, описана с уравнението (1.1), за която се предполага, че е управляема. Освен това се приема, че матрицата

$$B = [b_1, b_2, \dots, b_m]$$

има линейно независими колони, т.е. тази матрица има пълен колонен ранг.

Тъй като матрицата на управляемостта

$$C = [b_1, \dots, b_m, Ab_1, \dots, Ab_m, \dots, A^{n-1}b_m] \quad (1.3)$$

има ранг  $n$ , то винаги е възможно да се изберат  $n$  линейно независими колони в (1.3). Ако тези колони се изберат по реда, в който те се срещат от ляво на дясно, което по предположение включва всички колони на  $B$ , то се получава матрицата

$$S = [b_1, \dots, A^{p_1-1}b_1, b_2, \dots, A^{p_2-1}b_2, \dots, b_m, \dots, A^{p_m-1}b_m].$$

Числата  $p_1, p_2, \dots, p_m$ , свързани с двойката  $(A, B)$ , се наричат *Кронекерови инварианти* или индекси на управляемост, когато  $p_1 \geq p_2 \geq \dots \geq p_m$ . Числото  $p = \max p_i$  ( $i = 1, \dots, m$ ) се нарича *индекс на управляемост* на двойката  $(A, B)$ , съответно на системата (1.1). Индексите на управляемост са инвариантни към неособени преобразувания на състоянието, входа и изхода на системата.

Да конструираме матрицата

$$T = \begin{bmatrix} s_1 \\ \vdots \\ s_1 A^{p_1-1} \\ \vdots \\ s_m \\ \vdots \\ s_m A^{p_m-1} \end{bmatrix}$$

където  $s_i$  ( $i = 1, \dots, m$ ) е  $\sigma_i$ -тият ред на матрицата  $S^{-1}$ ,  $\sigma_i = \sum_{j=1}^i p_j$ .

Използвайки неособеното преобразуване  $\tilde{x} = Tx$ , получаваме еквивалентно-то представяне в пространство на състоянията  $\dot{\tilde{x}} = A_c \tilde{x} + B_c u$ , където

$$A_c = TAT^{-1} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mm} \end{bmatrix} =$$



$$\left[ \begin{array}{cccc|cccc|c|cccc}
0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & & 0 & 0 & \dots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \dots & \vdots & \vdots & & \vdots \\
0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & & 0 & 0 & \dots & 0 \\
\times & \times & \dots & \times & \times & \times & \dots & \times & & \times & \times & \dots & \times \\
\hline
0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & & 0 & 0 & \dots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \dots & \vdots & \vdots & & \vdots \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & & 0 & 0 & \dots & 0 \\
\times & \times & \dots & \times & \times & \times & \dots & \times & & \times & \times & \dots & \times \\
\hline
& & \vdots & & & & \dots & & \ddots & & & \vdots & \\
\hline
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & & 0 & 1 & \dots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \dots & \vdots & \vdots & & \vdots \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & & 0 & 0 & \dots & 1 \\
\times & \times & \dots & \times & \times & \times & \dots & \times & & \times & \times & \dots & \times
\end{array} \right], \quad (1.4)$$

$$B_c = \left[ \begin{array}{cccc}
0 & 0 & \dots & 0 \\
\vdots & \vdots & & \vdots \\
0 & 0 & \dots & 0 \\
1 & \times & \dots & \times \\
\hline
0 & 0 & \dots & 0 \\
\vdots & \vdots & & \vdots \\
0 & 0 & \dots & 0 \\
0 & 1 & \dots & \times \\
\hline
& & \vdots & \\
\hline
0 & 0 & \dots & 0 \\
\vdots & \vdots & & \vdots \\
0 & 0 & \dots & 0 \\
0 & 0 & \dots & 1
\end{array} \right] \quad (1.5)$$

и с  $\times$  са означени елементи, които нямат фиксирана стойност. Диагоналните блокове  $A_{ii}$  на  $A_c$  са съпровождащи матрици с размерности  $p_i \times p_i$ , а извъндиагоналните блокове  $A_{ij}$  за  $i \neq j$  са нулеви, с изключение на последните им редове.

Двойката  $(A_c, B_c)$  се нарича *многомерна фазово-координатна канонична форма* или *канонична форма на Луенбергер* на двойката  $(A, B)$ , съответно на системата (1.1). Трябва да се отбележи, че цялата информация относно матрицата на състоянието  $A_c$  може да се извлече от индексите на управляемост  $p_i$  и  $m$ -те нетривиални редове на  $A_c$  в позициите  $\sigma_i$ . Същото може да се каже и за матрицата  $B_c$ , тъй като и при нея същите редове са нетривиални.

**Пример 1.2** Нека е дадена система, матриците на която са

$$A = \begin{bmatrix} 0 & -6 & 2 & -3 & 10 \\ 1 & -17 & 7 & -10 & 27 \\ -4 & -11 & 4 & -7 & 21 \\ -3 & -7 & 2 & -4 & 14 \\ 1 & -10 & 4 & -6 & 16 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 0 & 3 \\ 0 & 2 \\ 1 & 0 \end{bmatrix}.$$

Първите пет линейно независими колони на матрицата на управляемостта са

$$[b_1, b_2, Ab_1, Ab_2, A^2b_1],$$

така че

$$S = [b_1, Ab_1, A^2b_1, b_2, Ab_2] = \begin{bmatrix} 1 & -2 & 5 & 0 & 0 \\ 2 & -6 & 14 & 0 & 1 \\ 0 & -5 & 12 & 3 & -2 \\ 0 & -3 & 8 & 2 & -2 \\ 1 & -3 & 8 & 0 & 0 \end{bmatrix}$$

и

$$p_1 = 3, \quad p_2 = 2, \quad \sigma_1 = p_1 = 3, \quad \sigma_2 = p_1 + p_2 = n = 5.$$

За да получим каноничната форма на Луенбергер е необходимо да изчислим матрицата на преобразуванието  $T$ . За целта най-напред трябва да се определят  $s_1$  и  $s_2$ , третият и петият (съответно на  $\sigma_1$  и  $\sigma_2$ ) редове на  $S^{-1}$ . В случая

$$s_1 = [1 \quad -2 \quad 2 \quad -3 \quad 3], \quad s_2 = [2 \quad -3 \quad 4 \quad -6 \quad 4],$$

от което следва, че

$$T = \begin{bmatrix} s_1 \\ s_1 A \\ s_1 A^2 \\ s_2 \\ s_2 A \end{bmatrix} = \begin{bmatrix} 1 & -2 & 2 & -3 & 3 \\ 2 & -3 & 2 & -3 & 4 \\ 2 & -2 & 1 & -2 & 3 \\ 2 & -3 & 4 & -6 & 4 \\ 3 & -3 & 3 & -4 & 3 \end{bmatrix}.$$

Ето защо

$$A_c = TAT^{-1} = \begin{bmatrix} 0 & 1 & 0 & | & 0 & 0 \\ 0 & 0 & 1 & | & 0 & 0 \\ 1 & 2 & 0 & | & -1 & 0 \\ - & - & - & - & - & - \\ 0 & 0 & 0 & | & 0 & 1 \\ 1 & 1 & 0 & | & 0 & -1 \end{bmatrix}, \quad B_c = TB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & -1 \\ - & - \\ 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

За този пример  $A_{11}$  и  $A_{22}$  са съпровождащи матрици с размерности съответно  $3 \times 3$  и  $2 \times 2$ .

По аналогичен начин напълно наблюдаема система (1.1),(1.2) може да се преобразува така, че матриците  $A$  и  $C$  да се получат в наблюдаема канонична

форма, като се използва дуалността. В този случай съответните цели числа  $q_1, q_2, \dots, q_r$  се наричат *индекси на наблюдаемост* и ако системата е напълно наблюдаема, то

$$\sum_{i=1}^r q_i = n.$$

Числото  $q = \max q_i$  ( $i = 1, \dots, r$ ) се нарича *индекс на наблюдаемост* на двойката  $(C, A)$ , съответно на системата (1.1). Индексите на наблюдаемост са инвариантни към неособени преобразувания на входа, състоянието и изхода на системата.

Получаването на каноничната форма на Луенбергер изисква изчисляване на матрицата на управляемостта (съответно наблюдаемостта), което е нежелателно от изчислителна гледна точка, тъй като може да е съпроводено с възникване на големи грешки от закръгляване, особено при големи норми на  $A$  и  $B$  и голямо  $n$ .

Изчислителните трудности, свързани с получаване на каноничната форма на Луенбергер, са присъщи на всички методи, при които се извършва преобразуване в "строго" канонична форма, поради големия брой на нулите и зададеното положение на единиците в тези форми. От друга страна, в редица случаи е достатъчно да се познава само структурата на системата, определена от индексите на управляемост (наблюдаемост), като по-нататъшното преобразуване в "строго" канонична форма не е необходимо. Това дава възможност, вместо неособени преобразувания от общ вид, да се използват някои специални преобразувания, такива като ортогоналните, при които се въвеждат малки грешки от закръгляване. В анализа и синтеза на линейните системи широко са разпространени т.н. *ортогонални канонични форми*, които се получават посредством числено устойчиви алгоритми. При едноходови системи двойката  $(A, b)$  може да се приведе в ортогонална канонична форма  $(A_c, b_c)$ , като

$$A_c = U^T A U = \begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ & a_{32} & \cdots & \cdots & a_{3n} \\ & & \ddots & & \vdots \\ 0 & & & a_{n,n-1} & a_{nn} \end{bmatrix}, \quad b_c = U^T B = \begin{bmatrix} b_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

където  $U$  е ортогонална матрица. Ако системата е управляема, то

$$b_1 \neq 0, \quad a_{i,i-1} \neq 0, \quad i = 2, 3, \dots, n.$$

Тази форма е известна също така под името Хесенбергова форма на двойката  $(A, b)$ .

В случай на многоходова система управляемата двойка  $(A, B)$  може да се приведе в ортогоналната канонична форма

$$A_c = U^T A U = \begin{bmatrix} A_{11} & A_{12} & \cdots & \cdots & A_{1p} \\ A_{21} & A_{22} & \cdots & \cdots & A_{2p} \\ & A_{32} & \cdots & \cdots & A_{3p} \\ & & \ddots & & \vdots \\ 0 & & & A_{p,p-1} & A_{pp} \end{bmatrix}, \quad B_c = U^T B = \begin{bmatrix} B_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (1.6)$$

където  $B_1$  е  $m_1 \times m$  матрица,  $A_{i,i-1}$ ,  $i = 2, 3, \dots, p$  са  $m_i \times m_{i-1}$  матрици,

$$\begin{aligned} \text{rank} B_1 &= m_1, \\ \text{rank} A_{i,i-1} &= m_i, \quad i = 2, 3, \dots, p \end{aligned}$$

и числата

$$m_1 \geq m_2 \geq \dots \geq m_p, \quad m_1 + m_2 + \dots + m_p = n,$$

са така наречените *спрегнати индекси на Кронекер* на двойката  $(A, B)$ . Формата (1.6) се нарича също блочна Хесенбергова форма на двойката  $(A, B)$ .

Ако системата е неуправляема, блокът  $A_{p,p-1}$  е равен на нула и блокът  $A_{pp}$  представя напълно неуправляемата част на системата.

Спрегнатите индекси на Кронекер са определени от

$$\begin{aligned} m_1 &= \text{rank} B, \\ m_i &= \text{rank} [B, AB, \dots, A^{i-1}B] - \\ &\quad \text{rank} [B, AB, \dots, A^{i-2}B] \quad (i = 2, \dots, p). \end{aligned}$$

Тези индекси могат да се определят от индексите на управляемост съгласно правилото:

$m_i$  = броят на числата от множеството  $\{p_1, p_2, \dots, p_m\}$ , които са по-големи или равни на  $i$  за  $i = 1, 2, \dots, p$ .

По подобен начин наблюдаемата двойка  $(C, A)$  в случай на един изход се привежда във вида

$$A = U^T A U = \begin{bmatrix} a_{11} & a_{12} & & 0 \\ \vdots & \vdots & \ddots & \\ a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n} \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \quad c_o = cU = [c_1 \ 0 \ \dots 0]$$

и при многоизходна система във вида

$$A = U^T A U = \begin{bmatrix} A_{11} & A_{12} & & 0 \\ \vdots & \vdots & \ddots & \\ A_{q-1,1} & A_{q-1,2} & \cdots & A_{q-1,n} \\ A_{q1} & A_{q2} & \cdots & A_{qq} \end{bmatrix}, \quad C_o = C U = [C_1 \ 0 \ \dots 0]$$

където  $q$  е индексът на наблюдаемост на системата,  $C_1$  е  $r \times r_1$  и  $A_{i,i+1}$ ,  $i = 1, 2, \dots, q-1$  са  $r_i \times r_{i+1}$  матрици,

$$\begin{aligned} \text{rank} C_1 &= r_1, \\ \text{rank} A_{i,i+1} &= r_{i+1}, \quad i = 1, 2, \dots, q-1 \end{aligned}$$

и

$$r_1 \geq r_2 \geq \dots \geq r_q, \quad r_1 + r_2 + \dots + r_q = n,$$

са спрегнатите индекси на Кронекер за двойката  $(C, A)$ .

Ако системата е ненаблюдаема, блкът  $A_{q-1,q}$  е нулев и блокът  $A_{qq}$  представя напълно ненаблюдаемата част на системата.

Числен алгоритъм за привеждане в ортогонална канонична форма е разгледан в т. 7.9.

**Пример 1.3** Нека

$$A = \begin{bmatrix} 5. & -16. & 20. & -12. \\ 8. & -21. & 22. & -8. \\ 6. & -14. & 12. & 0. \\ 2. & -4. & 2. & 3. \end{bmatrix}, \quad B = \begin{bmatrix} 2. & 5. \\ 3. & -1. \\ 1. & 0. \\ -2. & -4. \end{bmatrix}.$$

С точност до четвъртия знак за матриците на ортогоналната канонична форма се получава

$$A_c = \left[ \begin{array}{cc|cc} 8.1905 & -1.5561 & -4.5455 & 13.0375 \\ 16.2364 & -13.7385 & 3.6518 & 30.0449 \\ \hline 10.4885 & -9.4992 & 2.8445 & 20.3003 \\ .0000 & -3.8514 & 4.5040 & 1.7035 \end{array} \right],$$

$$B_c = \begin{bmatrix} -2.3146 & -6.4807 \\ -3.5557 & .0000 \\ \hline .0000 & .0000 \\ .0000 & .0000 \end{bmatrix}$$

и матрицата на ортогоналното преобразуване е

$$U = \begin{bmatrix} -.7715 & -.0603 & -.6136 & .1571 \\ .1543 & -.9442 & -.0271 & .2898 \\ .0000 & -.2812 & -.2119 & -.9359 \\ .6172 & .1607 & -.7602 & .1239 \end{bmatrix}.$$

Системата е напълно управляема, тъй като

$$m_1 = 2, \quad m_2 = 2$$

и

$$m_1 + m_2 = 4.$$

Вижда се, че използването на ортогонални преобразувания на подобие води до намаляване на броя на нулевите елементи в каноничната форма, в сравнение с формата на Луенбергер, която се получава с общи преобразувания. Преимуществото на ортогоналната канонична форма е, че тя може да се получи с числено устойчив алгоритъм (т. 7.9).

## 1.5 Операции над системи

Да предположим, че  $G_1$  и  $G_2$  са две подсистеми с представяния в пространство на състоянията:

$$G_1 = \left[ \begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array} \right], \quad G_2 = \left[ \begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D_2 \end{array} \right].$$

Тогава последователното свързване на тези две подсистеми е система с изходът на втората подсистема като вход на първата подсистема, както е показано на следната схема:



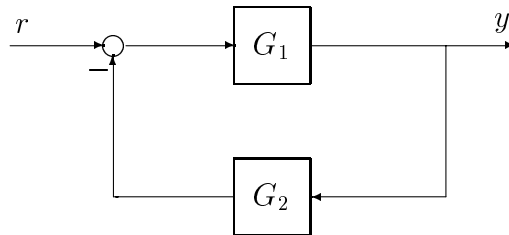
Тази операция по отношение на предавателните матрици на двете подсистеми е всъщност произведение на двете предавателни матрици. Следователно, представянето на последователното съединение може да се получи като

$$\begin{aligned} G_1 G_2 &= \left[ \begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array} \right] \left[ \begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D_2 \end{array} \right] \\ &= \left[ \begin{array}{cc|c} A_1 & B_1 C_2 & B_1 D_2 \\ 0 & A_2 & B_2 \\ \hline C_1 & D_1 C_2 & D_1 D_2 \end{array} \right] = \left[ \begin{array}{cc|c} A_2 & 0 & B_2 \\ B_1 C_2 & A_1 & B_1 D_2 \\ \hline D_1 C_2 & C_1 & D_1 D_2 \end{array} \right]. \end{aligned}$$

По подобен начин, паралелното съединение на  $G_1$  и  $G_2$  може да се получи като

$$G_1 + G_2 = \left[ \begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array} \right] + \left[ \begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D_2 \end{array} \right] = \left[ \begin{array}{cc|c} A_1 & 0 & B_1 \\ 0 & A_2 & B_2 \\ \hline C_1 & C_2 & D_1 + D_2 \end{array} \right].$$

По-нататък разглеждаме съединение с обратна връзка на  $G_1$  и  $G_2$ , както е показано по-долу:



Тогава предавателната матрица на затвореното съединение от  $r$  към  $y$  се дава от

$$T = \left[ \begin{array}{cc|c} A_1 - B_1 D_2 R_{12}^{-1} C_1 & -B_1 R_{21}^{-1} C_2 & B_1 R_{21}^{-1} \\ B_2 R_{12}^{-1} C_1 & A_2 - B_2 D_1 R_{21}^{-1} C_2 & B_2 D_1 R_{21}^{-1} \\ \hline R_{12}^{-1} C_1 & -R_{12}^{-1} D_1 C_2 & D_1 R_{21}^{-1} \end{array} \right]$$

където  $R_{12} = I + D_1 D_2$  и  $R_{21} = I + D_2 D_1$ . Отбележете, че тези представяния в пространство на състоянията не е необходимо да са управляеми и/или наблюдаеми, дори ако първоначалните подсистеми са такива.

По-нататък ще се използват също следните определения.

**Определение 1.7** Транспонираната на предавателна матрица  $G(s)$  или *дуалната система* се дефинира като

$$G \longmapsto G^T(s) \stackrel{\text{def}}{=} B^H(sI - A^H)^{-1}C^H + D^H$$

или еквивалентно

$$\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] \longmapsto \left[ \begin{array}{c|c} A^H & C^H \\ \hline B^H & D^H \end{array} \right].$$

**Определение 1.8** *Спрегнатата* система на  $G(s)$  се дефинира като

$$G \longmapsto G^\sim(s) \stackrel{\text{def}}{=} G^T(-s) = B^H(-sI - A^H)^{-1}C^H + D^H$$

или еквивалентно

$$\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] \longmapsto \left[ \begin{array}{c|c} -A^H & -C^H \\ \hline B^H & D^H \end{array} \right].$$

В частност, имаме че  $G^H(j\omega) \stackrel{\text{def}}{=} [G(j\omega)]^H = G^\sim(j\omega)$ .

**Определение 1.9** Реалната рационална матрица  $\hat{G}$  се нарича *дясна (лява) обратна* на предавателната матрица  $G(s)$  ако  $G(s)\hat{G}(s) = I$  ( $\hat{G}(s)G(s) = I$ ). Освен това, ако  $\hat{G}(s)$  е едновременно и дясна обратна и лява обратна на  $G(s)$ , то тя се нарича *обратна* на  $G(s)$ .

**Лема 1.11** Нека  $D^\dagger$  означава *дясната (лявата) обратна* на  $D$  ако  $D$  има пълен редови (колонен) ранг. Тогава

$$G^\dagger = \left[ \begin{array}{c|c} A - BD^\dagger C & -BD^\dagger \\ \hline D^\dagger C & D^\dagger \end{array} \right]$$

е дясна (лява) обратна на  $G$ .

## 1.6 Реализация на предавателни матрици

В някои случаи е по-естествено или по-удобно описанието на линейна система да се извърши с предавателни функции. Така например, съществуват сложни системи от висок ред, за които е трудно да се намери аналитично описание с диференциални уравнения. В такива случаи поведението на част от системата може да се характеризира с честотни характеристики, снети по експериментален път, които могат да се апроксимират с подходящи предавателни функции. Тъй като изчисленията в пространство на състоянията са най-удобни при използването на компютър, то необходимо е подходящо представяне в пространство на състоянията на получените предавателни функции.

Нека предположим, че  $G(s)$  е реална рационална предавателна матрица, която е *правилна*, т.е. редът на полинома в числителя на всеки елемент на предавателната матрица е по-малък или равен на реда на полинома в знаменателя

на този елемент. Тогава моделът в пространство на състоянията  $(A, B, C, D)$  наричаме *реализация* на  $G(s)$ , ако

$$G(s) = \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right].$$

**Определение 1.10** Реализацията в пространство на състоянията  $(A, B, C, D)$  на  $G(s)$  се нарича *минимална реализация* на  $G(s)$  ако  $A$  има възможно най-малката размерност.

**Теорема 1.12** Реализацията в пространство на състоянията е минимална, ако и само ако  $(A, B)$  е управляема и  $(C, A)$  е наблюдаема.

Възможно е да се докаже и следното свойство на минималните реализации.

**Теорема 1.13** Нека  $(A_1, B_1, C_1, D)$  и  $(A_2, B_2, C_2, D)$  са две минимални реализации на реалната рационална предавателна матрица  $G(s)$  и нека  $\mathcal{C}_1, \mathcal{C}_2, \mathcal{O}_1$  и  $\mathcal{O}_2$  са съответните матрици на управляемостта и наблюдаемостта. Тогава съществува единствено неособено  $T$ , такова че

$$A_2 = T A_1 T^{-1}, \quad B_2 = T B_1, \quad C_2 = C_1 T^{-1}.$$

Освен това,  $T$  може да се определи от  $T = (\mathcal{O}_2^H \mathcal{O}_2)^{-1} \mathcal{O}_2^H \mathcal{O}_1$  или  $T^{-1} = \mathcal{C}_1 \mathcal{C}_2^H (\mathcal{C}_2 \mathcal{C}_2^H)^{-1}$ .

Нека разгледаме възможните начини за получаване на реализация в пространство на състоянията на дадена многомерна предавателна матрица  $G(s)$ . Най-простият и най-пряк начин за получаване на реализация е да се намери реализацията на всеки елемент на матрицата  $G(s)$  и след това да се комбинират тези отделни реализации, за да се формира реализацията на  $G(s)$ . За илюстрация на този подход да разгледаме  $2 \times 2$  (блочна) предавателна матрица, такова че

$$G(s) = \left[ \begin{array}{cc} G_1(s) & G_2(s) \\ G_3(s) & G_4(s) \end{array} \right]$$

и да предположим, че  $G_i(s)$  има реализация в пространство на състоянията

$$G_i(s) = \left[ \begin{array}{c|c} A_i & B_i \\ \hline C_i & D_i \end{array} \right], \quad i = 1, \dots, 4.$$

Да отбележим, че  $G_i(s)$  може да е сама по себе си многомерна предавателна матрица. Тогава реализация на  $G(s)$  се дава от

$$G(s) = \left[ \begin{array}{cccc|cc} A_1 & 0 & 0 & 0 & B_1 & 0 \\ 0 & A_2 & 0 & 0 & 0 & B_2 \\ 0 & 0 & A_3 & 0 & B_3 & 0 \\ 0 & 0 & 0 & A_4 & 0 & B_4 \\ \hline C_1 & C_2 & 0 & 0 & D_1 & D_2 \\ 0 & 0 & C_3 & C_4 & D_3 & D_4 \end{array} \right].$$



Проблем, който възниква при този тип процедури за намиране на реализацията, е, че получената реализация в общия случай не е минимална. За да се получи минимална реализация трябва да се извършат декомпозиции по отношение на управляемостта и наблюдаемостта, с които да се елиминират неуправляемите и/или ненаблюдаемите състояния. (Алтернативен подход за отстраняване на неуправляемите и ненаблюдаемите състояния е методът на *уравновесената реализация*, който се разглежда по-нататък.)

## 1.7 Уравнения на Ляпунов

Уравненията на Ляпунов играят важна роля в почти всички задачи за анализ и синтез на многомерни системи за управление.

Нека разгледаме уравнението на Ляпунов

$$A^H X + X A + Q = 0 \quad (1.7)$$

за дадени реални матрици  $A$  и  $Q$ . Може да се покаже, че това уравнение има единствено решение, ако и само ако  $\lambda_i(A) + \bar{\lambda}_j(A) \neq 0$  за всяко  $i, j$ . Тук ще разгледаме връзката между устойчивостта на  $A$  и решението  $X$ , като за целта ще ни е необходима следната терминология.

Квадратната ермитова матрица  $Q = Q^H$  се нарича *положително определена* (*полуопределена*), означавано с  $Q > 0$  ( $\geq 0$ ), ако  $x^H Q x > 0$  ( $\geq 0$ ) за всяко  $x \neq 0$ .

**Лема 1.14** *Нека предположим, че  $A$  е устойчива. Тогава са в сила следните твърдения*

(a)  $X = \int_0^\infty e^{A^H t} Q e^{A t} dt.$

(б)  $X > 0$  ако  $Q > 0$  и  $X \geq 0$  ако  $Q \geq 0$ .

(в) Ако  $Q \geq 0$ , то  $(Q, A)$  е наблюдаема ако и само ако  $X > 0$ .

Непосредствено следствие от пункт (в) е, че за дадена устойчива матрица  $A$  двойката  $(C, A)$  е наблюдаема, ако и само ако решението на уравнението на Ляпунов

$$A^H L_o + L_o A + C^H C = 0$$

е положително определено.

Решението  $L_o$  се нарича *грамиан на наблюдаемостта*. По подобен начин, двойката  $(A, B)$  е управляема, ако и само ако решението на

$$A L_c + L_c A^H + B B^H = 0$$

е положително определено.  $L_c$  се нарича *грамиан на управляемостта*.

В редица приложения се разполага с решението на уравнението на Ляпунов и трябва да се направи заключение относно устойчивостта на матрицата  $A$ .

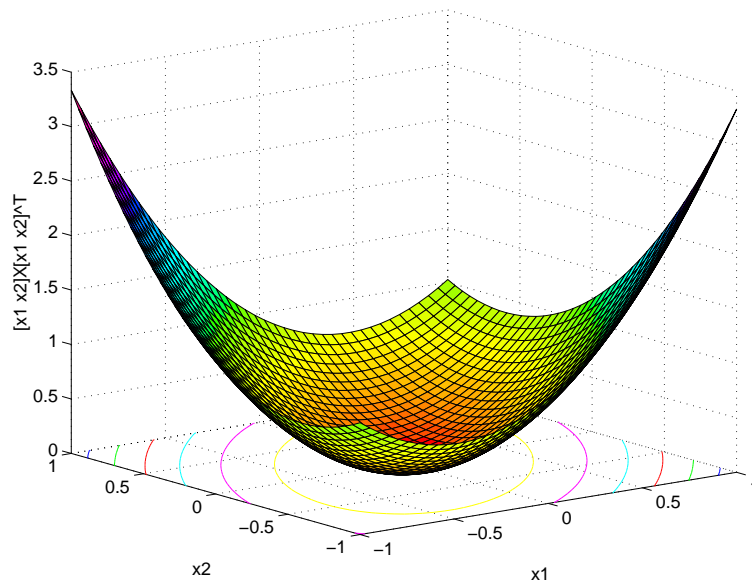
**Лема 1.15** *Нека предположим, че  $X$  е решението на уравнението на Ляпунов (1.7). Тогава*

- (а)  $\operatorname{Re} \lambda_i(A) \leq 0$ , ако  $X > 0$  и  $Q \geq 0$ .  
 (б)  $A$  е устойчива, ако  $X > 0$  и  $Q > 0$ .  
 (в)  $A$  е устойчива, ако  $X \geq 0$ ,  $Q \geq 0$  и  $(Q, A)$  е детектируема.

**Доказателство.** Нека  $\lambda$  е собствена стойност на  $A$  и  $v \neq 0$  е съответният собствен вектор, тогава  $Av = \lambda v$ . Умножавайки уравнението (1.7) отляво с  $v^H$  и отдясно с  $v$ , получаваме

$$2(\operatorname{Re} \lambda)(v^H X v) + v^H Q v = 0.$$

Сега ако  $X > 0$ , то  $v^H X v > 0$  и е ясно, че  $\operatorname{Re} \lambda \leq 0$ , ако  $Q \geq 0$  и  $\operatorname{Re} \lambda < 0$  ако  $Q > 0$ . Следователно, (а) и (б) са в сила. За да се докаже (в), да предположим, че  $\operatorname{Re} \lambda \geq 0$ . Тогава трябва да имаме, че  $v^H Q v = 0$ , т.е.  $Qv = 0$ . Това означава, че  $\lambda$  е неустойчива и ненаблюдаема, което противоречи на предположението, че  $(Q, A)$  е детектируема.



Фигура 1.1: Функция на Ляпунов за система от втори ред

На фиг. 1.1 са показани стойностите на функцията на Ляпунов  $V(x) = x^T X x$  за различни вектори на състоянието  $x$  на устойчива система от втори ред  $\dot{x} = Ax$ ,

$$A = \begin{bmatrix} -0.6 & -0.3 \\ -0.4 & -0.7 \end{bmatrix}.$$

Матрицата

$$X = \begin{bmatrix} 1.1282 & -0.5897 \\ -0.5897 & 1.0513 \end{bmatrix},$$

(елементите са показани с точност до четвъртия знак) е получена като решение на уравнението на Ляпунов за  $Q = I_2$ .

Числен алгоритъм за решаване на уравнението на Ляпунов е разгледан в т. 7.8.

## 1.8 Уравновесени реализации

Въпреки че съществуват безкрайно много различни реализации в пространство на състоянията за дадена предавателна матрица, някои специални реализации могат да са много полезни при анализа и синтеза на многомерни системи. За да се мотивира въвеждането на някои от тези реализации, най-напред ще разгледаме някои прости факти.

**Лема 1.16** *Нека  $\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]$  е реализация в пространство на състоянията на една (не непременно устойчива) предавателна матрица  $G(s)$ . Да предположим, че съществува симетрична матрица*

$$P = P^H = \begin{bmatrix} P_1 & 0 \\ 0 & 0 \end{bmatrix}$$

с  $P_1$  неособена, такава че

$$AP + PA^H + BB^H = 0.$$

Нека сега разделим реализацията  $(A, B, C, D)$  в съответствие с разделянето на  $P$ , така че

$$\left[ \begin{array}{cc|c} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \\ \hline C_1 & C_2 & D \end{array} \right].$$

Тогава

$$\left[ \begin{array}{c|c} A_{11} & B_1 \\ \hline C_1 & D \end{array} \right]$$

също е реализация на  $G$ . Освен това,  $(A_{11}, B_1)$  е управляема, ако  $A_{11}$  е устойчива.

Имаме също така

**Лема 1.17** *Нека  $\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]$  е реализация в пространство на състоянията на една (не непременно устойчива) предавателна матрица  $G(s)$ . Да предположим, че съществува симетрична матрица*

$$Q = Q^H = \begin{bmatrix} Q_1 & 0 \\ 0 & 0 \end{bmatrix}$$

с  $Q_1$  неособена, такава че

$$QA + A^H Q + C^H C = 0.$$

Нека сега разделим реализацията  $(A, B, C, D)$  в съответствие с разделянето на  $Q$ , така че

$$\left[ \begin{array}{cc|c} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \\ \hline C_1 & C_2 & D \end{array} \right].$$

Тогава

$$\left[ \begin{array}{c|c} A_{11} & B_1 \\ \hline C_1 & D \end{array} \right]$$

също е реализация на  $G$ . Освен това,  $(C_1, A_{11})$  е наблюдаема, ако  $A_{11}$  е устойчива.

Горните две лемии подсказват, че за да се получи минимална реализация от устойчива неминимална реализация, е необходимо само да се елиминират всички състояния, съответстващи на нулеви блок-диагонални членове на грамиана на управляемостта  $P$  и грамиана на наблюдаемостта  $Q$ . В случай, че  $P$  не е блок-диагонална, може да се използва следната процедура за елиминирание на неуправляемите подсистеми:

1. Нека

$$G(s) = \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]$$

е устойчива реализация.

2. Изчислява се грамиана на управляемостта  $P \geq 0$  от

$$AP + PA^H + BB^H = 0.$$

3. Диагонализира се  $P$ , за да се получи

$$P = [U_1 \ U_2] \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} [U_1 \ U_2]^H$$

с  $\Lambda_1 > 0$  и  $[U_1 \ U_2]$  унитарна.

4. Тогава

$$G(s) = \left[ \begin{array}{c|c} U_1^H A U_1 & U_1^H B \\ \hline C U_1 & D \end{array} \right]$$

е управляема реализация.

Дуална процедура може да се използва за да се елиминират ненаблюдаемите подсистеми.

Нека сега да предположим, че  $\Lambda_1 > 0$  е диагонална и е разделена като  $\Lambda_1 = \text{diag}(\Lambda_{11}, \Lambda_{22})$ , така че  $\lambda_{\max}(\Lambda_{22}) \ll \lambda_{\min}(\Lambda_{11})$ . Тогава би могло да се предположи, че могат да се пренебрегнат състоянията, които отговарят на  $\Lambda_{22}$ , без да се направи много голяма грешка. Това обаче не е непременно вярно, както показва следния пример. Да разгледаме устойчивата предавателна функция

$$G(s) = \frac{3s + 18}{s^2 + 3s + 18}.$$

Тогава  $G(s)$  има реализация в пространство на състоянията, която се дава от

$$G(s) = \left[ \begin{array}{cc|c} -1 & -4/\alpha & 1 \\ 4\alpha & -2 & 2\alpha \\ \hline -1 & 2/\alpha & 0 \end{array} \right]$$

където  $\alpha$  е кое да е ненулево число. Лесно може да се провери, че грамиана на управляемостта на реализацията е

$$P = \begin{bmatrix} 0.5 & \\ & \alpha^2 \end{bmatrix}.$$

Тъй като последния диагонален елемент на  $P$  може да се направи произволно малък правейки  $\alpha$  малко, управляемостта на съответното състояние може да се направи произволно слаба. Ако състоянието, съответстващо на последния диагонален елемент на  $P$  се премахне, получаваме предавателната функция

$$\hat{G} = \left[ \begin{array}{c|c} -1 & 1 \\ \hline -1 & 0 \end{array} \right] = \frac{-1}{s+1}$$

която в никакъв смисъл не е близо до първоначалната предавателна функция. Проблемът може лесно да се открие, ако се провери грамиана на наблюдаемостта  $Q$ , който е

$$Q = \begin{bmatrix} 0.5 & \\ & 1/\alpha^2 \end{bmatrix}.$$

Тъй като  $1/\alpha^2$  е много голямо ако  $\alpha$  е малко, това показва, че състоянието, съответстващо на последния диагонален елемент е силно наблюдаемо. По такъв начин примерът показва, че грамиана на управляемостта (или наблюдаемостта) сам по себе си не може да даде точна индикация за доминирането на състоянията на системата във входно/изходното поведение.

Това мотивира въвеждането на уравновесена реализация, която дава уравновесени грамиани на управляемостта и наблюдаемостта.

Нека да предположим, че

$$G(s) = \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]$$

е устойчива, т.е.  $A$  е устойчива. Нека  $P$  и  $Q$  означават съответно грамианите на управляемостта и наблюдаемостта. Тогава съгласно Лема 1.14  $P$  и  $Q$  удовлетворяват следните уравнения на Ляпунов

$$AP + PA^H + BB^H = 0$$

$$A^H Q + QA + C^H C = 0,$$

и  $P \geq 0$ ,  $Q \geq 0$ . Освен това, двойката  $(A, B)$  е управляема ако и само ако  $P > 0$ , и  $(C, A)$  е наблюдаема, ако и само ако  $Q > 0$ .

Да предположим, че състоянието е преобразувано чрез неособено  $T$  в  $\hat{x} = Tx$  за да се получи реализацията

$$\hat{G}(s) = \left[ \begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & \hat{D} \end{array} \right] = \left[ \begin{array}{c|c} TAT^{-1} & TB \\ \hline CT^{-1} & D \end{array} \right].$$

Тогава грамианите се преобразуват в  $\hat{P} = TPT^H$  и  $\hat{Q} = (T^{-1})^H QT^{-1}$ . Да отбележим, че  $\hat{P}\hat{Q} = TPQT^{-1}$  и следователно собствените стойности на произведението на грамианите са инвариантни относно преобразувания на състоянието.

Да разгледаме преобразувание на подобие  $T$  което дава декомпозицията по собствени вектори

$$PQ = T^{-1}\Lambda T, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n).$$

Тогава колоните на  $T^{-1}$  са собствени вектори на  $PQ$  съответстващи на собствените стойности  $\{\lambda_i\}$ . Може да се покаже, че  $PQ$  има реална диагонална жорданова форма и че  $\Lambda > 0$ , които са следствия от  $P \geq 0$  и  $Q \geq 0$ .

Въпреки че собствените вектори не са единствени, в случая на минимална реализация те винаги могат да се изберат така, че

$$\hat{P} = TPT^H = \Sigma,$$

$$\hat{Q} = (T^{-1})^H QT^{-1} = \Sigma,$$

където  $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$  и  $\Sigma^2 = \Lambda$ . Тази нова реализация с грамиани на управляемостта и наблюдаемостта  $\hat{P} = \hat{Q} = \Sigma$  ще бъде наричана *уравновесена реализация* (наричана също вътрешно уравновесена реализация). Числата с не-нарастващ порядък  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ , се наричат *Ханкелови сингулярни числа* на системата.

По-общо, ако реализацията на устойчива система не е минимална, тогава съществува преобразувание, такова че грамианите на управляемостта и наблюдаемостта на преобразуваната реализация са диагонални и управляемата и наблюдаемата подсистема са уравновесени. Това е следствие от следния факт в теория на матриците.

**Теорема 1.18** *Нека  $P$  и  $Q$  са две положително определени матрици. Тогава съществува неособена матрица  $T$ , такова че*

$$TPT^H = \begin{bmatrix} \Sigma_1 & & & \\ & \Sigma_2 & & \\ & & 0 & \\ & & & 0 \end{bmatrix} = (T^{-1})^H QT^{-1} = \begin{bmatrix} \Sigma_1 & & & \\ & 0 & & \\ & & \Sigma_3 & \\ & & & 0 \end{bmatrix},$$

съответно, със  $\Sigma_1, \Sigma_2, \Sigma_3$  диагонални и положително определени.

**Следствие 1.19** *Произведението на две положително полуопределени матрици е подобно на положително полуопределена матрица.*

**Следствие 1.20** *За всяка устойчива система*

$$G(s) = \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]$$

съществува неособено  $T$ , такова че

$$\left[ \begin{array}{c|c} TAT^{-1} & TB \\ \hline CT^{-1} & D \end{array} \right]$$

има грамиан на управляемостта  $P$  и грамиан на наблюдаемостта  $Q$ , които се дават от

$$P = \begin{bmatrix} \Sigma_1 & & & \\ & \Sigma_2 & & \\ & & 0 & \\ & & & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} \Sigma_1 & & & \\ & 0 & & \\ & & \Sigma_3 & \\ & & & 0 \end{bmatrix},$$

съответно, със  $\Sigma_1, \Sigma_2, \Sigma_3$  диагонални и положително определени.

В специалния случай, когато  $\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]$  е минимална реализация, уравновесената реализация може да се получи чрез следната опростена процедура:

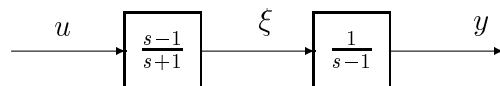
1. Изчисляват се грамианите на управляемостта и наблюдаемостта  $P > 0, Q > 0$ .
2. Намира се матрица  $R$ , такава че  $P = R^H R$ .
3. Диагонализира се  $RQR^H$  за да се получи  $RQR^H = U\Sigma^2 U^H$ .
4. Нека  $T^{-1} = R^H U \Sigma^{-1/2}$ . Тогава  $TPT^H = (T^H)^{-1}QT^{-1} = \Sigma$  и  $\left[ \begin{array}{c|c} TAT^{-1} & TB \\ \hline CT^{-1} & D \end{array} \right]$  е уравновесена.

Нека Ханкеловите сингулярни числа на системата са подредени по намаляващ ред, така че  $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$  и  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$  и да предположим, че  $\sigma_r \gg \sigma_{r+1}$  за някакво  $r$ . Тогава уравновесената реализация означава, че тези състояния, които съответстват на сингулярните числа  $\sigma_{r+1}, \dots, \sigma_n$ , са в някакъв смисъл по-малко управляеми и наблюдаеми, отколкото състоянията, съответстващи на  $\sigma_1, \dots, \sigma_r$ . Следователно, отхвърлянето на тези състояния няма да доведе до загубата на много информация относно системата.

Други две подобни реализации са *входната нормална реализация* с  $P = I$  и  $Q = \Sigma^2$ , и *изходната нормална реализация* с  $P = \Sigma^2$  и  $Q = I$ . Тези реализации могат да се получат от уравновесената реализация посредством подходящо мащабиране на състоянията.

## 1.9 Скрити моди и съкращаване на полюси и нули

Друг важен въпрос, свързан с теорията на реализацията, е проблемът за неуправляемите и/или ненаблюдаеми неустойчиви моди на линейна система. Този проблем се илюстрира със следния пример. Да разгледаме последователното съединение на две подсистеми, както е показано на следната схема



Предавателната функция на тази система,

$$g(s) = \frac{s-1}{s+1} \frac{1}{s-1} = \frac{1}{s+1},$$

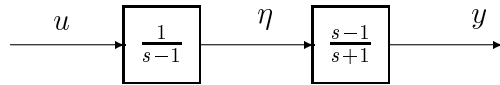
е устойчива и има минимална реализация от първи ред. От друга страна, нека

$$\begin{aligned} x_1 &= y \\ x_2 &= u - \xi. \end{aligned}$$

Тогава описанието в пространство на състоянията на тази система се дава от

$$\begin{aligned} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} &= \begin{bmatrix} 1 & -1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} u \\ y &= [1 \ 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{aligned}$$

и е система от втори ред. Освен това може да се покаже, че неустойчивата мода 1 е неуправляема, но е наблюдаема. Следователно, изходът може да е неограничен, ако началното състояние  $x_1(0)$  е ненулево. Трябва също да се отбележи, че този проблем не се премахва посредством промяна в реда на съединението:



В последния случай неустойчивата мода 1 става управляема, но ненаблюдаема. Неустойчивата мода пак може да доведе до неограничен вътрешен сигнал  $\eta$ , ако началното състояние  $\eta(0)$  е ненулево. Разбира се, съществуват фундаментални разлики между тези два типа съединения по отношение на синтеза на управлението. Например, ако състоянието е достъпно за управление с обратна връзка, то второто съединение може да бъде стабилизирано, докато първото не може.

Този пример показва, че трябва да бъдем много внимателни при съкращаването на неустойчиви моди в процедурата за формиране на предавателна функция; в противен случай получените резултати могат да са заблуждаващи и тези неустойчиви моди стават скрити моди, които могат да предизвикат неограничено нарастване на изхода. Един извод от този пример е, че проблемът в действителност се причинява от неустойчивата нула на подсистемата  $\frac{s-1}{s+1}$ . Въпреки че нулите на система с един вход и един изход лесно могат да се видят, това не е точно така за предавателна матрица на многомерна система. В действителност, понятието "нула на системата" не може да се обобщи по естествен начин от нулите на скаларна предавателна функция. Да разгледаме например следната предавателна матрица

$$G(s) = \begin{bmatrix} \frac{1}{s+1} & \frac{1}{s+2} \\ \frac{2}{s+2} & \frac{1}{s+1} \end{bmatrix}$$



която е устойчива и всеки елемент на  $G(s)$  няма крайни нули. Нека

$$K = \begin{bmatrix} \frac{s+2}{s-\sqrt{2}} & -\frac{s+1}{s-\sqrt{2}} \\ 0 & 1 \end{bmatrix}$$

която е неустойчива. Обаче,

$$KG = \begin{bmatrix} -\frac{s+\sqrt{2}}{(s+1)(s+2)} & 0 \\ \frac{2}{s+2} & \frac{1}{s+1} \end{bmatrix}$$

е устойчива. Това означава, че  $G(s)$  трябва да има неустойчива нула в  $\sqrt{2}$  която се съкращава с неустойчивия полюс на  $K$ . Това води до следващия въпрос: полюси и нули на многомерна система.

## 1.10 Полюси и нули на многомерни системи

Нека  $\mathbb{R}[s]$  означава пръстенът на полиномите с реални коефициенти. Дадена матрица се нарича полиномиална матрица, ако всеки елемент на матрицата принадлежи на  $\mathbb{R}[s]$ . Квадратна полиномиална матрица се нарича *унимодуларна матрица*, ако нейната детерминанта е ненулева константа (не е полином от  $s$ ). Възможно е да се покаже, че квадратна полиномиална матрица е унимодуларна, ако и само ако тя е обратима в  $\mathbb{R}[s]$ , т.е. нейната обратна също е полиномиална матрица.

**Определение 1.11** Нека  $Q(s) \in \mathbb{R}[s]$  е  $(r \times m)$  полиномиална матрица. Тогава *нормалният ранг* на  $Q(s)$  означава *normalrank*( $Q(s)$ ), е максимално възможният ранг на  $Q(s)$  поне за едно  $s \in \mathbb{C}$ .

Понякога ще се записва накратко, че полиномиалната матрица  $Q(s)$  има  $\text{rank}(Q(s))$  в  $\mathbb{R}[s]$  като се има предвид нормалния ранг на  $Q(s)$ .

За да се покаже разликата между нормалният ранг на полиномиална матрица и рангът на полиномиална матрица, изчислен в някаква точка, да разгледаме

$$Q(s) = \begin{bmatrix} s & 1 \\ s^2 & 1 \\ s & 1 \end{bmatrix}.$$

$Q(s)$  има нормален ранг 2, тъй като  $\text{rank } Q(2) = 2$ . Обаче,  $Q(0)$  има ранг 1.

В линейната алгебра се доказва, че всяка полиномиална матрица може да се приведе в т.н. *форма на Смит* чрез извършване на унимодуларни операции отляво и отдясно.

**Лема 1.21 (Форма на Смит)** Нека  $P(s) \in \mathbb{R}[s]$  е коя да е полиномиална матрица. Тогава съществуват унимодуларни матрици  $U(s), V(s) \in \mathbb{R}[s]$ ,

такива че

$$U(s)P(s)V(s) = S(s) \stackrel{\text{def}}{=} \begin{bmatrix} \gamma_1(s) & 0 & \dots & 0 & 0 \\ 0 & \gamma_2(s) & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \gamma_p(s) & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

и  $\gamma_i(s)$  дели  $\gamma_{i+1}(s)$ .

$S(s)$  се нарича *форма на Смит* на  $P(s)$ . Ясно е, че  $p$  е нормалният ранг на  $P(s)$ .

Процедурата за получаване на формата на Смит ще бъде илюстрирана с пример.

**Пример 1.4** Нека

$$P(s) = \begin{bmatrix} s+1 & (s+1)(2s+1) & s(s+1) \\ s+2 & (s+2)(s^2+5s+3) & s(s+2) \\ 1 & 2s+1 & s \end{bmatrix}.$$

Полиномиалната матрица  $P(s)$  има нормален ранг 2, тъй като

$$\det(P(s)) \equiv 0, \quad \det \begin{bmatrix} s+1 & (s+1)(2s+1) \\ s+2 & (s+2)(s^2+5s+3) \end{bmatrix} = (s+1)^2(s+2)^2 \neq 0.$$

Най-напред се размястват първият и третият ред и се извършва елементарна операция върху редовете, за да се нулират елементите  $s+1$  и  $s+2$  на  $P(s)$ . Този процес може да се извърши посредством умножаване на  $P(s)$  отляво с унимодуларната матрица  $U$ :

$$U = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & -(s+2) \\ 1 & 0 & -(s+1) \end{bmatrix}.$$

Тогава

$$P_1(s) \stackrel{\text{def}}{=} U(s)P(s) = \begin{bmatrix} 1 & 2s+1 & s \\ 0 & (s+1)(s+2)^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

След това се използва операция върху колоните, за да се нулират елементите  $2s+1$  и  $s$  на  $P_1$ . Този процес може да се осъществи чрез умножаването на  $P_1$  отдясно с унимодуларната матрица  $V$ :

$$V(s) = \begin{bmatrix} 1 & -2(s+1) & -s \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

и

$$P_1(s)V(s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & (s+1)(s+2)^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Тогава имаме, че

$$S(s) = U(s)P(s)V(s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & (s+1)(s+2)^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

По подобен начин, нека  $\mathcal{R}_p(s)$  означава множеството на рационалните правилни предавателни матрици. Казва се, че предавателната матрица  $G(s) \in \mathcal{R}_p(s)$  има нормален ранг  $p$ , ако  $G(s)$  има максимално възможният ранг  $p$  поне за едно  $s \in \mathbb{C}$ . Всяка реална рационална предавателна матрица може да се приведе в т.н *форма на Макмилън* чрез унимодуларни операции отляво и отдясно.

**Лема 1.22 (Форма на Макмилън)** *Нека  $G(s) \in \mathcal{R}_p(s)$  е коя да е правилна реална рационална предавателна матрица. Тогава съществуват унимодуларни матрици  $U(s), V(s) \in \mathbb{R}[s]$ , такива че*

$$U(s)G(s)V(s) = M(s) \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\alpha_1(s)}{\beta_1(s)} & 0 & \dots & 0 & 0 \\ 0 & \frac{\alpha_2(s)}{\beta_2(s)} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{\alpha_p(s)}{\beta_p(s)} & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

и  $\alpha_i(s)$  дели  $\alpha_{i+1}(s)$  и  $\beta_{i+1}(s)$  дели  $\beta_i(s)$ .

**Доказателство.** Ако запишем предавателната матрица  $G(s)$  като  $G(s) = N(s)/d(s)$ , така че  $d(s)$  е скаларен полином и  $N(s)$  е  $r \times m$  полиномиална матрица и ако положим формата на Смит на  $N(s)$  да е  $S(s) = U(s)N(s)V(s)$ , то резултатът следва, полагайки  $M(s) = S(s)/d(s)$ .

**Определение 1.12** Числото  $\sum_i \deg(\beta_i(s))$  се нарича *степен на Макмилън* на  $G(s)$ , където  $\deg(\beta_i(s))$  означава степента на полинома  $\beta_i(s)$ , т.е. най-високата степен на  $s$  в  $\beta_i(s)$ .

Степента на Макмилън на предавателна матрица е тясно свързана с размерността на минималната реализация на  $G(s)$ . Действително, може да се покаже, че размерността на минималната реализация на  $G(s)$  е точно степента на Макмилън на  $G(s)$ .

**Определение 1.13** Корените на полиномите  $\beta_i(s)$  във формата на Макмилън се наричат *полюси* на  $G$ .

Нека  $(A, B, C, D)$  да е минималната реализация на  $G(s)$ . Тогава може да се покаже, че комплексно число е полюс на  $G(s)$ , ако и само ако то е собствена стойност на  $A$ .

**Определение 1.14** Корените на всички полиноми  $\alpha_i(s)$  във формата на Макмилън на  $G(s)$  се наричат *предавателни нули* на  $G(s)$ . Комплексно число  $z_0 \in \mathbb{C}$  се нарича *блокираща нула* на  $G(s)$ , ако  $G(z_0) = 0$ .

Ясно е, че блокиращата нула е предавателна нула. Нещо повече, за скаларна предавателна функция блокиращите нули и предавателните нули съвпадат.

Горните понятия ще илюстрираме с пример.

**Пример 1.5** Да разгледаме  $3 \times 3$  предавателната матрица

$$G(s) = \begin{bmatrix} \frac{1}{(s+1)(s+2)} & \frac{2s+1}{(s+1)(s+2)} & \frac{s}{(s+1)(s+2)} \\ \frac{1}{(s+1)^2} & \frac{s^2+5s+3}{(s+1)^2} & \frac{s}{(s+1)^2} \\ \frac{1}{(s+1)^2(s+2)} & \frac{2s+1}{(s+1)^2(s+2)} & \frac{s}{(s+1)^2(s+2)} \end{bmatrix}.$$

Тази матрица може да се запише като

$$G(s) = \frac{1}{(s+1)^2(s+2)} \begin{bmatrix} s+1 & (s+1)(2s+1) & s(s+1) \\ s+2 & (s+2)(s^2+5s+3) & s(s+2) \\ 1 & 2s+1 & s \end{bmatrix} \stackrel{\text{def}}{=} \frac{N(s)}{d(s)}.$$

Тъй като  $N(s)$  е точно същото, както  $P(s)$  в предишния пример, то очевидно  $G(s)$  има формата на Макмилън

$$M(s) = U(s)G(s)V(s) = \begin{bmatrix} \frac{1}{(s+1)^2(s+2)} & 0 & 0 \\ 0 & \frac{s+2}{s+1} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

и  $G(s)$  има степен на Макмилън равна на 4. Полусите на предавателната матрица са  $\{-1, -1, -1, -2\}$  и предавателната нула е  $\{-2\}$ . Обърнете внимание, че предавателната матрица има полюс и нула на едно и също място  $\{-2\}$ ; това е уникално свойство на многомерните системи.

Съществуват и други определения за нули и полюси на система. Определенията, въведени тук, се използват най-често.

Въпреки че Лема 1.21 и 1.22 дават метод за намиране на полюсите и нулите, този метод не е удобен за практическо използване. Сигурни и практични алгоритми за тази цел са разработени в пространство на състоянията.

Нека  $(A, B, C, D)$  е реализация в пространство на състоянията на  $G(s)$ .

**Определение 1.15** Собствените стойности на  $A$  се наричат полюси на реализацията  $G(s)$ .

За да дефинираме нулите, нека разгледаме следната системна матрица

$$Q(s) = \begin{bmatrix} A - sI & B \\ C & D \end{bmatrix}.$$

**Определение 1.16** Комплексното число  $z_0 \in \mathbb{C}$  се нарича *инвариантна нула* на реализацията на системата, ако то удовлетворява

$$\text{rank} \begin{bmatrix} A - z_0 I & B \\ C & D \end{bmatrix} < \text{normalrank} \begin{bmatrix} A - sI & B \\ C & D \end{bmatrix}.$$

Инвариантните нули не се променят от постоянна обратна връзка по състоянието, тъй като

$$\begin{aligned} \text{rank} \begin{bmatrix} A + BF - z_0 I & B \\ C + DF & D \end{bmatrix} &= \text{rank} \begin{bmatrix} A - z_0 I & B \\ C & D \end{bmatrix} \begin{bmatrix} I & 0 \\ F & I \end{bmatrix} \\ &= \text{rank} \begin{bmatrix} A - z_0 I & B \\ C & D \end{bmatrix}. \end{aligned}$$

Инвариантните нули не се променят също така при неособени преобразувания на състоянието, входа и изхода.

Възможно е да се докаже следния резултат.

**Теорема 1.23** Нека  $G(s)$  е реална рационална предавателна матрица и нека  $(A, B, C, D)$  е съответната минимална реализация. Тогава комплексното число  $z_0$  е предавателна нула на  $G(s)$ , ако и само ако то е инвариантна нула на минималната реализация.

**Следствие 1.24** Всяка предавателна нула на предавателната матрица  $G(s)$  е инвариантна нула на всичките ѝ реализации и всеки полюс на предавателната матрица  $G(s)$  е полюс на всичките ѝ реализации.

И накрая, имаме следният резултат, който дава динамична интерпретация на предавателна нула на системата.

**Следствие 1.25** Нека  $G(s) \in \mathcal{R}_p(s)$  е  $p \times t$  предавателна матрица и нека  $(A, B, C, D)$  е минимална реализация. Да предположим, че  $z_0 \in \mathbb{C}$  е предавателна нула на  $G(s)$  и не е полюс на  $G(s)$ . Тогава за всеки ненулев вектор  $u_0 \in \mathbb{C}^m$  изходът на системата, дължащ се на началното състояние  $x_0 = (z_0 I - A)^{-1} B u_0$  и на входа  $u = u_0 e^{z_0 t}$ , е идентично равен на нула:  $y(t) = G(z_0) u_0 e^{z_0 t} = 0$ .



## Глава 2

# УСТОЙЧИВОСТ И КАЧЕСТВО

Най-важната цел на системата за управление е удовлетворяването на зададени изисквания към качеството в допълнение към осигуряването на вътрешна устойчивост. Един от начините да се опишат изискванията към качеството на системата за управление е чрез големините на някакви сигнали, представляващи интерес. Ето защо в началото на тази глава се разглеждат различни начини за дефиниране на големина на сигналите, т.е. различни норми на сигналите. Коя от тези норми е най-подходяща зависи от конкретната ситуация. За целта най-напред се въвеждат пространствата на Харди  $\mathcal{H}_2$  и  $\mathcal{H}_\infty$ , след което се разглежда качеството на системата при различни входни сигнали. Показва се, че  $\mathcal{H}_2$ - и  $\mathcal{H}_\infty$ -нормите възникват по естествен начин като мерки на възможно най-лошото качество за редица класове от входни сигнали. По-нататък се въвежда стандартната структурна схема на многомерна система с обратна връзка и се намират условия за реализуемост на контура и вътрешна устойчивост на системата. Разглеждат се свойствата на система с обратна връзка и се въвежда концепция за формиране честотните характеристики на контура, която се използва при синтеза на управляващо устройство.

## 2.1 Норми на сигнали и системи

### Норми на сигнали

Сигналят може да се разглежда като функция  $f$ , която изобразява множеството на реалните числа  $\mathbb{R}$  в  $\mathbb{R}^n$ . Множеството на сигналите се дава от

$$S = \{f : \mathbb{R} \rightarrow \mathbb{R}^n\}.$$

Сигналите формират линейно векторно пространство под действието на събиране и умножение със скалар, които се дефинират чрез

$$(f + g)(t) = f(t) + g(t)$$

$$(\alpha f)(t) = \alpha f(t).$$

Удобно е да се въведат двете подпространства

$$S_+ = \{f \in S : f(t) = 0 \text{ за всяко } t < 0\}$$

$$S_- = \{f \in S : f(t) = 0 \text{ за всяко } t > 0\}.$$

Пространството на Лебег на сигнали се дефинира от

$$\mathcal{L}_2(-\infty, \infty) = \{f \in S : \|f\|_2 < \infty\}$$

в което 2-нормата е

$$\|f\|_2 = \left\{ \int_{-\infty}^{\infty} f^T(t)f(t)dt \right\}^{\frac{1}{2}}.$$

Пространствата  $\mathcal{L}_{2+} = \mathcal{L}_2[0, \infty)$  и  $\mathcal{L}_{2-} = \mathcal{L}_2(-\infty, 0]$  се дефинират като подпространства на  $\mathcal{L}_2(-\infty, \infty)$  с функция  $f(t)$  равна на нула за  $t < 0$ , съответно за  $t > 0$ .

Пространството  $\mathcal{L}_2(-\infty, \infty)$  е Хилбертово пространство със скалярно произведение, дефинирано от

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f^T(t)g(t)dt.$$

Два сигнала  $f$  и  $g$  са ортогонални, ако  $\langle f, g \rangle = 0$ . Това е естествено разширение на ортогоналността в  $\mathbb{R}^n$ .

Пространствата  $\mathcal{L}_2[0, \infty)$  и  $\mathcal{L}_2(-\infty, 0]$  също са Хилбертови пространства сами по себе си с интеграл на скалярното произведение, взет над съответния временен интервал.

Аналогично се дефинира пространството на Лебег  $\mathcal{L}_2(-\infty, \infty)$  на матрични функции над  $\mathbb{R}$  със скалярно произведение

$$\langle f, g \rangle = \int_{-\infty}^{\infty} \text{trace}[f^H(t)g(t)]dt.$$

Сигнал в честотната област е функция  $f(j\omega)$ , която има свойството  $(f(j\omega))^H = f^T(-j\omega)$ . Променливата  $\omega$  е реална променлива на честотата, която се измерва в радиани за секунда. 2-нормата в честотната област се дефинира като

$$\|f\|_2 = \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} f^H(j\omega)f(j\omega)d\omega \right\}^{\frac{1}{2}}.$$

Пространството на Лебег в честотната област се състои от сигналите, които имат крайна норма

$$\mathcal{L}_2 = \{f : \|f\|_2 < \infty\}.$$

$\mathcal{L}_2$  е Хилбертово пространство със скалярно произведение

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} f^H(j\omega)g(j\omega)d\omega.$$



Причината да се използва същият символ за нормата и скаларното произведение и във временната и в честотната област е, че преобразуването на Фурие запазва скаларното произведение и 2-нормата. За  $f \in \mathcal{L}_2(-\infty, \infty)$  преобразуването на Фурие на  $f$  е

$$\hat{f}(j\omega) = \lim_{T \rightarrow \infty} \int_{-T}^T f(t) e^{-j\omega t} dt.$$

Имаме, че

$$\langle f, g \rangle = \langle \hat{f}, \hat{g} \rangle,$$

което е известно като *равенство на Парсевал*. Следствие от това равенство е, че  $\|f\|_2 = \|\hat{f}\|_2$ .

По-нататък няма да се прави разлика между означенията на сигнали във временната и честотната област, като разликата ще бъде ясна от контекста.

Пространството на Харди  $\mathcal{H}_2$  се състои от функции на комплексната променлива, които са аналитични (т.е. нямат полюси) в отворената дясна полуравнина на комплексната равнина и такива, че нормата

$$\|f\|_2 = \left\{ \sup_{\sigma > 0} \frac{1}{2\pi} \int_{-\infty}^{\infty} f^H(\sigma + j\omega) f(\sigma + j\omega) d\omega \right\}^{\frac{1}{2}}$$

е крайна, т.е.

$$\mathcal{H}_2 = \{f : f(s) \text{ е аналитична в } \operatorname{Re}(s) > 0 \text{ и } \|f\|_2 < \infty\}.$$

Може да се покаже, че

$$\|f\|_2 = \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} f^H(j\omega) f(j\omega) d\omega \right\}^{\frac{1}{2}}.$$

Следователно, нормата в  $\mathcal{H}_2$  може да се изчисли по същия начин, както в  $\mathcal{L}_2$ .

Аналогично се дефинира пространството на Харди  $\mathcal{H}_2$  на матрични функции  $F(s)$  аналитични в  $\operatorname{Re}(s) > 0$  с 2-норма

$$\|F\|_2 = \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{trace}[F^H(j\omega) F(j\omega)] d\omega \right\}^{\frac{1}{2}}.$$

Реалното рационално подпространство на  $\mathcal{H}_2$ , което се състои от всички строго правилни и реални рационални устойчиви предавателни матрици се означава с  $\mathcal{RH}_2$ .

### Норми на системи

Системата може да се разглежда като изображение от едно пространство на сигнали (входното пространство) в друго пространство на сигнали (изходното пространство):

$$\begin{aligned} G &: S_1 \mapsto S_2 \\ &: w \mapsto z = Gw. \end{aligned}$$

Системите образуват линейно пространство под действието на събиране (паралелно съединение) и умножение със скалар, които се дефинират с

$$(G_1 + G_2)w = G_1w + G_2w$$

$$(\alpha G)w = \alpha(Gw).$$

Системата  $G$  е устойчива, ако  $z = Gw$  е в  $\mathcal{L}_2[0, \infty)$ , когато  $w$  е в  $\mathcal{L}_2[0, \infty)$ .

Тъй като пространството  $\mathcal{L}_2(-\infty, \infty)$  е изоморфно на  $\mathcal{L}_2$ , линейната стационарна система изобразява  $\mathcal{L}_2(-\infty, \infty)$  в  $\mathcal{L}_2(-\infty, \infty)$ , ако и само ако предавателната матрица  $G$  е такава, че  $Gw \in \mathcal{L}_2$  за всяко  $w \in \mathcal{L}_2$ . Достатъчно условие за това е  $\sup_{\omega} \bar{\sigma}(G(j\omega)) < \infty$  (тук и по-нататък  $\sup$  означава супремумът на множество, вж. приложение П.1), тъй като

$$\begin{aligned} \|Gw\|_2^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \|G(j\omega)w(j\omega)\|^2 d\omega \\ &\leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \bar{\sigma}(G(j\omega))^2 \|w(j\omega)\|^2 d\omega \\ &\leq \sup_{\omega} \bar{\sigma}(G(j\omega))^2 \frac{1}{2\pi} \int_{-\infty}^{\infty} \|w(j\omega)\|^2 d\omega. \end{aligned}$$

Класът системи, за които супремумът е краен, е известен като  $\mathcal{L}_{\infty}$ :

$$\mathcal{L}_{\infty} = \{G : \|G\|_{\infty} < \infty\},$$

в което  $\mathcal{L}_{\infty}$ -нормата се дефинира от

$$\|G\|_{\infty} = \sup_{\omega} \bar{\sigma}(G(j\omega)).$$

Имаме, че

$$\|Gw\|_2 \leq \|G\|_{\infty} \|w\|_2 \text{ за всички } w \in \mathcal{L}_2,$$

а така също се удовлетворява важното субмултипликативно свойство

$$\|GH\|_{\infty} \leq \|G\|_{\infty} \|H\|_{\infty}.$$

Когато  $G$  е рационална,  $G \in \mathcal{L}_{\infty}$  ако и само ако  $G$  няма полюси върху имагинерната ос.

Ако  $G$  е предавателната матрица на линейна стационарна система, то  $G$  дефинира устойчива система, ако и само ако  $z = Gw \in \mathcal{H}_2$ , когато  $w \in \mathcal{H}_2$ . Това е така, тъй като  $\mathcal{L}_2[0, \infty)$  е изоморфно на  $\mathcal{H}_2$ .

Пространството на Харди  $\mathcal{H}_{\infty}$  се дефинира като

$$\mathcal{H}_{\infty} = \{G : G \text{ е аналитична в } \operatorname{Re}(s) > 0 \text{ и } \|G\|_{\infty} < \infty\},$$

в което

$$\|G\|_{\infty} = \sup_{\omega} \bar{\sigma}(G(j\omega)).$$

Система, която има предавателна матрица в  $\mathcal{H}_{\infty}$ , е устойчива система. Действително, предавателната матрица  $G$  дефинира устойчива система, ако и само ако  $G \in \mathcal{H}_{\infty}$ .

В случая, когато  $G$  е рационална,  $G \in \mathcal{H}_\infty$  ако и само ако  $G$  няма полюс в дясната комплексна полуравнина.

Терминът ”  $\mathcal{H}_\infty$  теория на управлението ” произлиза от факта, че се желае постигането на определни цели върху  $\|\cdot\|_\infty$  при условието, че затворената система е устойчива.

Реалното рационално подпространство на  $\mathcal{H}_\infty$ , означавано с  $\mathcal{RH}_\infty$ , се състои от всички правилни и реални рационални устойчиви предавателни матрици.

### $\mathcal{H}_2$ - и $\mathcal{H}_\infty$ -норми на система

Нека  $G : S_1 \mapsto S_2$ , където  $S_1$  и  $S_2$  са нормирани пространства, такива като  $\mathcal{L}_2[0, \infty)$ . Индуцираната норма на  $G$  се дефинира от

$$\|G\| = \sup_{w \neq 0} \frac{\|Gw\|_{S_2}}{\|w\|_{S_1}}. \quad (2.1)$$

Когато системата  $G$  е линейна и стационарна, тя може да се представи с предавателната функция  $G(s)$ . Тъй като  $z(s) = G(s)w(s)$ , можем да дефинираме  $\|G(s)\|$  като норма, индуцирана от нормата на сигнал в честотната област, като например  $\mathcal{L}_2$  нормата. От равенството на Парсевал следва, че тази индуцирана норма на предавателната матрица е идентична на нормата на системата, индуцирана от  $\mathcal{L}_2(-\infty, \infty)$  и следователно можем да напишем  $\|G\| = \|G(s)\|$ . За  $w \in \mathcal{L}_2$  и  $z = Gw$  имаме, че

$$\begin{aligned} \|z\|_2^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \|z(j\omega)\|^2 d\omega \\ &\leq \sup_{\omega} \bar{\sigma}(G(j\omega))^2 \frac{1}{2\pi} \int_{-\infty}^{\infty} \|w(j\omega)\|^2 d\omega \\ &= \|G\|_\infty^2 \|w\|_2^2. \end{aligned}$$

Следователно, съгласно (2.1) имаме, че  $\|G\| \leq \|G\|_\infty$ . В действителност е изпълнено, че

$$\|G\| = \|G\|_\infty.$$

По такъв начин за линейна стационарна система нормата, която е индуцирана от 2-нормата, е точно норма безкрайност на нейната предавателна матрица.

Възможно е да се даде следната физическа интерпретация на  $\mathcal{H}_\infty$ -нормата на система. Да напомним, че енергията на сигнала  $w(t)$  се дефинира като  $\|w(t)\|_2^2$ . Следователно, множеството  $S_1$  се състои от всички входове  $w$ , които имат крайна енергия. По такъв начин  $\mathcal{H}_\infty$ -нормата на система е максималната изходна енергия за всички входове в  $S_1$  с единична енергия.

$\mathcal{H}_2$ -нормата на система с правилна устойчива предавателна матрица се дава от

$$\|G\|_2 = \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \text{trace}[G^H(j\omega)G(j\omega)] d\omega \right\}^{\frac{1}{2}}.$$

Съгласно теоремата на Парсевал, тази норма е равна на  $\mathcal{H}_2$ -нормата на импулсната преходна матрица на системата

$$\|G\|_2 = \|g\|_2 \stackrel{\text{def}}{=} \left\{ \int_{-\infty}^{\infty} \text{trace}[g^H(t)g(t)]dt \right\}^{\frac{1}{2}}. \quad (2.2)$$

Тъй като

$$\text{trace}[g^H(t)g(t)] = \|g(t)\|_F^2 = \sum_{ij} |g_{ij}(t)|^2,$$

където  $g_{ij}(t)$  е  $ij$ -тият елемент на импулсната преходна матрица и тъй като редът на интегриране и сумиране в (2.2) може да се смени, то имаме, че

$$\|G\|_2 = \|g\|_2 \stackrel{\text{def}}{=} \left\{ \sum_{ij} \int_{-\infty}^{\infty} |g_{ij}(t)|^2 dt \right\}^{\frac{1}{2}}.$$

Оттук се вижда, че  $\mathcal{H}_2$ -нормата може да се интерпретира като 2-нормата на изходния сигнал, който се получава при прилагането на единични импулси  $\delta_j(t)$  към всеки вход, един след друг (изчаквайки изходът да затихне преди да се приложи импулс към следващия вход). Това се вижда ясно записвайки  $\|G\|_2^2 = \sum_{i=1}^m \|z_i(t)\|_2^2$ , където  $z_i(t)$  е изходният вектор, получен при прилагането на единичен импулс  $\delta_i(t)$  към  $i$ -тия вход.

$\mathcal{H}_2$ -нормата има и стохастическа интерпретация. Тази норма е стойността на квадратния корен от средния квадрат на изхода, когато входът  $w$  е реализация на бял шум с единична спектрална плътност  $S_{ww}$ , т.е.

$$\|G\|_2 = \mathcal{E} \{z^T z\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \text{trace}[S_{zz}(\omega)]d\omega, \quad (2.3)$$

където  $S_{zz}(\omega) = G(j\omega)S_{ww}(\omega)G^H(j\omega)$ ,  $S_{ww}(\omega) = I$  и  $\mathcal{E}$  е операторът на математическото очакване. В (2.3) е използван факта, че  $z^T z = \text{trace}(zz^T)$ . По такъв начин  $\|G\|_2$  дава информация за усилването по мощност на  $G$  когато входът е бял шум.

Използвайки декомпозицията по сингулярни числа, за 2-нормата на  $r \times m$  предавателна матрица  $G(s)$  получаваме, че

$$\|G\|_2 = \left\{ \int_{-\infty}^{\infty} \sum_{i=1}^p (\sigma_i(G(j\omega)))^2 d\omega \right\}^{\frac{1}{2}},$$

където  $p = \min\{m, r\}$ .

Въпреки че  $\|\cdot\|_2$  дефинира норма на система, субмултипликативното свойство, което се удовлетворява от  $\|\cdot\|_{\infty}$  не е в сила за  $\|\cdot\|_2$ . Това означава, че  $\|GH\|_2$  може да е по-голямо или по-малко от  $\|G\|_2\|H\|_2$ .

### Изчисляване на $\mathcal{H}_2$ - и $\mathcal{H}_\infty$ -нормите

Нека  $G(s) \in \mathcal{L}_2$ .  $\mathcal{L}_2$ -нормата на  $G$  по дефиниция е равна на

$$\begin{aligned} \|G\|_2 &\stackrel{\text{def}}{=} \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \text{trace}[G^H(j\omega)G(j\omega)]d\omega \right\}^{\frac{1}{2}} \\ &= \|g\|_2 \\ &= \left\{ \int_{-\infty}^{\infty} \text{trace}[g^H(t)g(t)]dt \right\}^{\frac{1}{2}}. \end{aligned} \quad (2.4)$$

където  $g(t)$  означава импулсната преходна матрица, чието Лапласово преобразуване е равно на  $G(s)$ .

Вижда се, че  $\mathcal{L}_2$ -нормата е крайна, ако и само ако предавателната матрица  $G$  е строго правилна, т.е.  $G(\infty) = 0$ . Следователно, по-нататък винаги ще се предполага, че предавателната матрица е строго правилна, когато се говори за  $\mathcal{L}_2$ - или  $\mathcal{H}_2$ -нормата на  $G$ .

Въпреки че  $\|G\|_2$  може по принцип да се изчисли от (2.4), в редица приложения е по-удобно да се използва алтернативен израз чрез представянето на  $G$  в пространство на състоянията. Изчисляването на  $\mathcal{RH}_2$ -нормата на предавателна матрица в този случай е съвсем просто.

**Лема 2.1** *Да разгледаме предавателната матрица*

$$G(s) = \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right]$$

с  $A$  устойчива матрица. Тогава имаме, че

$$\|G\|_2^2 = \text{trace}(B^H L_o B) = \text{trace}(C L_c C^H),$$

където  $L_o$  и  $L_c$  са грамианите на управляемост и наблюдаемост, които могат да се получат от следните уравнения на Ляпунов

$$A L_c + L_c A^H + B B^H = 0, \quad A^H L_o + L_o A + C^H C = 0.$$

По такъв начин изчисляването на  $\mathcal{L}_2$ -нормата на предавателната матрица се свежда до решаване на уравнения на Ляпунов.

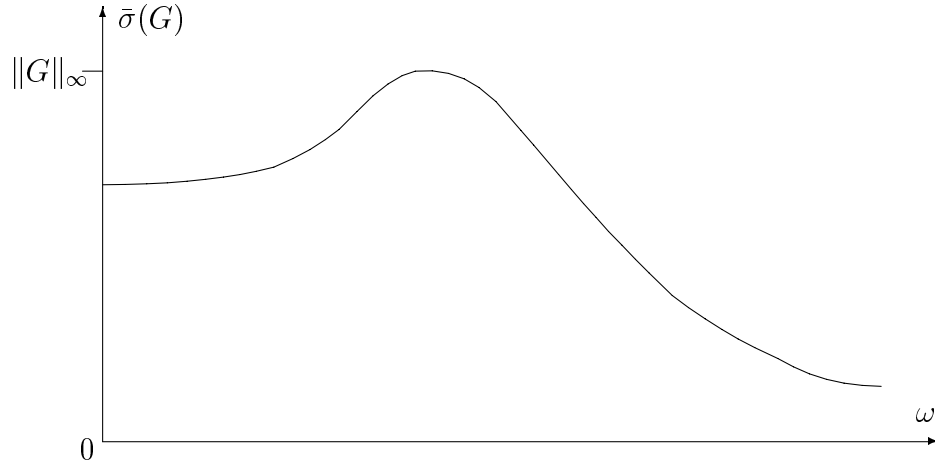
Нека сега разгледаме изчисляването на норма  $\mathcal{L}_\infty$  на предавателна матрица. Нека  $G(s) \in \mathcal{L}_\infty$ . Имаме, че

$$\|G\|_\infty \stackrel{\text{def}}{=} \sup_{\omega} \bar{\sigma} \{G(j\omega)\}.$$

Изчисляването на  $\mathcal{L}_\infty$  по този начин е сравнително сложно и изисква процедура за търсене. В едноходовия случай нормата на скаларната предавателна функция  $G$  е максимумът на  $|G(j\omega)|$  и може да се получи по графичен път (Фиг. 2.1).

За да получим оценка за  $\|G\|_\infty$  може да се зададе множество от стойности на честотата

$$\{\omega_1, \dots, \omega_N\}.$$



Фигура 2.1: Определяне на  $\|G\|_\infty$  за едновходова система

Тогава оценка за  $\|G\|_\infty$  е

$$\max_{1 \leq k \leq N} \bar{\sigma} \{G(j\omega_k)\}.$$

$\mathcal{L}_\infty$ -нормата може също да се изчисли в пространство на състоянията, ако  $G$  е рационална.

**Лема 2.2** *Нека  $\gamma > 0$  и*

$$G(s) = \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] \in \mathcal{RH}_\infty.$$

*Тогава  $\|G\|_\infty < \gamma$  ако и само ако  $\bar{\sigma}(D) < \gamma$  и  $H$  няма собствени стойности върху имагинерната ос, където*

$$H \stackrel{\text{def}}{=} \begin{bmatrix} A + BR^{-1}D^H C & BR^{-1}B^H \\ -C^H(I + DR^{-1}D^H)C & -(A + BR^{-1}D^H C)^H \end{bmatrix}$$

*и  $R = \gamma^2 I - D^H D$ .*

Лема 2.2 води до следния алгоритъм на бисекцията за изчисляване на  $\mathcal{H}_\infty$ -нормата.

**Алгоритъм на бисекцията**

- (а) избират се горна граница  $\gamma_u$  и долна граница  $\gamma_l$ , такива че  $\gamma_l \leq \|G\|_\infty \leq \gamma_u$ ;
- (б) ако  $(\gamma_u - \gamma_l)/\gamma_l \leq$  определен праг, край на изчисленията;  $\|G\| \approx (\gamma_u + \gamma_l)/2$ . В противен случай се отива на следващата стъпка;
- (в) проверява се дали  $\|G\|_\infty < \gamma$  чрез изчисляване на собствените стойности на  $H$  за дадено  $\gamma$ ;
- (г) ако  $H$  има собствена стойност върху  $j\mathbb{R}$  се полага  $\gamma_l = \gamma$ ; в противен случай се полага  $\gamma_u = \gamma$ ; отива се на обратно на стъпка (б).

Вижда се, че изчисляването на  $\mathcal{H}_\infty$ -нормата изисква търсене или по  $\gamma$ , или по  $\omega$ , за разлика от  $\mathcal{H}_2$ -нормата, която може да се намери пряко. Подобна ситуация възниква при постоянни матрици с норми  $\|M\|_F^2 = \text{trace}(M^H M)$  и  $\|M\|_2 = \bar{\sigma}(M)$ .  $\|M\|_F$  може да се получи с краен брой изчисления, но изчисляването на  $\bar{\sigma}(M)$  изисква някъкъв итеративен алгоритъм.

$\mathcal{H}_\infty$ -нормата на устойчива предавателна функция може да се оцени по експериментален път, използвайки факта, че  $\mathcal{H}_\infty$ -нормата е максималната амплитуда на реакцията в установен режим на всички възможни синусоидални входни сигнали с единична амплитуда.

**Пример 2.1** Да разгледаме система от пети ред с матрици

$$A = \begin{bmatrix} 9 + p_1 & 10 & 8 & 6 & 4 + p_2 \\ -21 & -21 & -16 & -12 & -8 \\ 15 & 14 & 11 & 9 & 6 \\ -4 & -4 & -4 & -3 & 0 \\ -2 & -2 & -2 & -3 & -5 \end{bmatrix}, B = \begin{bmatrix} 2 & -4 \\ -3 & -1 \\ 5 & 1 \\ -2 & 6 \\ 1 & 3 \end{bmatrix},$$

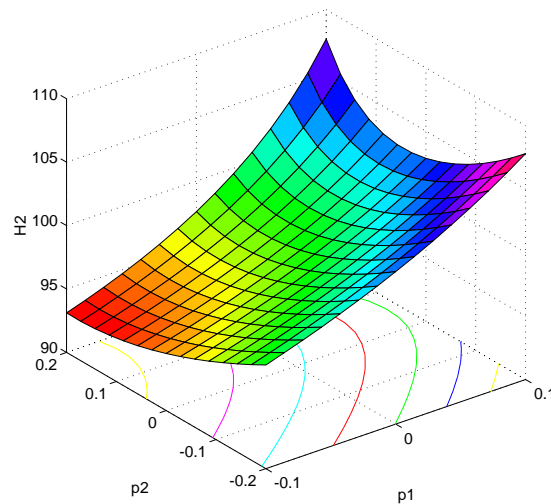
$$C = \begin{bmatrix} 9 & -3 & 1 & 8 & 2 \\ -2 & -4 & 5 & 3 & -7 \end{bmatrix},$$

където  $p_1$  и  $p_2$  са свободни параметри.

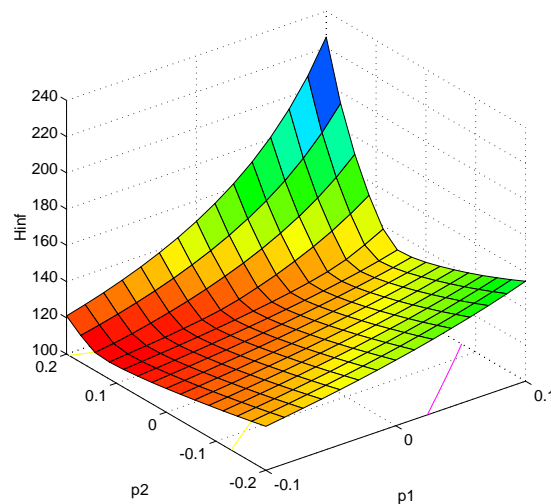
На фиг. 2.2 са показани стойностите на  $H_2$ -нормата за стойности на  $p_1$  и  $p_2$  в интервалите

$$p_1 \in [-0.1, 0.1] \quad p_2 \in [-0.2, 0.2],$$

а на фиг. 2.3 – стойностите на  $H_\infty$ -нормата за същите стойности на  $p_1$  и  $p_2$ . При тези стойности на параметрите системата е асимптотически устойчива.  $H_2$ -нормата е получена чрез решаване на уравнението на Ляпунов, а  $H_\infty$ -нормата е изчислена с алгоритъма на бисекцията при праг равен на  $10^{-8}$ . Трябва да се отбележи, че при  $p_1 = 0.2$ ,  $p_2 = 0.25$ , системата е неустойчива, което обяснява защо  $H_2$ - и  $H_\infty$ -нормите имат по-високи стойности при по-големи  $p_1$  и  $p_2$ .



Фигура 2.2:  $H_2$ -нормата като функция на  $p_1$  и  $p_2$



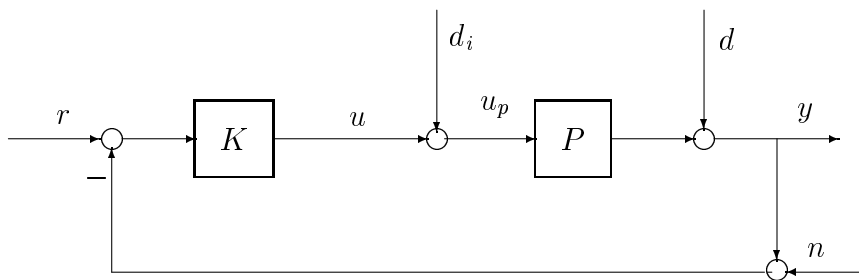
Фигура 2.3:  $H_\infty$ -нормата като функция на  $p_1$  и  $p_2$

## 2.2 Многомерни системи с обратна връзка

### Структура на обратната връзка

Основната задача при синтеза на системи за управление е изискването да се осигури удовлетворително качество на процесите в присъствие на грешки при моделирането, вариации в параметрите на системата и неопределеност. Действително, това изискване е било първоначалната мотивация за създаването на сис-





Фигура 2.4: Стандартна конфигурация с обратна връзка

темите с обратна връзка. Обратната връзка се изисква само когато качеството на системата не може да се постигне поради неопределеност в характеристиките на системата.

Нека предположим, че е даден модел, който включва представяне на неопределеността, което адекватно отразява основните свойства на обекта. Следващата стъпка в синтеза на управляващо устройство е да се определи каква структура е необходима, за да се осигури желаното качество. Управлението в отворен контур може да промени динамичната реакция на модела (по-точно на множеството от модели), но не може да намали ефекта на неопределеността. Ако неопределеността е толкова голяма, че не може да се постигне желаната точност на реакцията, тогава се изисква структура с обратна връзка. Простото използване на обратна връзка, обаче, не гарантира намаляване на неопределеността, а съществуват и други пречки за намаляване на тази неопределеност. В частност, тъй като във всяко приемливо множество от модели, представящо физическа система, неопределеността става голяма и фазовото закъснение става напълно неизвестно при достатъчно високи честоти, усиляването в контура трябва да е малко при тези честоти, за да се избегне дестабилизирането на системата от немоделираната високочестотна динамика. Това, което може да се постигне при разумното използване на обратна връзка, е значително намаляване на неопределеността за известни сигнали, които са важни, с малко увеличение, което е разпределено над останалите сигнали. По такъв начин задачата за синтез на обратна връзка се състои в компромис, включващ намаляване на цялостния ефект от неопределеността. Този компромис възниква също когато се използва обратна връзка за намаляване на грешката по задание или по смущение с едновременно минимизиране на влошаването на качеството вследствие на шумове при измерването. За да има практическа стойност, методът за синтез трябва да осигурява средства за осъществяване на тези компромиси.

Нека разгледаме стандартна конфигурация с обратна връзка, показана на фиг. 2.4.

Тя се състои от многомерен обект  $P$  и управляващо устройство  $K$ , върху което въздействат задание  $r$ , шум в чувствителния елемент  $n$ , входно смущение на обекта  $d_i$  и изходно смущение  $d$ . В общия случай се предполага, че всички сигнали са векторни и се предполага, че всички предавателни матрици имат подходящи размерности.

### Реализуемост на контура на обратната връзка

Да предположим, че обектът  $P$  и управляващото устройство  $K$  са фиксирани реални рационални правилни предавателни матрици. Първият въпрос, който тогава възниква, е дали свързването с обратна връзка има смисъл или е физически реализуемо. За определеност да разгледаме прост пример, за който

$$P = -\frac{s-1}{s+2}, \quad K = 1.$$

Имаме обаче, че

$$u = \frac{(s+2)}{3}(r-n-d) - \frac{s-1}{3}d_i$$

т.е. предавателните функции от външните сигнали  $r-n-d$  и  $d_i$  към  $u$  не са правилни. Следователно, системата с обратна връзка не е реализуема.

**Определение 2.1** Система с обратна връзка се нарича *реализуема*, ако всички предавателни матрици на затворената система са реализуеми и правилни.

Нека сега да предположим, че външните сигнали  $r, n, d$  и  $d_i$  са известни и че матричните предавателни функции на затворената система от тях към  $u$  са съответно реализуеми и правилни. Тогава  $y$  и всички други сигнали са също реализуеми и съответните предавателни матрици са правилни. Освен това, тъй като предавателните матрици от  $d$  и  $n$  към  $u$  са еднакви и се различават от предавателните матрици от  $r$  към  $u$  само по знака си, системата е реализуема, ако и само ако предавателната матрица от  $\begin{bmatrix} d_i \\ d \end{bmatrix}$  към  $u$  съществува и е правилна.

За еднаквост с означенията, използвани по-нататък, да означим

$$\hat{K} = -K.$$

Да регрупираме външните входни сигнали в контура на обратната връзка като  $w_1$  и  $w_2$  и да регрупираме входните сигнали на обекта и управляващото устройство като  $e_1$  и  $e_2$ . Тогава контурът на обратната връзка с обекта и управляващото устройство може да се представи просто както е показано на фиг. 2.5 и системата е реализуема, ако и само ако предавателната матрица от  $\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$  към  $e_1$  съществува и е правилна.

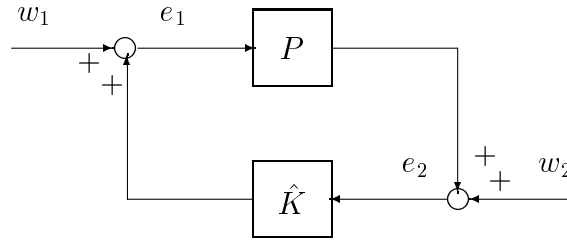
**Лема 2.3** Системата с обратна връзка на фиг. 2.5 е реализуема ако и само ако

$$I - \hat{K}(\infty)P(\infty) \quad (2.5)$$

е неособена матрица.

**Доказателство.** Системата на горната схема може да се представи във вид на уравнения като

$$e_1 = w_1 + \hat{K}e_2$$



Фигура 2.5: Схема за анализ на вътрешната устойчивост

$$e_2 = w_2 + P e_1.$$

Тогава за  $e_1$  се получава, че

$$(I - \hat{K}P)e_1 = w_1 + \hat{K}e_2.$$

Реализуемостта е еквивалентна на условието, че  $(I - \hat{K}P)^{-1}$  съществува и е правилна. Но това е еквивалентно на условието, че постоянният член на предавателната матрица  $I - \hat{K}P$  е обратим.

Може да се покаже, че (2.5) е еквивалентно на едно от следните две условия:

$$\begin{bmatrix} I & -\hat{K}(\infty) \\ -P(\infty) & I \end{bmatrix} \text{ е обратима; } \quad (2.6)$$

$$I - P(\infty)\hat{K}(\infty) \text{ е обратима.}$$

Реализацията може да се дефинира просто в пространство на състоянията. Да въведем реализациите на  $P$  и  $\hat{K}$ :

$$P = \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] \quad (2.7)$$

$$\hat{K} = \left[ \begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & \hat{D} \end{array} \right]. \quad (2.8)$$

Тогава  $P(\infty) = D$  и  $\hat{K}(\infty) = \hat{D}$ . Реализуемостта в (2.6) е еквивалентна на условието

$$\begin{bmatrix} I & -\hat{D} \\ -D & I \end{bmatrix} \text{ е обратима.}$$

В повечето практически случаи  $D = 0$  и реализуемостта за повечето реални системи е гарантирана.

### Вътрешна устойчивост

Да разгледаме система, описана със стандартната блок-схема на фиг. 2.4 и да предположим, че системата е реализуема. Освен това приемаме, че реализациите на  $P(s)$  и  $\hat{K}(s)$ , дадени в уравнения (2.7) и (2.8) са стабилизируеми и детектируеми.

Нека  $x$  и  $\hat{x}$  означават векторите на състоянията за  $P$  и  $\hat{K}$  съответно и да запишем уравненията на състоянията на фиг. 2.5 с  $w_1$  и  $w_2$  положени равни на нула:

$$\dot{x} = Ax + Be_1 \quad (2.9)$$

$$e_2 = Cx + De_1 \quad (2.10)$$

$$\dot{\hat{x}} = \hat{A}\hat{x} + \hat{B}e_2 \quad (2.11)$$

$$e_1 = \hat{C}\hat{x} + \hat{D}e_2. \quad (2.12)$$

**Определение 2.2** Системата на фиг. 2.5 се нарича *вътрешно устойчива*, ако координатното начало  $(x, \hat{x}) = (0, 0)$  е асимптотически устойчиво, т.е. състоянията  $(x, \hat{x})$  се стремят към нула за всички начални състояния при  $w_1 = 0$  и  $w_2 = 0$ .

Трябва да се отбележи, че вътрешната устойчивост е понятие в пространство на състоянията. За да се получи конкретно характеризирание на вътрешната устойчивост, да решим уравненията (2.10) и (2.12) за  $e_1$  и  $e_2$ :

$$\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} I & -\hat{D} \\ -D & I \end{bmatrix}^{-1} \begin{bmatrix} 0 & \hat{C} \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ \hat{x} \end{bmatrix}.$$

Съществуването на обратна матрица се гарантира от условието за реализуемост. Заместваме този израз в (2.9) и (2.11) и получаваме

$$\begin{bmatrix} \dot{x} \\ \dot{\hat{x}} \end{bmatrix} = \tilde{A} \begin{bmatrix} x \\ \hat{x} \end{bmatrix}$$

където

$$\tilde{A} = \begin{bmatrix} A & 0 \\ 0 & \hat{A} \end{bmatrix} + \begin{bmatrix} B & 0 \\ 0 & \hat{B} \end{bmatrix} \begin{bmatrix} I & -\hat{D} \\ -D & I \end{bmatrix}^{-1} \begin{bmatrix} 0 & \hat{C} \\ C & 0 \end{bmatrix}.$$

Вътрешната устойчивост е еквивалентна на условието всичките собствени стойности на  $\tilde{A}$  да са в лявата полуравнина.

**Лема 2.4** Системата на фиг. 2.5 с дадени стабилизируеми и детектируеми реализации на  $P$  и  $\hat{K}$  е вътрешно устойчива ако и само ако матрицата  $\tilde{A}$  е устойчива матрица.

Лесно може да се провери, че това определение на вътрешната устойчивост зависи само от  $P$  и  $\hat{K}$ , а не от техни конкретни реализации при условие, че реализациите на  $P$  и  $\hat{K}$  са едновременно стабилизируеми и детектируеми, т.е. няма допълнителни неустойчиви моди, въведени от реализациите.

Даденото по-горе определение на вътрешната устойчивост е дефинирано чрез реализации в пространство на състоянията на  $P$  и  $\hat{K}$ . Вътрешната устойчивост може също така да се характеризира от гледна точка на предавателните матрици.

**Лема 2.5** Системата на фиг. 2.5 е вътрешно устойчива, ако и само ако предавателната матрица

$$\begin{aligned} \begin{bmatrix} I & -\hat{K} \\ -P & I \end{bmatrix}^{-1} &= \begin{bmatrix} (I - \hat{K}P)^{-1} & \hat{K}(I - P\hat{K})^{-1} \\ P(I - \hat{K}P)^{-1} & (I - P\hat{K})^{-1} \end{bmatrix} \\ &= \begin{bmatrix} I + \hat{K}(I - P\hat{K})^{-1}P & \hat{K}(I - P\hat{K})^{-1} \\ (I - P\hat{K})^{-1}P & (I - P\hat{K})^{-1} \end{bmatrix} \end{aligned} \quad (2.13)$$

от  $(w_1, w_2)$  към  $(e_1, e_2)$  принадлежи на  $\mathcal{RH}_\infty$ .

За да се провери вътрешната устойчивост е необходимо и достатъчно да се провери дали всяка една от четирите предавателни матрици в (2.13) е в  $\mathcal{RH}_\infty$ . За устойчивостта не може да се даде заключение даже ако три от четирите предавателни матрици в (2.13) са в  $\mathcal{RH}_\infty$ . Например, нека

$$P = \frac{s-1}{s+1}, \quad \hat{K} = -\frac{1}{s-1}.$$

Тогава

$$\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} \frac{s+1}{s+2} & -\frac{s+1}{(s-1)(s+2)} \\ \frac{s-1}{s+2} & \frac{s+1}{s+2} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix},$$

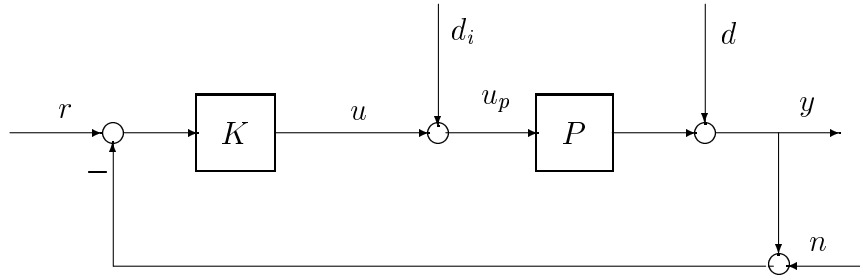
което показва, че системата не е вътрешно устойчива, въпреки че три от четирите предавателни функции са устойчиви. Това може да се види също така чрез изчисляване на матрицата на състоянието на затворената система с кои да са стабилизируеми и детектируеми реализации на  $P$  и  $\hat{K}$ .

Трябва да се отбележи, че вътрешната устойчивост е основно изискване към реална система с обратна връзка. Това е така, тъй като всички свързани системи неизбежно са обект на някакви ненулеви начални условия и някакви (евентуално малки) грешки и на практика е неприемливо такива грешки в някои места да водят до неограничени сигнали в други места на затворената система. Вътрешната устойчивост гарантира, че всички сигнали в системата са ограничени, при условие, че въведените сигнали (в кои да са точки) са ограничени.

Съществуват някои специални случаи, когато определянето на вътрешната устойчивост е просто.

**Следствие 2.6** Да предположим, че  $\hat{K} \in \mathcal{RH}_\infty$ . Тогава системата на фиг. 2.5 е вътрешно устойчива, ако и само ако тя е реализуема и  $P(I - \hat{K}P)^{-1} \in \mathcal{RH}_\infty$ .

Това следствие в действителност е в основата на класическата теория на управлението, където устойчивостта се проверява само за една предавателна функция при неявното предположение, че управляващото устройство само по себе си е устойчиво.



Фигура 2.6: Стандартна конфигурация с обратна връзка

**Следствие 2.7** *Да предположим, че  $P \in \mathcal{RH}_\infty$ . Тогава системата на фиг. 2.5 е вътрешно устойчива, ако и само ако тя е реализуема и  $\hat{K}(I - P\hat{K})^{-1} \in \mathcal{RH}_\infty$ .*

**Следствие 2.8** *Да предположим, че  $P \in \mathcal{RH}_\infty$  и  $\hat{K} \in \mathcal{RH}_\infty$ . Тогава системата на фиг. 2.5 е вътрешно устойчива, ако и само ако  $(I - P\hat{K})^{-1} \in \mathcal{RH}_\infty$ .*

## 2.3 Свойства на обратната връзка

В тази точка се обсъждат свойствата на многомерни системи с обратна връзка. В частност, разглежда се ползата от структура с обратна връзка и концепцията за компромис при синтеза в случая на конфликтни цели, а именно как да се постигне полза от обратната връзка в присъствието на неопределеност.

Нека разгледаме отново системата с обратна връзка, показана на фиг. 2.4. За удобство схемата на системата е показана отново на фиг. 2.6. За по-нататъшното разглеждане е удобно да се дефинират *предавателната матрица на входния контур*  $L_i$  и *предавателната матрица на изходния контур*  $L_o$ , като

$$L_i = KP, \quad L_o = PK,$$

съответно, където  $L_i$  се получава чрез разкъсване на контура на входа ( $u$ ) на обекта, а  $L_o$  се получава от разкъсване на контура на изхода ( $y$ ) на обекта. Матрицата на *входната чувствителност* се дефинира като предавателна матрица от  $d_i$  към  $u_p$ :

$$S_i = (I + L_i)^{-1}, \quad u_p = S_i d_i.$$

Матрицата на *изходната чувствителност* се дефинира като предавателна матрица от  $d$  към  $y$ :

$$S_o = (I + L_o)^{-1}, \quad y = S_o d.$$

Матриците на *входната и изходната допълнителна чувствителност* се дефинират като

$$T_i = I - S_i = L_i(I + L_i)^{-1}$$

$$T_o = I - S_o = L_o(I + L_o)^{-1},$$

съответно. (Думата *допълнителна* се използва за означаване на факта, че  $T$  е допълнителна на  $S$ ,  $T = I - S$ .) Матрицата  $I + L_i$  се нарича *матрица на входната възвратна разлика* и  $I + L_o$  се нарича *матрица на изходната възвратна разлика*.

Лесно се установява, че затворената система, ако тя е вътрешно устойчива, удовлетворява следните уравнения:

$$y = T_o(r - n) + S_o P d_i + S_o d \quad (2.14)$$

$$r - y = S_o(r - d) + T_o n - S_o P d_i \quad (2.15)$$

$$u = K S_o(r - n) - K S_o d - T_i d_i \quad (2.16)$$

$$u_p = K S_o(r - n) - K S_o d + S_i d_i. \quad (2.17)$$

Тези четири уравнения показват фундаменталните ползи и цели на синтеза, присъщи на контурите с обратна връзка. Например, уравнение (2.14) показва, че ефекта на смущението  $d$  върху изхода на обекта може да се направи "малък" правейки функцията на изходната чувствителност малка. По-подобен начин уравнение (2.17) показва, че ефектът на смущението  $d_i$  върху входа на обекта може да се направи малък чрез правене на функцията на входната чувствителност малка. Понятието "малко" за предавателна матрица в някъкъв честотен диапазон може да се изрази явно използвайки честотно зависими сингулярни числа. Например,  $\bar{\sigma}(S_o) < 1$  за някъкъв честотен диапазон ще означава, че ефекта на смущението  $d$  на изхода на обекта става малък за този честотен диапазон.

Следователно, добро подтискане на смущението на изхода  $y$  на обекта би изисквало

$$\bar{\sigma}(S_o) = \bar{\sigma}((I + PK)^{-1}) = \frac{1}{\underline{\sigma}(I + PK)},$$

за смущението на изхода на обекта  $d$  и

$$\bar{\sigma}(S_o P) = \bar{\sigma}((I + PK)^{-1} P) = \bar{\sigma}(P S_i),$$

за смущението на входа на обекта  $d_i$ , да се направят малки, а добро подтискане на смущението на входа на обекта  $u_p$  би изисквало

$$\bar{\sigma}(S_i) = \bar{\sigma}((I + KP)^{-1}) = \frac{1}{\underline{\sigma}(I + KP)},$$

за смущението на входа на обекта  $d_i$  и

$$\bar{\sigma}(S_i K) = \bar{\sigma}(K(I + PK)^{-1}) = \bar{\sigma}(K S_o),$$

за смущението на изхода на обекта  $d$ , да се направят малки, особено за ниско-честотния диапазон, където  $d$  и  $d_i$  обикновено са значителни.

Да отбележим, че

$$\underline{\sigma}(PK) - 1 \leq \underline{\sigma}(I + PK) \leq \underline{\sigma}(PK) + 1$$

$$\underline{\sigma}(KP) - 1 \leq \underline{\sigma}(I + KP) \leq \underline{\sigma}(KP) + 1$$

така че

$$\frac{1}{\underline{\sigma}(PK) + 1} \leq \bar{\sigma}(S_o) \leq \frac{1}{\underline{\sigma}(PK) - 1}, \text{ ако } \sigma(PK) > 1$$

$$\frac{1}{\underline{\sigma}(KP) + 1} \leq \bar{\sigma}(S_i) \leq \frac{1}{\underline{\sigma}(KP) - 1}, \text{ ако } \sigma(KP) > 1.$$

Тези уравнения означават, че

$$\bar{\sigma}(S_o) \ll 1 \iff \underline{\sigma}(PK) \gg 1$$

$$\bar{\sigma}(S_i) \ll 1 \iff \underline{\sigma}(KP) \gg 1.$$

Ако сега предположим, че  $P$  и  $K$  са обратими, то

$$\underline{\sigma}(PK) \gg 1 \text{ или } \underline{\sigma}(KP) \gg 1 \iff \bar{\sigma}(S_o P) = \bar{\sigma}((I + PK)^{-1} P)$$

$$\approx \bar{\sigma}(K^{-1}) = \frac{1}{\underline{\sigma}(K)}$$

$$\underline{\sigma}(PK) \gg 1 \text{ или } \underline{\sigma}(KP) \gg 1 \iff \bar{\sigma}(K S_o) = \bar{\sigma}(K(I + PK)^{-1})$$

$$\approx \bar{\sigma}(P^{-1}) = \frac{1}{\underline{\sigma}(P)}.$$

Следователно, добро качество на изхода  $y$  на обекта изисква изобщо голямо усиление на изходния контур,  $\underline{\sigma}(L_o) = \underline{\sigma}(PK) \gg 1$  в честотния диапазон, където  $d$  е значително, за да се намали влиянието на  $d$  и достатъчно голямо усиление на управляващото устройство  $\underline{\sigma}(K) \gg 1$  в честотния диапазон, където  $d_i$  е значително, за да се намали влиянието на  $d_i$ . По подобен начин, добро качество на входа на обекта ( $u_p$ ) изисква изобщо по-голямо усиление на входния контур  $\underline{\sigma}(L_i) = \underline{\sigma}(KP) \gg 1$  в честотния диапазон, където  $d_i$  е значително, за да се намали влиянието на  $d_i$  и достатъчно голямо усиление на обекта  $\underline{\sigma}(P) \gg 1$  в честотния диапазон, където  $d$  е значително, което не може да се промени при синтеза на управляващото устройство, за да се намали влиянието на  $d$ . (Трябва да се отбележи, че изобщо  $S_o \neq S_i$ , освен ако  $K$  и  $P$  са квадратни и диагонални, което е вярно, ако  $P$  се състои от несвързани помежду си едномерни системи. Следователно, малко  $\bar{\sigma}(S_o)$  не означава непременно малко  $\bar{\sigma}(S_i)$ ; иначе казано, добро подтискане на смущението на изхода не означава непременно добро подтискане на смущението на входа на обекта.)

По такъв начин *успешният синтез на многомерния контур с обратна връзка се свежда до постигане на високи коефициенти на усиление на контура (и евентуално на управляващото устройство) в необходимия честотен диапазон.*

Независимо от простотата на това твърдение, синтезът на обратна връзка не е тривиален. Това е така, тъй като коефициентите на усиление на контура не могат да се направят произволно големи в произволно широки честотни диапазони. Вместо това, те трябва да удовлетворяват известен компромис с качеството и ограниченията при синтеза. Един голям компромис по отношение на



качеството, например, се отнася до намаляване на грешките по задание и смущение спрямо устойчивостта при неопределеност в модела. Да предположим, че моделът на обекта е променен на  $(I + \Delta)P$  с  $\Delta$  устойчива и да предположим, че номиналната система е устойчива, т.е. затворената система с  $\Delta = 0$  е устойчива. Тогава номиналната затворена система е устойчива, ако

$$\det(I + (I + \Delta)PK) = \det(I + PK) \det(I + \Delta T_o)$$

няма нули в дясната комплексна полуравнина. В общия случай това означава, че  $\Delta T_o$  трябва да е малко или че  $\bar{\sigma}(T_o)$  трябва да е малко за тези честоти, където  $\Delta$  е значително, типично във високочестотния диапазон. Това на свой ред означава, че усиляването на контура  $\bar{\sigma}(L_o)$  трябва да е малко за тези честоти.

Друг компромис засяга намаляването на грешката от шумовете на чувствителните елементи. Конфликтът между подтискането на смущението и намаляването на шумовете от чувствителните елементи се вижда от уравнение (2.14). Големи стойности на  $\underline{\sigma}(L_o(j\omega))$  в широк честотен диапазон прави грешките, дължащи се на  $d$ , малки. Те обаче правят също грешките, дължащи се на  $n$ , големи, тъй като този шум се пропуска в същия честотен диапазон, т.е.

$$y = T_o(r - n) + S_o P d_i + S_o d \approx (r - n).$$

Отбележете, че типично  $n$  е значително във високочестотния диапазон. Още по-лошо, големи коефициенти на усиляване на контура извън лентата на пропускане на  $P$ , т.е.  $\underline{\sigma}(L_o(j\omega)) \gg 1$  или  $\underline{\sigma}(L_i(j\omega)) \gg 1$  при  $\bar{\sigma}(P(j\omega)) \ll 1$  може да направи управляващото въздействие ( $u$ ) неприемливо, тъй като може да доведе до насищане на изпълнителните елементи. Това следва от

$$u = K S_o(r - n - d) - T_i d = S_i K(r - n - d) - T_i d_i \approx P^{-1}(r - n - d) - d_i.$$

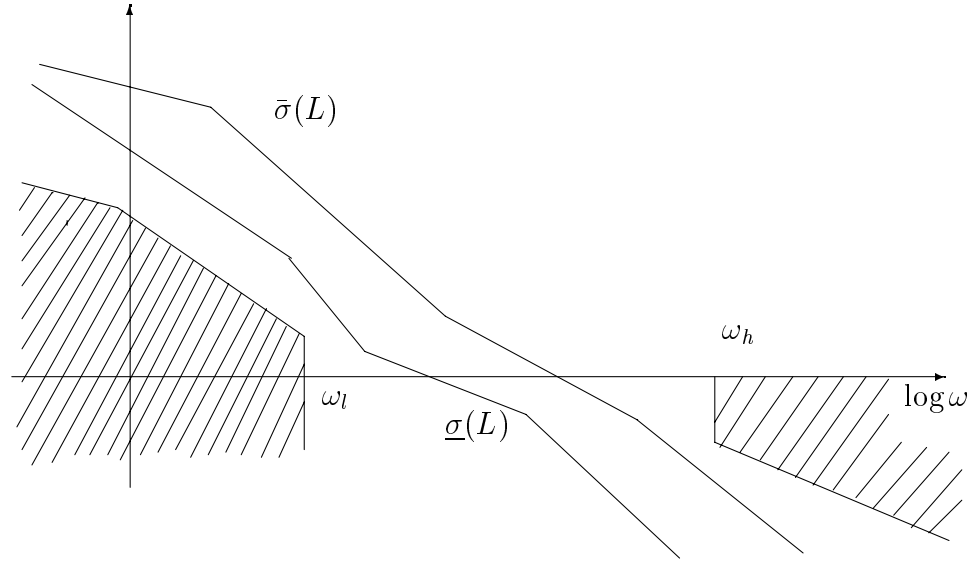
Тук бе предположено за удобство, че  $P$  е квадратна и обратима. Полученото уравнение показва, че смущаващото въздействие и шумът на чувствителния елемент в действителност се усилят в  $u$  когато честотния диапазон значително надхвърля лентата на пропускане на  $P$ , тъй като за  $\omega$  такава, че  $\bar{\sigma}(P(j\omega)) \ll 1$ , имаме че

$$\underline{\sigma}[P^{-1}(j\omega)] = \frac{1}{\bar{\sigma}[P(j\omega)]} \gg 1.$$

По подобен начин, усиляването на управляващото устройство,  $\sigma(K)$ , трябва също да се поддържа да не е толкова голямо в честотния диапазон, където усиляването на контура е малко, за да не се наситят изпълнителните устройства. Това е така, тъй като за малко усиляване на контура  $\bar{\sigma}(L_o(j\omega)) \ll 1$  или  $\bar{\sigma}(L_i(j\omega)) \ll 1$  е изпълнено

$$u = K S_o(r - n - d) - T_i d_i \approx K(r - n - d).$$

Следователно, желателно е да се поддържа  $\bar{\sigma}(K)$  да не е много голямо, когато усиляването на контура е малко.



Фигура 2.7: Желано усилване на контура

За да се резюмират горните разглеждания, да подчертаем, че доброто качество изисква в някъкъв честотен диапазон, типично някъкъв нискочестотен диапазон  $(0, \omega_l)$ , да е изпълнено

$$\underline{\sigma}(PK) \gg 1, \underline{\sigma}(KP) \gg 1, \underline{\sigma}(K) \gg 1,$$

а добра робастност и добро подтискане на шума на чувствителния елемент изисква в някъкъв честотен диапазон, типично някъкъв височестотен диапазон  $(\omega_h, \infty)$ , да е изпълнено

$$\bar{\sigma}(PK) \ll 1, \bar{\sigma}(KP) \ll 1, \bar{\sigma}(K) \leq M$$

където  $M$  не е много голямо. Тези изисквания при синтеза са показани графически на фиг. 2.7. Конкретните честоти  $\omega_l$  и  $\omega_h$  зависят от конкретното приложение и от познаването на характеристиките на смущението, неопределеността при моделирането и нивото на шумовете на чувствителния елемент.

## 2.4 Формиране честотните характеристики на контура

Анализът в предишната точка мотивира една концептуално проста техника за синтез на управляващо устройство: формиране честотните характеристики на контура. Синтезът на управляващо устройство чрез формиране честотните характеристики на контура включва основно намиране на управляващо устройство  $K$ , което формира предавателната матрица на контура  $L$  така, че коефициентите на контура  $\underline{\sigma}(L)$  и  $\bar{\sigma}(L)$  да заобикалят границите, определени от изискванията

към качеството при ниските честоти и изискванията към робастността при високите честоти, както е показано на фиг. 2.7.

В едномерния случай техниката за синтез чрез формиране честотните характеристики на контура е особено ефективна, тъй като  $\bar{\sigma}(L) = \underline{\sigma}(L) = |L|$ . Процедурата за синтез може да се извърши на две стъпки:

#### Формиране честотните характеристики на контура на едномерна система

- (1) Намира се рационална строго правилна предавателна функция  $L$ , която съдържа всички полюси и нули на  $P$  в дясната полуравнина, така че  $|L|$  заобикаля границите, определени от изискванията към качеството на ниските честоти и от изискванията към робастността на високите честоти, както е показано на фиг. 2.7.

$L$  трябва също да е така избрано, че  $1 + L$  да има всички нули в отворената лява полуравнина, което обикновено може да се гарантира, правейки  $L$  с приемлива динамика в областта на пресичането с абсцисата, т.е.  $L$  не трябва да намалява много бързо в честотния диапазон около спрягащата честота  $|L(j\omega)| \approx 1$ .

- (2) Управляващото устройство се дава от  $K = L/P$ .

Формирането на честотните характеристики на контура на многомерна система може да се извърши по подобен начин, ако сингулярните числа на предавателната матрица на контура се използват като коефициенти на усилване на контура.

#### Формиране честотните характеристики на контура на многомерна система

- (1) Намира се рационална строго правилна предавателна матрица  $L$ , която съдържа всички полюси и нули на  $P$  в дясната полуравнина, така че производението на  $P$  и  $P^{-1}L$  (или  $LP^{-1}$ ) да няма съкращавания на неустойчиви полюси и/или нули и  $\underline{\sigma}(L)$  заобикаля границите, определени от изискванията към качеството при ниските честоти и  $\bar{\sigma}(L)$  заобикаля границите, определени от изискванията към робастността при високите честоти, както е показано на фиг. 2.7.

$L$  трябва също да е така избрана, че  $\det(I + L)$  да има всичките си нули в отворената лява полуравнина. (Това не е лесно при многомерните системи.)

- (2) Управляващото устройство се дава от  $K = P^{-1}L$ , ако  $L$  е изходната предавателна матрица на контура (или  $K = LP^{-1}$ , ако  $L$  е входната предавателна матрица на контура.)

Техниката за формиране честотните характеристики на контура може да е много полезна, особено при синтеза на едномерни системи. Съществуват обаче редица ограничения, когато тя се използва при синтеза на многомерни системи.

### Ограничения при формиране честотните характеристики на контура на многомерни системи

Някои от ограниченията при синтеза чрез формиране честотните характеристики на контура са:

- Техниката за формиране честотните характеристики на контура, описана по-горе, може да се справи ефикасно само със задачи с равномерно разпределени изисквания към качеството и робастността. По-конкретно, методът може да не е ефективен при задачи с различни изисквания в различните канали. За да се илюстрира тази трудност, да разгледаме неопределената динамична система

$$P_{\Delta} = (I + \Delta)P$$

където  $P$  е номиналният обект и  $\Delta$  е мултипликативната грешка при моделирането. Да предположим, че  $\Delta$  може да се запише в следния вид

$$\Delta = \tilde{\Delta}W_t, \quad \bar{\sigma}(\tilde{\Delta}) < 1.$$

Тогава за робастна устойчивост би трябвало да изискваме  $\bar{\sigma}(\Delta T_o) = \bar{\sigma}(\tilde{\Delta}W_t T_o) < 1$  или  $\bar{\sigma}(W_t T_o) \leq 1$ . Ако се изисква равномерна граница за усилването на контура, за да се използва техниката за формиране честотните характеристики на контура, е необходимо да се ограничи отгоре  $\bar{\sigma}(W_t T_o)$ :

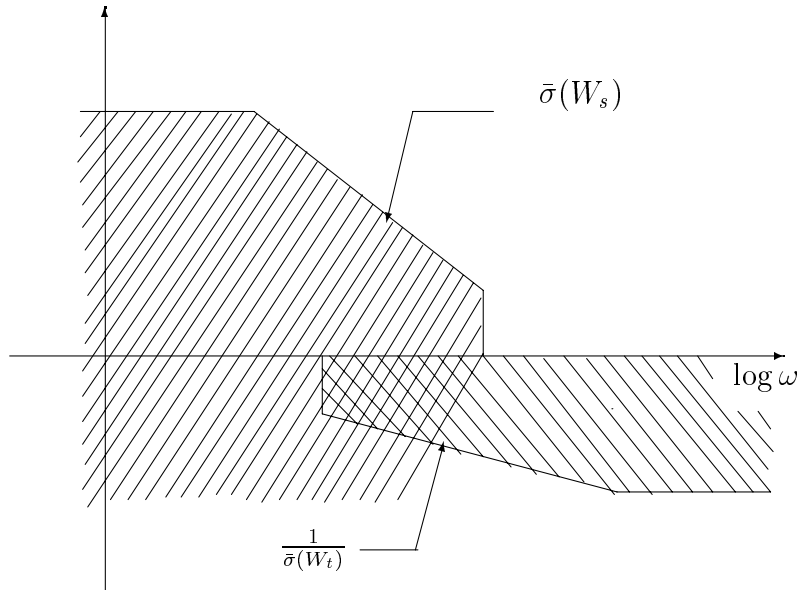
$$\bar{\sigma}(W_t T_o) \leq \bar{\sigma}(W_t) \bar{\sigma}(T_o) \leq \bar{\sigma}(W_t) \frac{\bar{\sigma}(L_o)}{1 - \bar{\sigma}(L_o)}, \text{ ако } \bar{\sigma}(L_o) < 1.$$

и изискването за робастна устойчивост е наложено чрез

$$\bar{\sigma}(L_o) \leq \frac{1}{\bar{\sigma}(W_t) + 1} \approx \frac{1}{\bar{\sigma}(W_t)}, \text{ ако } \bar{\sigma}(L_o) < 1.$$

По подобен начин, ако изискванията към качеството, например подтискането на изходното смущение, не са равномерно разпределени във всички канали, а чрез тегловна матрица  $W_s$  такава, че  $\bar{\sigma}(W_s S_o) \leq 1$ , то тогава също е необходимо да се ограничи отгоре  $\bar{\sigma}(W_s S_o)$ , за да може да се приложи техниката за формиране честотните характеристики на контура:

$$\bar{\sigma}(W_s S_o) \leq \bar{\sigma}(W_s) \bar{\sigma}(S_o) \leq \frac{\bar{\sigma}(W_s)}{\underline{\sigma}(L_o) - 1}, \text{ ако } \underline{\sigma}(L_o) > 1.$$



Фигура 2.8: Конфликтни изисквания

и изискването към качеството се подразбира от

$$\underline{\sigma}(L_o) \geq \bar{\sigma}(W_s) + 1 \approx \bar{\sigma}(W_s), \text{ ако } \underline{\sigma}(L_o) > 1.$$

Възможно е границите за формата на честотните характеристики на контура да си противоречат една на друга, както е показано на фиг. 2.8. Това обаче не означава, че не съществува управляващо устройство, което да удовлетворява едновременно номиналното качество и робастната устойчивост, освен при едновходови системи.

Това противоречие се получава, тъй като границите не използват структурата на тегловните функции  $W_s$  и  $W_t$ ; също така тези граници са само достатъчни условия за робастна устойчивост и номинално качество. Тази възможност може да се илюстрира чрез следния пример: Да предположим, че предавателната матрица на двувходова и двуизходна система се дава от

$$P(s) = \frac{1}{s+1} \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix},$$

и да предположим, че тегловните матрици се дават от

$$W_s = \begin{bmatrix} \frac{1}{s+1} & \frac{\alpha}{(s+1)(s+2)} \\ 0 & \frac{1}{s+1} \end{bmatrix}, W_t = \begin{bmatrix} \frac{s+2}{s+10} & \frac{\alpha(s+1)}{s+10} \\ 0 & \frac{s+2}{s+10} \end{bmatrix}.$$

Лесно може да се покаже, че за големи  $\alpha$  тегловните функции изглеждат така, както е показано на фиг. 2.8 и следователно техниката за формиране на честотните характеристики на контура не може да се приложи. Лесно е обаче да се покаже, че системата с управляващо устройство

$$K = I_2$$

дава

$$W_s S = \begin{bmatrix} \frac{1}{s+2} & 0 \\ 0 & \frac{1}{s+2} \end{bmatrix}, \quad W_t T = \begin{bmatrix} \frac{1}{s+10} & 0 \\ 0 & \frac{1}{s+10} \end{bmatrix}$$

и следователно изискванията за робастност и качество се удовлетворяват.

- Дори ако всички горепосочени трудности могат да се избегнат, все още може да е трудно намирането на такава предавателна матрица  $L_o$ , че  $K = P^{-1}L_o$  да стабилизира системата. Това може да стане много по-трудно, ако  $P$  е неминимално фазова или/и неустойчива.

По такъв начин е необходимо да се разглежда качеството на затворения контур директно в термините на предавателните матрици на затворената система, вместо предавателните матрици на отворената система.

## 2.5 Претеглено $\mathcal{H}_2$ и $\mathcal{H}_\infty$ качество

Както бе показано в т. 2.3, изискванията към качеството на система с обратна връзка могат обикновено да се формулират като изисквания към функциите на чувствителността или допълнителната чувствителност, или към някакви други предавателни матрици на затворената система. Например, критериите за качество на едномерни системи могат да се формулират чрез

$$\begin{cases} |s(j\omega)| \leq \alpha \ll 1 \text{ за всяко } \omega \leq \omega_0 \\ |s(j\omega)| \leq \beta > 1 \text{ за всяко } \omega > \omega_0 \end{cases}$$

където  $s(j\omega) = \frac{1}{1+p(j\omega)k(j\omega)}$ . Много по-удобно е, обаче, изискванията към качеството да се отразят, избирайки подходящи тегловни функции. Например, горните изисквания към качеството могат да се формулират като

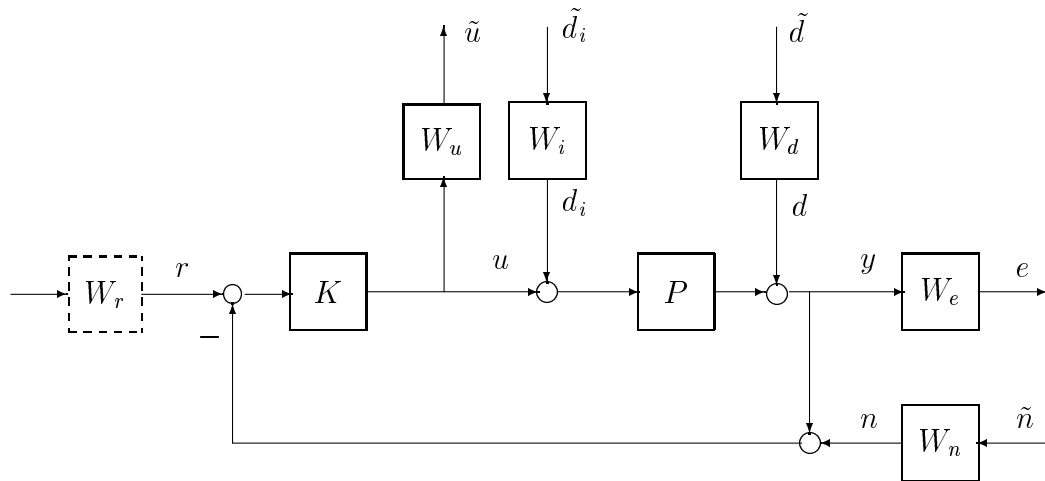
$$|w_s(j\omega)s(j\omega)| \leq 1, \text{ за всяко } \omega$$

с

$$|w_s(j\omega)| = \begin{cases} \alpha^{-1} \text{ за всяко } \omega \leq \omega_0, \\ \beta^{-1} \text{ за всяко } \omega > \omega_0. \end{cases}$$

За да се използва  $w_s$  при синтеза на управляващо устройство, обикновено се избира рационална предавателна функция, за да се апроксимира горепосочената честотна характеристика.

Предимството при използване на претеглени показатели на качеството е очевидно при синтеза на многомерни системи. Преди всичко, някои компоненти на векторния сигнал са обикновено по-важни от останалите. Второ, всяка компонента на сигнала може да не се измерва в една и съща метрична система; например, някои компоненти на изходния сигнал на грешката могат да се измерват в мерни единици за дължина, а други могат да се измерват в мерни единици за напрежение. Следователно, тегловните функции играят съществена роля, за



Фигура 2.9: Стандартна конфигурация с обратна връзка с тегла

да направят тези компоненти сравними. Също така, може да сме заинтересовани от намаляване на грешката в някъкъв честотен диапазон (например ниските честоти), поради което трябва да изберем някакви честотнозависими тегла.

В общият случай ще модифицираме стандартната схема с обратна връзка на фиг. 2.6 в схемата, показана на фиг. 2.9. Тегловните функции на фиг. 2.9 са избрани така, че да отразят целите на синтеза и познаването на смущенията и шума в чувствителния елемент. Например,  $W_d$  и  $W_i$  могат да се изберат така, че да отразят честотния състав на смущенията  $d$  и  $d_i$  или могат да се използват да моделират спектъра на мощността на смущението, в зависимост от природата на сигналите в реалната система. Тегловната матрица  $W_n$  се използва за моделиране честотния състав на шума в чувствителните елементи, а  $W_e$  може да се използва за отразяване на изискванията към формата на изходната функция на чувствителността. По подобен начин  $W_u$  може да се използва, за да се отразят ограниченията върху управляващите или изпълнителните сигнали. Предкомпенсаторът  $W_r$  е елемент, който евентуално се избира за да се постигне подходяща форма на предавателната матрица по задание или да се представи система с неединична обратна връзка в еквивалентна форма с единична обратна връзка.

По-нататък ще бъде показано, че използването на подходящи тегловни матрици е съществено при прилагането на теорията на оптималното управление, по-специално при  $\mathcal{H}_2$  и  $\mathcal{H}_\infty$  теорията. Ето защо много важна стъпка в процеса на синтез на управляващо устройство е да се изберат подходящи тегла  $W_e$ ,  $W_d$ ,  $W_u$  и евентуално  $W_n$ ,  $W_i$  и  $W_r$ . Подходящият избор на тегла за конкретна практическа задача не е тривиален. В много случаи, както при скалярни системи, теглата се избират просто като параметри при синтеза без никакви физически основания, така че тези тегла могат да се третират като настройвани параметри, които се избират от проектанта за да се постигне най-добър компромис между противоречивите изисквания. Изборът на тегловните матрици трябва да се насочва в зависимост от очакваните входове на системата и относителната важност на

изходите.

Следователно, процесът на синтеза на многомерна система може да се разглежда като процес на избор на управляващо устройство  $K$ , такова че известни претеглени сигнали се правят малки в някакъв смисъл. Съществуват различни начини за дефиниране на големината на сигнал или предавателна матрица, както бе показано в т. 2.1. Различните дефиниции водят до различни методи на синтез, които се различават по своята сложност.

По-долу се въвеждат два класа формулировки на качеството:  $\mathcal{H}_2$  и  $\mathcal{H}_\infty$  критерии, които се използват най-често при синтеза на многомерните системи. За опростяване на разглежданията ще предполагаме, че  $d_i = 0$  и  $n = 0$ .

### $\mathcal{H}_2$ качество

Нека предположим, че смущението  $\bar{d}$  може приближено да се моделира като импулс със случайно входно направление, т.е.

$$\bar{d}(t) = \eta \delta(t)$$

и

$$\mathcal{E}(\eta \eta^H) = I$$

където  $\mathcal{E}$  означава математическото очакване. Ние можем да изберем да минимизираме очакваната енергия на грешката  $e$ , дължаща се на смущението  $\bar{d}$ :

$$\mathcal{E} \{ \|e\|_2^2 \} = \mathcal{E} \left\{ \int_0^\infty \|e\|^2 dt \right\} = \|W_e S_o W_d\|_2^2.$$

В общият случай, управляващо устройство, което минимизира само горния критерий, може да доведе до много голям управляващ сигнал  $u$ , който би могъл да предизвика насищане на изпълнителните елементи, както и много други нежелани проблеми. Ето защо за реалистичен синтез е необходимо в целевата функция да се включи и управляващият сигнал. По такъв начин показателят на качеството обикновено е във вида

$$\mathcal{E} \{ \|e\|_2^2 + \rho^2 \|\tilde{u}\|_2^2 \} = \left\| \begin{bmatrix} W_e S_o W_d \\ \rho W_u K S_o W_d \end{bmatrix} \right\|_2^2$$

с подходящо избрана тегловна матрица  $W_u$  и скалар  $\rho$ . Параметърът  $\rho$  очевидно определя компромиса, дискутиран по-горе, между добро подтискане на смущението на изхода и управляващото усилие (или подтискането на смущението и шума на чувствителния елемент в точката на изпълнителните устройства). Отбележете, че  $\rho$  може да се положи равно на единица посредством подходящ избор на  $W_u$ . Тази задача може да се разглежда като минимизиране на енергията, консумирана от системата, за да се подтисне смущението  $d$ .

Този тип от задачи бе доминираща тематика през шестдесетте и седемдесетте години и обикновено е известен под името *линейно-квадратично гаусово управление* или LQG. Разработването на тази тематика доведе до редица нови



резултати, особено в областта на оценяването. Теоретичните приноси включват по-дълбоко разбиране на същността на линейните системи и усъвършенствувани числени методи за комплексни системи, основани на подхода на пространство на състоянията. Най-голямото ограничение на тази теория е отсъствието на формално третиране на неопределеността в самия обект. Тази неизвестност играе важна роля в системите с обратна връзка.

### $\mathcal{H}_\infty$ качество

Въпреки че  $\mathcal{H}_2$  нормата (или  $\mathcal{L}_2$  нормата) може да бъде значима мярка на качеството и въпреки че LQG-теорията може да даде ефективен компромис при синтеза при някои предположения за смущението и обекта,  $\mathcal{H}_2$  нормата страда от един голям недостатък. Този недостатък се дължи на факта, че компромиса между намаляване на грешката от смущението и намаляване на грешката от шума на чувствителния елемент не е единственото ограничение при синтеза на система с обратна връзка. Проблемът се състои в това, че тези компромиси по отношение на качеството често стават в сянката на едно друго ограничение на високите коефициенти на усилване на контура, а именно на изискването за допустимост на неопределеностите. Свойствата на физическите системи поставят строги ограничения върху честотните диапазони, където усилванията в контура могат да са високи.

Едно решение на тази задача е да се поставят явни ограничения върху усилванията на контура в наказателната функция. Например, можем да изберем да минимизираме

$$\sup_{\|\tilde{d}\|_2 \leq 1} \|e\|_2 = \|W_e S_o W_d\|_\infty,$$

при някои ограничения върху енергията на управлението или честотната лента на управляващото устройство:

$$\sup_{\|\tilde{d}\|_2 \leq 1} \|\tilde{u}\|_2 = \|W_u K S_o W_d\|_\infty.$$

По-често, би могло да се въведе параметър  $\rho$  и смесен критерий

$$\sup_{\|\tilde{d}\|_2 \leq 1} \{\|e\|_2^2 + \rho^2 \|\tilde{u}\|_2^2\} = \left\| \begin{bmatrix} W_e S_o W_d \\ \rho W_u K S_o W_d \end{bmatrix} \right\|_\infty^2.$$

Алтернативно, ако най-същественният проблем е осигуряването на робастна устойчивост, може да се ограничи претеглената допълнителна чувствителност. По такъв начин цялата функция на загубите може да е

$$\left\| \begin{bmatrix} W_e S_o W_d \\ \rho W_1 T_o W_2 \end{bmatrix} \right\|_\infty^2,$$

където  $W_1$  и  $W_2$  са честотно зависимите матрици, мащабиращи неопределеността. Тези задачи на синтеза обикновено се наричат  $\mathcal{H}_\infty$  задачи на смесената чувствителност. За скаларна система, задачата за минимизация на  $\mathcal{H}_\infty$  нормата може също да се разглежда като минимизиране на максимума на амплитудата на установената реакция на системата по отношение на най-лошите синусоидални входи.



## Глава 3

# НЕОПРЕДЕЛЕННОСТ И РОБАСТНОСТ

В тази глава са изложени различни начини за представяне на неопределеността в модела на многомерната система, въз основа на които се получават условия за робастност на устойчивостта (т. 3.2) и робастност на качеството на затворената система (т. 3.3). В т. 3.4 е разгледан метод за синтез на оптимално управляващо устройство, което осигурява робастност на качеството на затворената система в присъствието на неопределеност.

### 3.1 Представяне на неопределеността

Терминът *неопределеност* касае разликите или грешките между моделите и реалността и който и да е механизъм, използван за изразяване на тези грешки, се нарича *представяне на неопределеността*.

#### Неструктурирана неопределеност

При неструктурираната неопределеност се ограничава големината на възможните смущения, но смущенията не се свързват с конкретни елементи на обекта. (Обърнете внимание, че тук под *смущения* се имат предвид параметрични смущения, а не смущаващи въздействия.)

Трите най-често използвани модели на неструктурирана неопределеност за многомерен обект са следните. Нека  $P_0(s)$  е номиналната предавателна матрица, която представлява най-добрата в някъкъв смисъл оценка на истинското поведение на обекта и нека  $P(s)$  означава истинската предавателна матрица на обекта. Тогава

$$P(s) = P_0(s) + \Delta_a(s) \quad (3.1)$$

$$P(s) = P_0(s)[I + \Delta_i(s)] \quad (3.2)$$

или

$$P(s) = [I + \Delta_o(s)]P_o(s) \quad (3.3)$$

където  $\Delta_a$  представлява *адитивно смущение*,  $\Delta_i$  е *входно мултипликативно смущение* и  $\Delta_o$  е *изходно мултипликативно смущение*.

Единственото ограничение върху тези смущения е тяхната "големина", която се измерва с  $\|\Delta\|_\infty$ . Ако желаем да направим големината честотно зависима, то можем да използваме  $\bar{\sigma}(\Delta)$  или еквивалентно можем да положим  $\Delta = W_1 \tilde{\Delta} W_2$ , където  $W_1$  и  $W_2$  са минимално фазови предавателни матрици (без полюси и нули в дясната комплексна полуравнина), които служат като честотно зависими тегловни функции. В този случай винаги може да се вземе  $\|\tilde{\Delta}\|_\infty \leq 1$ .

Адитивният модел (3.1) може да се използва при решаването на някои задачи за робастна стабилизация, които имат лесно решение, но мултипликативните модели (3.2) и (3.3) често са по-реалистични, тъй като  $\|\Delta_i\|_\infty$  и  $\|\Delta_o\|_\infty$  представляват относителни, а не абсолютни големина. Например,  $\|\Delta_i\|_\infty \leq 0.1$  означава, че големината на смущението е най-много 10% от "големината" на  $P_0$ , тъй като

$$\begin{aligned} \|P - P_0\|_\infty &= \|P_0 \Delta_i\|_\infty \\ &\leq \|P_0\|_\infty \|\Delta_i\|_\infty \\ &\leq 0.1 \|P_0\|_\infty \end{aligned}$$

Задавайки обаче  $\|\Delta_a\|_\infty \leq 0.1$  получаваме че

$$\begin{aligned} \|P - P_0\|_\infty &= \|\Delta_a\|_\infty \\ &\leq 0.1. \end{aligned}$$

При описанието на мултипликативни смущения се нуждаем и от двата модела (3.2) и (3.3), тъй като умножението на предавателни матрици не е комутативно. Представянето (3.2) предполага, че цялата неопределеност възниква на входа на обекта, докато (3.3) означава, че цялата неопределеност възниква на изхода. Използването на две тегловни предавателни матрици ( $W_1$  и  $W_2$ ) вместо една прави възможно границата върху смущението да бъде винаги нормализирана до

$$\|W_1^{-1} \Delta W_2^{-1}\|_\infty \leq 1.$$

В някои случаи не е подходящо да се отнася цялата неопределеност към входа или изхода. Тогава е необходимо да се използва или адитивния модел (3.1) или някъкъв мултипликативен модел, като например

$$P(s) = [I + \Delta_o(s)] P_0(s) [I + \Delta_i(s)]$$

или

$$P(s) = P_0(s) \cdot [I + \Delta(s)]$$

където " $\cdot$ " представлява *произведението на Шур*, т.е. поелементно умножаване. Други възможни модели са *обратният мултипликативен модел*

$$P(s) = [I + \Delta_R(s)]^{-1} P_0(s)$$

и модел, основан на описанието с матрични дроби

$$P(s) = N_r(s) M_r^{-1}(s) \tag{3.4}$$

където

$$N_r(s) = N_{r0}(s) + \Delta_N(s) \text{ и } M_r(s) = M_{r0}(s) + \Delta_M(s). \quad (3.5)$$

Моделите (3.4) и (3.5) могат да се разширят до т.н. *линейни дробни представления* където  $N_r(s)$  и  $M_r(s)$  могат да са устойчиви предавателни матрици вместо полиномиални матрици.

Голямо преимущество при използването на операторни норми за описване неопределеността на модела е, че не е необходимо да се постулира съществуването на "истинска" предавателна матрица на модела. Реалните обекти винаги са нелинейни и често нестационарни, така че не може да съществува "истинско" описание с предавателна матрица. Ние обаче можем да интерпретираме модела на неопределеността, например (3.1), по следния начин. Да предположим, че прилагаме вход  $u(t)$  към обекта, такъв че  $\|u\|_2 = 1$  и че полученият изход е  $y(t)$ . Да предположим също така, че същият вход, приложен към модел с предавателна матрица  $P_0(s)$ , дава изхода  $y_0(t)$ . Тогава (3.1) означава, че

$$\|y - y_0\|_2 \leq \|\Delta_a\|_\infty.$$

### Структурирана неопределеност

В някои случаи разполагаме с информация за *структурата* на входната или изходната неопределеност, като неопределеността в даден елемент е независима от неопределеността в останалите елементи. Например, входната неопределеност може да бъде зададена във вида

$$\Delta_i = \begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix}, \quad |\delta_1| \leq 0.1, \quad |\delta_2| \leq 0.1.$$

Ако вместо това представяне на неопределеността сега запишем  $\|\Delta_i\|_\infty \leq 0.1$ , ние ще загубим цялата структурна информация, тъй като този запис позволява също смущения като

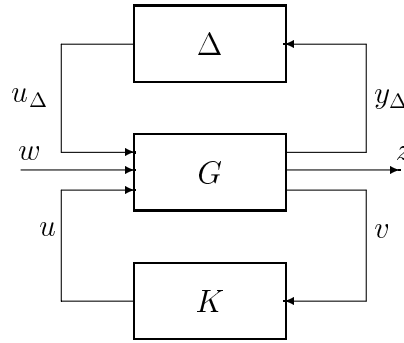
$$\Delta_i = \begin{bmatrix} 0 & 0.1 \\ 0 & 0 \end{bmatrix}$$

и

$$\Delta_i = \frac{1}{\sqrt{2}} \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \end{bmatrix},$$

които не отговарят на никакви реални смущения. Ето защо използването на неструктурирано описание в общия случай води до синтез на управляващи устройства, които са ненужно консервативни, тъй като те работят удовлетворително (в някакъв смисъл) дори при смущения, които никога не възникват.

На практика може да е налична както структурирана, така и неструктурирана информация за неопределеността в обекта. Например, могат да са известни конкретни параметри в модела в пространство на състоянията, които могат да се изменят в известни граници, което е структурирана информация, като в същото време може да е известно, че всички елементи на модела стават извънредно неточни над известна честота поради немоделирани закъснения, нелинейности,



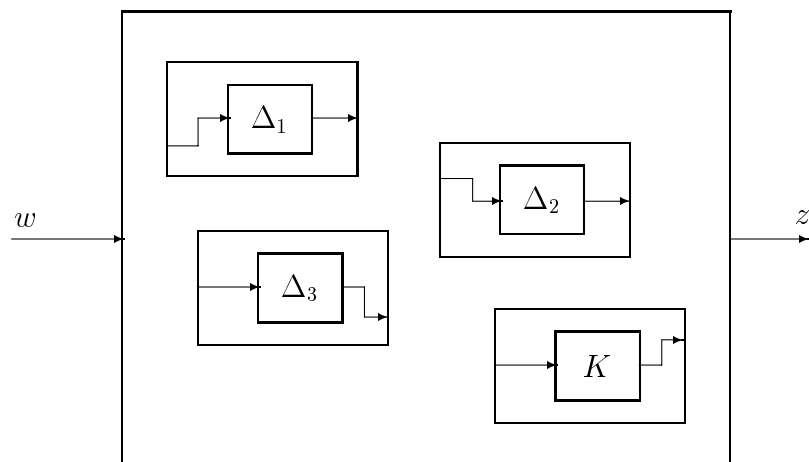
Фигура 3.1: Стандартно представяне на неточно известен обект с обратна връзка

резонанс, паразитни връзки и т.н., което е неструктурирана информация, тъй като практически не е възможно всичко това да се опише в подробности. Ето защо проблемът е как да се обединят двата вида информация едновременно в точно описание на неизвестността на обекта.

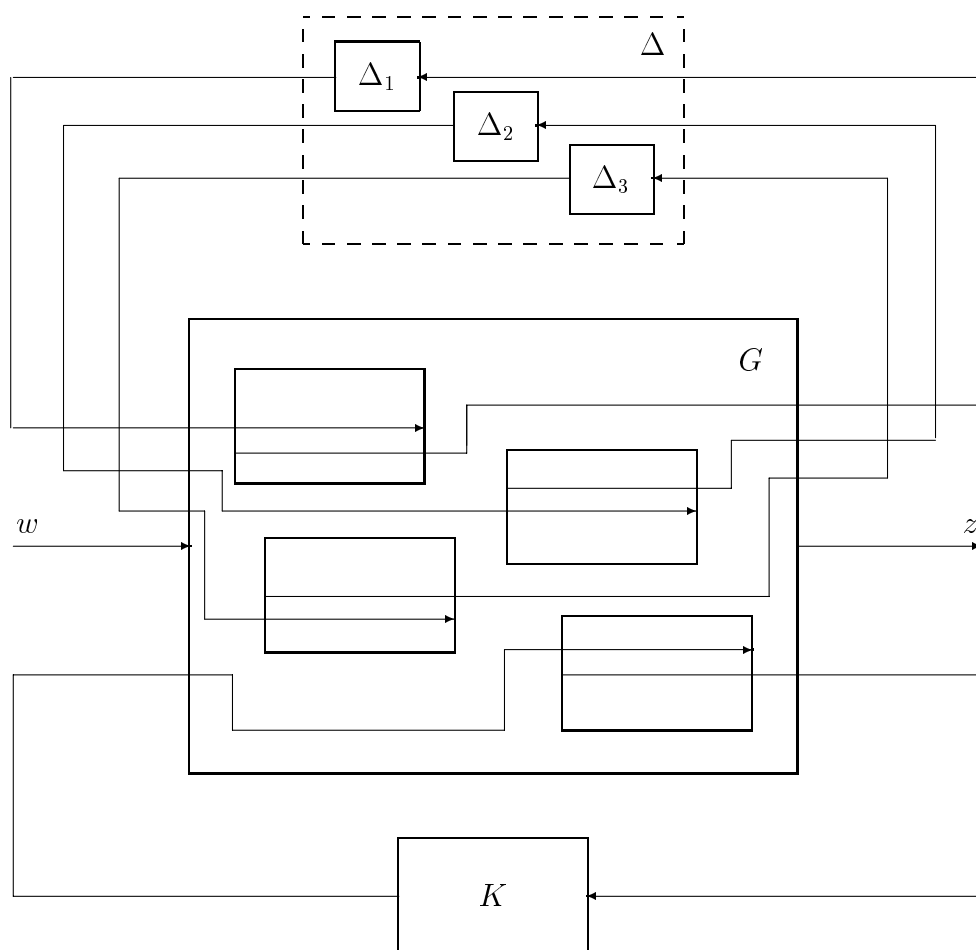
На практика винаги е възможно да се представи много точно информацията за неопределеността по следния начин. Разглеждаме обект с три множества от входове и три множества от изходи, както е показано с блока  $G$  на фиг. 3.1.

Първото множество от входове се състои от всички управляеми входни променливи, представени с вектора  $u$ , а второто множество се състои от всички други външни сигнали, такива като заданията, смущенията и шумовете от измерване, представени с вектора  $w$ . Първото множество от изходи се състои от всички измервани сигнали, представени с вектора  $v$ , които са достъпни за подаване към управляващото устройство  $K$ , а второто множество се състои от всички други изходи, представени с вектора  $z$ , чието поведение може да представлява интерес (те може да не са реални сигнали, които могат да се измерват в някаква точка на обекта, но могат да включват концептуални сигнали, такива като сигнали на грешката). Фиг. 3.1 показва управляващо устройство  $K$ , свързано между първата двойка от изходи и входове.

Третото множество от входове  $u_\Delta$  и изходи  $y_\Delta$  на обекта  $G$  е ново. Където и да има неизвестност относно някаква част на системата, ние бихме могли да я представим "локално", използвайки някое от описанията, разгледани във връзка с неструктурираната неопределеност. Например, ако параметърът  $\theta$  има номинална стойност  $\theta_0$  с неопределеност от 20%, то  $\theta = \theta_0(1 + \delta)(|\delta| \leq 0.2)$ . По такъв начин можем да получим блок-схема на обекта с редица  $\Delta$ -блокове, разположени около него. Всеки от тези блокове има вход (който може да е скаларен или векторен) и изход. Събирайки всички тези входове заедно, получаваме третото множество  $y_\Delta$  от изходи на  $G$ . По подобен начин събираме всички изходи заедно, за да формираме третото множество  $u_\Delta$  от входове на  $G$ . Този процес на преобразуване на неопределената система е показан на фиг. 3.2. В резултат всички  $\Delta$ -блокове се изваждат навън от обекта; на фиг. 3.2 (б) те се съдържат вътре в блока, означен с  $\Delta$ . Разглеждан като предавателна матрица, този външен блок  $\Delta$  има твърде специална структура: тя е блок-диагонална и блоковете



(а) Първоначалната система с няколко смущения



(б) Изваждане на смущенията

Фигура 3.2: Преобразуване на неопределена система в стандартен вид

върху диагонала са точно малките  $\Delta$ -и, които са "извадени" от вътрешността на обекта. Записваме

$$\Delta(s) = \text{diag} \{ \Delta_1(s), \dots, \Delta_n(s) \}$$

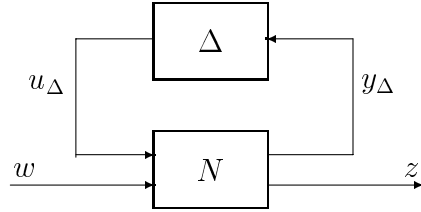
ако  $n$  такива блокове са "извадени" навън. Отбележете, че всяко  $\Delta_j$  може да е скалар или матрица.

По-нататък може да се извърши едно усъвършенствуване на този модел по следния начин. Въвеждаме мащабирания в обекта, така че всички неопределености се нормализират във вида

$$\|\Delta_j\|_\infty \leq 1.$$

Това винаги е възможно и улеснява формулирането на стандартни условия за постигане на робастен синтез. Да отбележим, че номиналният модел на обекта се получава полагайки  $\Delta = 0$ .

Ако управляващото устройство  $K$  вече е известно, то може да се комбинира с някаква част от обекта, за да се получи обща система  $N$ , както е показано на фиг. 3.3. От това става ясно, че всяка неопределеност, свързана с управляващото устройство, може да се моделира по същия начин както неопределеността в обекта, което ще доведе до добавяне на някакви блокове към  $\Delta$ .



Фигура 3.3: Обект  $G$  и управляващо устройство  $K$ , комбинирани в обща система  $N$

Нека предавателната матрица на обобщения обект  $G$  е разделена във вида

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix},$$

така че нейните блокове са съвместими със сигналите  $w, z, u$  и  $v$  в стандартното представяне на фиг. 3.1,

$$z = G_{11}w + G_{12}u, \quad (3.6)$$

$$v = G_{21}w + G_{22}u. \quad (3.7)$$

Отбележете, че  $G_{22}$  има размерности, съвместими с тези на управляващото устройство, т.е. ако  $K$  е  $n_u \times n_v$  матрица, то  $G_{22}$  е  $n_v \times n_u$  матрица.

За да се намери  $N$ , уравненията на обобщения обект се комбинират с уравнението на управляващото устройство

$$u = Kv \quad (3.8)$$



и се елиминират  $u$  и  $v$  от уравненията (3.6), (3.7) и (3.8). В резултат се получава  $z = Nw$ , където  $N$  се дава от

$$N = G_{11} + G_{12}K(I - G_{22}K)^{-1}G_{21} \stackrel{\text{def}}{=} F_\ell(G, K).$$

Тук  $F_\ell(G, K)$  означава долно *линейно дробно преобразуване (ЛДП)* на  $G$  и  $K$ . Това е еквивалентно на използването на  $K$  за затваряне на долния контур на фиг. 3.1 около  $G$ . Тъй като в стандартната конфигурация на фиг. 3.1 се използва положителна обратна връзка, членът  $(I - G_{22}K)^{-1}$  съдържа отрицателен знак.

За да се изчисли смутената (неопределената) предавателна матрица от външните входове  $w$  към външните изходи  $z$ , използваме  $\Delta$  за да затворим горния контур около  $N$  (вж. фиг. 3.3), което дава като резултат *горното линейно дробно преобразуване*

$$z = F_u(N, \Delta)w, \quad F_u(N, \Delta) \stackrel{\text{def}}{=} N_{22} + N_{21}\Delta(I - N_{11}\Delta)^{-1}N_{12}.$$

Линейните дробни преобразувания се разглеждат в Приложение П.7.

**Пример 3.1** Да разгледаме входно/изходното съотношение

$$z = \frac{a + b\delta_2 + c\delta_1\delta_2^2}{1 + d\delta_1\delta_2 + e\delta_1^2}w = Gw,$$

където  $a, b, c, d$  и  $e$  са дадени константи или предавателни функции. Искаме да представим  $G$  като линейно дробно преобразуване в термините на  $\delta_1$  и  $\delta_2$ . Това може да се извърши на три стъпки:

1. Начертава се блок-схема за входно/изходното съотношение, в която всяко  $\delta$  е отделено, така както е показано на фиг. 3.4.
2. Отбелязват се входовете и изходите на  $\delta$ -тите като  $y$ -и и  $u$ -та, съответно. (Това всъщност е изваждане навън на  $\delta$ -тите.)
3. Записват се  $z$ -вете и  $y$ -ите в термините на  $w$ -тата и  $u$ -тата, като всички  $\delta$ -и са извадени. (Тази стъпка е еквивалентна на изчисляване на преобразуването в блока  $G$  на фиг. 3.2.)

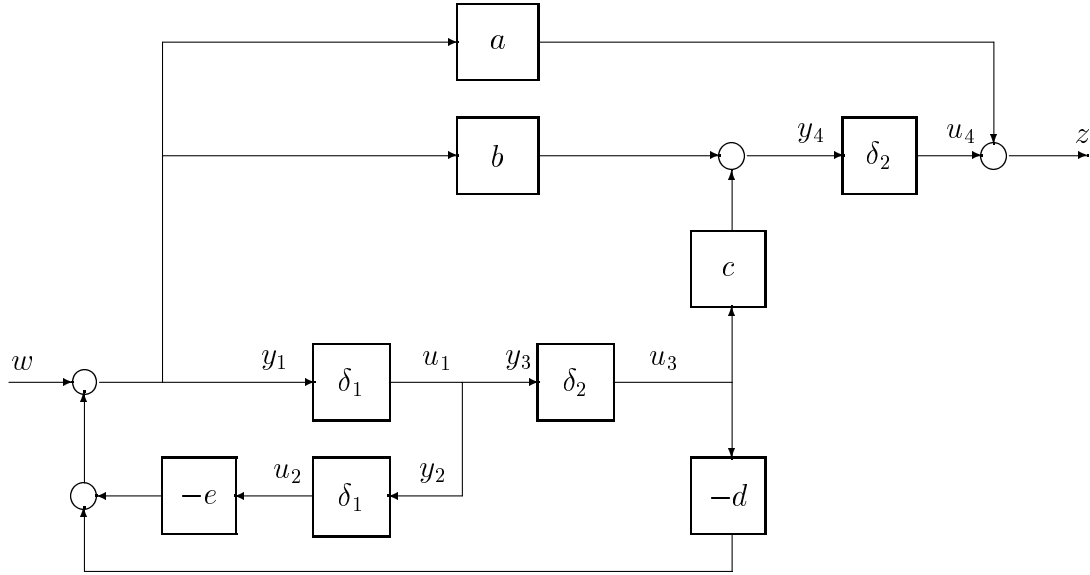
$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ z \end{bmatrix} = N \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ w \end{bmatrix},$$

където

$$N = \left[ \begin{array}{ccc|c|c} 0 & e & d & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & be & bd + c & 0 & b \\ \hline 0 & ae & ad & 1 & a \end{array} \right].$$

Тогава

$$z = F_u(N, \Delta)w, \quad \Delta = \begin{bmatrix} \delta_1 I_2 & 0 \\ 0 & \delta_2 I_2 \end{bmatrix}.$$



Фигура 3.4: Блок-схема на  $G$

## 3.2 Робастност на устойчивостта

След като имаме средства за описване на неопределеността в модела на системата, сега можем да насочим нашето внимание към начините за проверяване дали конкретна конфигурация с обратна връзка е робастна в присъствието на тази неопределеност. Фундаменталното свойство, което трябва да се запазва при всички възможни смущения в параметрите на обекта (и на управляващото устройство) е устойчивостта на системата с обратна връзка.

### Неструктурирана неопределеност

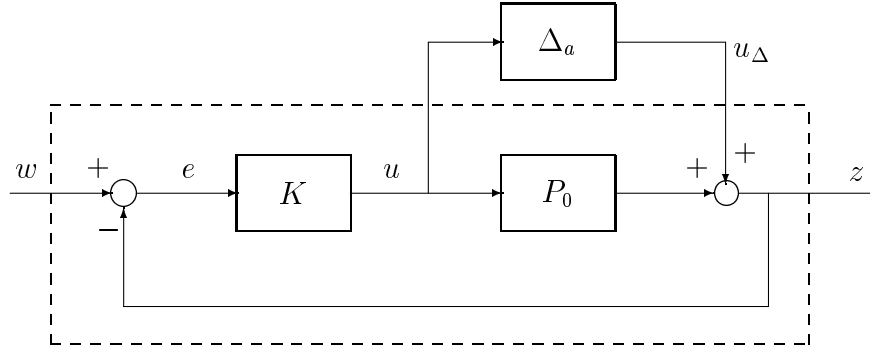
Да разгледаме адитивния модел на неструктурираната неопределеност, дефиниран с (3.1):

$$P = P_0 + \Delta_a$$

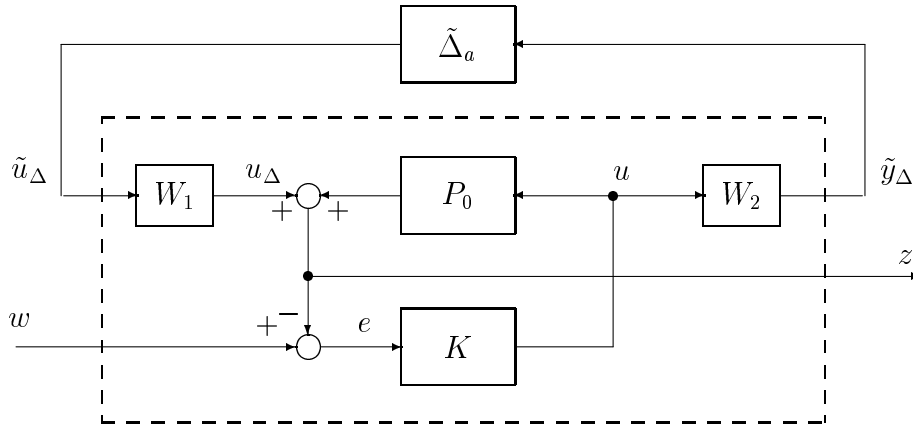
с

$$\Delta_a = W_1 \tilde{\Delta}_a W_2 \text{ и } \|\tilde{\Delta}_a\|_\infty \leq 1.$$

Да предположим, че синтезираме управляващо устройство  $K$ , който стабилизира номиналния модел  $P_0$ . Тогава контурът с действителния обект е показан на фиг. 3.5 и той може да се преобразува в стандартния вид с  $\tilde{\Delta}_a$  разположено в контура на обратната връзка, както е показано на фиг. 3.6. Отбележете, че и на двете фигури прекъснатата линия затваря устойчива система.



Фигура 3.5: Контур на обратната връзка, съдържащ обект с адитивна неопределеност



Фигура 3.6: Контурът на обратната връзка от фиг. 3.5, преобразуван във вида от фиг. 3.3

Ако запишем

$$\begin{bmatrix} \tilde{y}_\Delta \\ z \end{bmatrix} = N \begin{bmatrix} \tilde{u}_\Delta \\ w \end{bmatrix} = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} \tilde{u}_\Delta \\ w \end{bmatrix}$$

където  $\tilde{y}_\Delta, w, z$  и  $\tilde{u}_\Delta$  са както са дефинирани на фиг. 3.5 и 3.6, то лесно може да се покаже, че

$$\begin{aligned} N_{11} &= -W_2(I + KP_0)^{-1}KW_1, \\ N_{12} &= W_2(I + KP_0)^{-1}K, \\ N_{21} &= (I + P_0K)^{-1}W_1 \end{aligned}$$

и

$$N_{22} = P_0K(I + P_0K)^{-1}.$$

За да се провери дали системата с обратна връзка ще остане устойчива при допустими смущения, е необходимо да се провери дали съединението с обратна връзка на  $N_{11}$  и  $\tilde{\Delta}_a$  ще остане устойчиво за всички допустими  $\tilde{\Delta}_a$ . При това матрицата  $N_{11}$  е устойчива (тъй като  $K$  стабилизира  $P_0$ ) и *предполагаме, че*

$\tilde{\Delta}_a$  е устойчива. При това предположение може да се покаже, че системата с обратна връзка ще остане устойчива, ако

$$\bar{\sigma}(N_{11}\tilde{\Delta}_a) < 1$$

за всяка честота, или еквивалентно ако

$$\|N_{11}\tilde{\Delta}_a\|_{\infty} < 1.$$

Тъй като

$$\|N_{11}\tilde{\Delta}_a\|_{\infty} \leq \|N_{11}\|_{\infty}\|\tilde{\Delta}_a\|_{\infty}$$

и

$$\|\tilde{\Delta}_a\|_{\infty} \leq 1,$$

неравенството

$$\|N_{11}\|_{\infty} < 1 \tag{3.9}$$

дава едно достатъчно условие за гарантиране на устойчивостта при всички възможни смущения.

Ако  $W_1 = I$ ,  $W_2 = I$  и  $\|\Delta_a\|_{\infty} \neq 1$ , условието за робастна устойчивост при адитивна неопределеност се записва като

$$\bar{\sigma}(\Delta_a) < \frac{1}{\bar{\sigma}(R)}, \quad R = (I + KP_0)^{-1}K = KS,$$

където  $S = (I + P_0K)^{-1}$  е изходната чувствителност.

Ако вместо адитивно смущение се разглежда случая на мултипликативно смущение (3.2) или (3.3), условието (3.9) отново остава в сила, но предавателната матрица  $N$  се променя.

Ако предположението за устойчивост на смущението е нереалистично, може да се използва друг модел на неструктурираната неопределеност.

За някои модели на неопределеността изискването за устойчивост на смущението може да се отслаби. Да разгледаме моделът на изходното мултипликативно смущение (3.3). Ще допуснем, че  $\Delta_o$  е неустойчива, но ще наложим алтернативното изискване номиналният обект  $P_0$  и смутеният обект  $P$  да имат *един и същ брой неустойчиви полюси*. При това предположение може да се докаже, че системата ще остане устойчива, ако

$$\bar{\sigma}(\Delta_o) < \frac{1}{\bar{\sigma}(T)}, \quad T = P_0K(I + P_0K)^{-1} = I - S$$

или, еквивалентно, ако

$$\|\Delta_o\|_{\infty} < \frac{1}{\|T\|_{\infty}}. \tag{3.10}$$

Този резултат показва, че  $\mathcal{H}_{\infty}$ -нормата на допустимото от гледна точка на робастната устойчивост изходно мултипликативно смущение е обратно пропорционална на  $\mathcal{H}_{\infty}$ -нормата на предавателната матрица на затворената система между изхода  $z$  и външния вход  $w$ .

Условията за робастна устойчивост от типа (3.9) и (3.10) са специален случай от по-общ резултат, известен като *теорема за малкото усиление*, която твърди, че контур с обратна връзка, състоящ се от устойчиви оператори, ще остава устойчив, ако произведението на всички операторни усиления е по-малко от единица. Този резултат е в сила както за линейни, така и за нелинейни оператори.

### Структурирана неопределеност

Разглеждаме структурирана неопределеност, така както е показана на фиг. 3.3;  $\Delta$  има блок-диагонална структура, като всеки блок  $\Delta_j$  удовлетворява  $\|\Delta_j\|_\infty \leq 1$ . Ще предположим също така, че всяка предавателна матрица  $\Delta_j$  е устойчива. Разделяйки на блокове коригираната система  $N$  както преди това, намереното по-рано условие

$$\|N_{11}\|_\infty < 1 \quad (3.11)$$

е достатъчно за робастна устойчивост. Това условие обаче вече не е необходимо, тъй като повечето смущения, които удовлетворяват  $\|\Delta\|_\infty \leq 1$ , вече не са допустими: единствено допустими са тези, които имат подходящата блок-диагонална структура. Освен това оценката (3.11) в някои случаи може да е толкова консервативна, че да е безполезна. Следователно, трябва да се намери друга оценка, която да дава необходимо условие за робастна устойчивост, за разлика от (3.11).

Нека  $BD_\delta$  означава множеството на устойчивите блок-диагонални смущения с конкретна структура и с  $\|\Delta_j\|_\infty \leq \delta$ . За да се опише това множество напълно за всяка задача за синтез, би трябвало да се използва означение от вида  $BD_\delta(m_1, m_2, \dots, m_n, k_1, k_2, \dots, k_n)$ , за да се запише, че  $\Delta$  съдържа  $m = \sum_{i=1}^n m_i$  блокове, като всеки блок се повтаря  $m_i$  пъти и има размерности  $k_i \times k_i$ . (Блоковете, които се повтарят, позволяват да се моделират корелирани смущения.) Обикновено, обаче, ще използваме съкратено означение, с подходяща структура, която се подразбира че е фиксирана.

Дадено допустимо смущение дестабилизира системата, ако и само ако

$$\det[I - N_{11}(j\omega)\Delta(j\omega)] = 0$$

за някакво  $\omega$  и някакво  $\Delta \in BD_1$ . Ето защо дефинираме

$$\mu(N_{11}(j\omega)) = \begin{cases} 0 & \text{ако } \det[I - N_{11}\Delta] \neq 0 \text{ за всяко } \Delta \in BD_\infty \\ \{\min_{\Delta \in BD_\infty} (\bar{\sigma}(\Delta(j\omega)) : \det[I - N_{11}(j\omega)\Delta(j\omega)] = 0)\}^{-1} & \\ \text{в противния случай} \end{cases} \quad (3.12)$$

Дефинираме също така величината

$$\|N_{11}\|_\mu = \sup_\omega \mu(N_{11}(j\omega)).$$

Функцията  $\mu(N)$  се нарича *структурно сингулярно число* на  $N$  и зависи както от структурата на множеството  $BD_\delta$ , така и от  $N$ . Трябва да се има предвид, че в (3.12)  $\|\cdot\|_\mu$  не е норма.

Въпреки че изразът (3.12) дава определение за  $\mu(\cdot)$ , той не дава начин за изчисляването на тази величина. Възможно е да се покажат следните важни нейни свойства:

$$\mu(\alpha N) = |\alpha| \mu(N),$$

$$\mu(I) = 1,$$

$$\mu(AB) \leq \bar{\sigma}(A) \mu(B), \quad A, B \text{ са квадратни,}$$

$$\mu(\Delta) = \bar{\sigma}(\Delta), \Delta \in BD_\delta.$$

Ако  $n = 1$  и  $m_1 = 1$  (вж. по-горе), то  $\mu(N) = \bar{\sigma}(N)$ .

Ако  $n = 1$  и  $k_1 = 1$  (вж. по-горе), то  $\mu(N) = \rho(N)$ , където  $\rho(N) = \max_i |\lambda_i(N)|$  е спектралният радиус на  $N$ .

Нека  $D$  е произволна диагонална, реална положителна матрица със структурата

$$\text{diag}\{d_1 I_{k_1}, \dots, d_{m_1} I_{k_1}, d_{m_1+1} I_{k_2}, \dots, d_{m_n} I_{k_n}\}$$

и  $d_i > 0$ . Тогава

$$\mu(DND^{-1}) = \mu(N).$$

Нека  $U$  е унитарна матрица ( $UU^H = I$ ) със същата блок-диагонална структура, както множеството  $BD_\delta$ . Тогава

$$\max_U \rho(UN) \leq \mu(N) \leq \inf_D \bar{\sigma}(DND^{-1}). \quad (3.13)$$

Важността на  $\mu(\cdot)$  може да се види от следната теорема.

**Теорема 3.1 (Робастна устойчивост при структурирано смущение)**  
Системата, показана на фиг. 3.3, остава устойчива за всички  $\Delta \in BD_1$ , ако и само ако

$$\|N_{11}\|_\mu < 1.$$

Теорема 3.1 в действителност не дава тест за робастна устойчивост, понеже до момента не разполагаме със средство за изчисляване на  $\|N_{11}\|_\mu$ . Изразът (3.13), обаче, дава възможност да се изчислят приближения на  $\|N_{11}\|_\mu$ . Възможно е, например, първото неравенство в (3.13) да се усили до

$$\max_U \rho(UN) = \mu(N),$$

но задачата за намиране на максимума в този случай обикновено води до локален максимум. В действителност второто неравенство в (3.13) е по-подходящо за използване, още повече, че е опасно да се недооцени  $\mu$ , но не е опасно то да се преоцени.

Ако блоковете в  $\Delta$  са до три и не се повтарят, то може да се покаже, че

$$\mu(N) = \inf_D \bar{\sigma}(DND^{-1}). \quad (3.14)$$

Така че, ако имаме до три некорелирани смущения, то (3.14) е в сила. В случая размерността на всеки блок не е ограничена. При повече от три блока числените

експерименти показват, че долната и горната граница в (3.14) обикновено се различават с не повече от 5%, като почти винаги разликата е не повече от 15%. По такъв начин  $\inf \bar{\sigma}(DND^{-1})$  е добра оценка на  $\mu$  и може успешно да се използва при изчисленията.

За определяне на  $\mu$  може да се използва оптимизационна процедура, която намира  $D$  минимизиращо  $\bar{\sigma}(DND^{-1})$  за всяка честота. Тази процедура е привлекателна, тъй като  $\bar{\sigma}(DND^{-1})$  е изпъкнало по отношение на  $D$ .

### 3.3 Робастност на качеството

В предишната точка бе разгледана робастността на устойчивостта в присъствие на различни смущения. На практика сме заинтересувани от нещо повече. Ние бихме желали поддържането на приемливо ниво на качеството, както и устойчивостта, в присъствието на определен клас от смущения, т.е. бихме желали да постигнем *робастност на качеството*, както и робастност на устойчивостта.

В т. 2.4 бе показано, че е възможно да се формулират изисквания към качеството в честотната област в термините на операторната норма  $\|\cdot\|_\infty$ . Използвайки стандартното представяне на неопределеността, показано на фиг. 3.3, разделяме предавателната матрица  $N$  така че

$$\begin{bmatrix} y_\Delta \\ z \end{bmatrix} = N \begin{bmatrix} u_\Delta \\ w \end{bmatrix} = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \begin{bmatrix} u_\Delta \\ w \end{bmatrix} \quad (3.15)$$

и да дефинираме променливите  $w$  и  $z$ , така че

$$\|N_{22}\|_\infty < 1$$

да стане изискване към качеството.

Нека например да предположим, че изискването към качеството е

$$\left\| \begin{bmatrix} W_1 S \\ W_2 T \end{bmatrix} \right\|_\infty < 1$$

където  $S$  е функцията на изходната чувствителност,  $T$  е допълнителната изходна чувствителност ( $T = I - S$ ), и  $W_1$  и  $W_2$  са честотно зависими тегловни матрици. Тогава  $w$  и  $z = [z_1 \ z_2]^T$  се дефинират както на фиг. 3.7, на която  $P$  и  $K$  са *смутеният* обект и управляващо устройство, съответно.

При наличието на смущение  $\Delta$ , както е показано на фиг. 3.3, връзката между  $w$  и  $z$  става

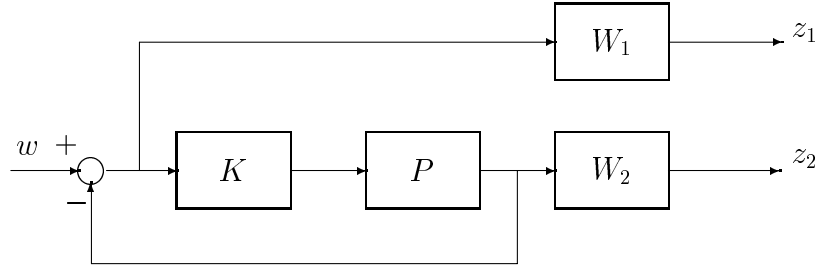
$$z = [N_{22} + N_{21}\Delta(I - N_{11}\Delta)^{-1}N_{12}]w,$$

така че критерият за робастно качество може да се зададе точно като

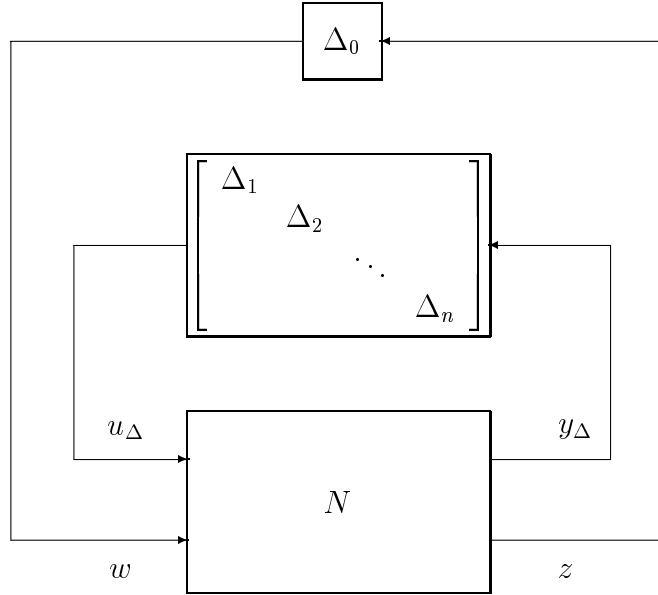
$$\|N_{22} + N_{21}\Delta(I - N_{11}\Delta)^{-1}N_{12}\|_\infty < 1 \quad (3.16)$$

и

$$\|N_{11}\|_\mu < 1, \quad (3.17)$$



Фигура 3.7: Дефиниране на променливите  $w$  и  $z$  в уравнение (3.15)



Фигура 3.8: Заместване на изискване към качеството с фиктивна неопределеност  $\Delta_0$

където  $\Delta = \text{diag} \{\Delta_1, \dots, \Delta_n\}$  и  $\|\Delta_j\| \leq 1$  и (3.17) идва от Теорема 3.1.

Да предположим, че вместо налагане на изискване към качеството, сме прибавили допълнителен неопределен блок  $\Delta_0$  между  $z$  и  $w$  с  $\|\Delta_0\|_\infty \leq 1$ , така както е показано на фиг. 3.8, и изискваме робастна устойчивост в присъствието на кое да е произволно смущение, включително фиктивното  $\Delta_0$ . Иначе казано, да предположим, че изискваме

$$\|N\|_\mu < 1 \quad (3.18)$$

където  $\|\cdot\|_\mu$  сега е изчислено по отношение на структурата

$$\text{diag} \{\Delta_0, \Delta\} = \text{diag} \{\Delta_0, \Delta_1, \dots, \Delta_n\}.$$

Тогава, тъй като устойчивостта е изисквана за смущения  $\text{diag} \{\Delta_0, 0\}$  и  $\text{diag} \{0, \Delta\}$ , (3.18) ще означава, че и двете условия  $\|N_{11}\|_\infty < 1$  и  $\|N_{22}\|_\mu < 1$  са удовлетворени. Много съществен факт е, че (3.18) също така означава, че (3.16) се удовлетворява за всички допустими  $\Delta$ ; нещо повече, (3.18) е в действителност необходимо условие (3.16) да е в сила.



Този резултат е от голяма важност в анализа на многомерните системи и открива пътя към систематични методи за синтез, както е показано по-нататък.

**Теорема 3.2 (Робастност на качеството)** *Условието (3.16) е в сила за всяко  $\Delta \in BD_1$  и (3.17) се удовлетворява (изчислено по отношение структурата на  $\Delta$ ), ако и само ако*

$$\|N\|_\mu < 1,$$

където

$$N = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}$$

и  $\|N\|_\mu$  е изчислено по отношение на структурата на  $\text{diag}\{\Delta_0, \Delta\}$  (при условието, че  $N$  е устойчива).

**Пример 3.2** Нека е даден номиналният обект

$$P_0(s) = \frac{1}{s} \begin{bmatrix} 10 & 9 \\ 9 & 8 \end{bmatrix}$$

с неструктурирана мултипликативна входна неопределеност, така че истинският обект е

$$P(s) = P_0(s)(I + \Delta_i)$$

където

$$\bar{\sigma}(\Delta_i) \leq \left| \frac{1 + j\omega}{\alpha} \right| \quad (3.19)$$

за някаква константа  $\alpha$ . Изискването към качеството е

$$\bar{\sigma}(S(j\omega)) \leq \left| \frac{j\omega\alpha}{1 + j\omega} \right| \quad (3.20)$$

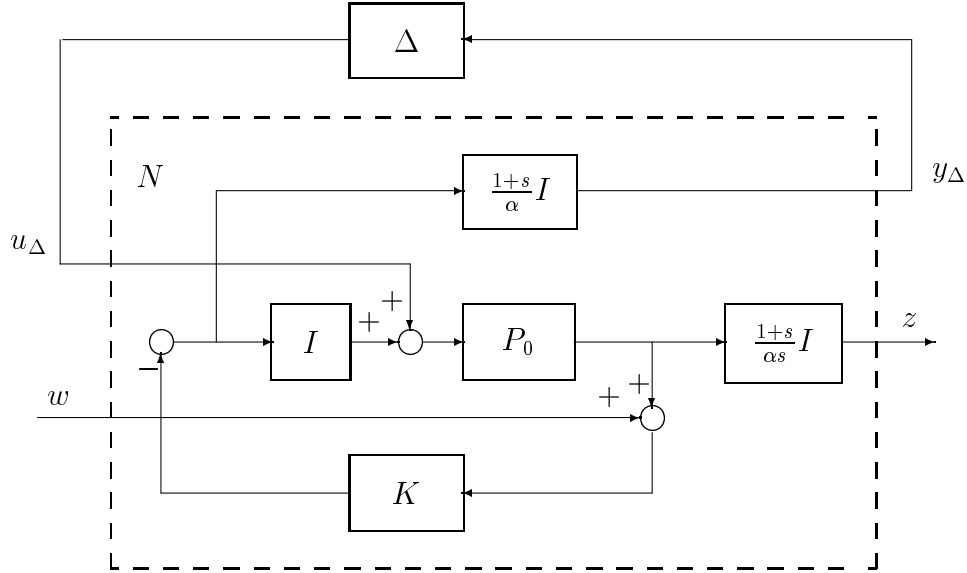
където  $S = (I + PK)^{-1}$ . Обърнете внимание, че увеличаването на  $\alpha$  води до улесняване постигането на робастност на качеството.

В случая желаем да определим стойностите на  $\alpha$ , за които изискването към качеството се постига в присъствието на неопределеност, при зададено управляващо устройство

$$K = \begin{bmatrix} 0.118 & 1 \\ 1 & -0.118 \end{bmatrix}. \quad (3.21)$$

За да направим това, най-напред реформулираме (3.19) и (3.20) като

$$\begin{aligned} \|\Delta\|_\infty &= \left\| \frac{\alpha}{1+j\omega} \Delta_i \right\|_\infty \leq 1 \\ \left\| \frac{1+j\omega}{j\omega\alpha} S \right\|_\infty &\leq 1 \end{aligned}$$



Фигура 3.9: Блок-схема, съответстваща на системата от Пример 3.2

и получаваме  $N$ , съответстващо на фиг. 3.3. Подробната блок-схема за този случай е показана на фиг. 3.9, от която получаваме

$$\begin{aligned}
 N &= \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} \\
 &= \begin{bmatrix} -\frac{1+s}{\alpha}(I + KP_0)^{-1}KP_0 & -\frac{1+s}{\alpha}(I + KP_0)^{-1}K \\ \frac{1+s}{\alpha s}(I + P_0K)^{-1}P_0 & \frac{1+s}{\alpha s}(I + P_0K)^{-1} \end{bmatrix}.
 \end{aligned}$$

Сега имаме, че

$$\mu(N) = \inf_D \bar{\sigma}(DND^{-1}), \quad (3.22)$$

където

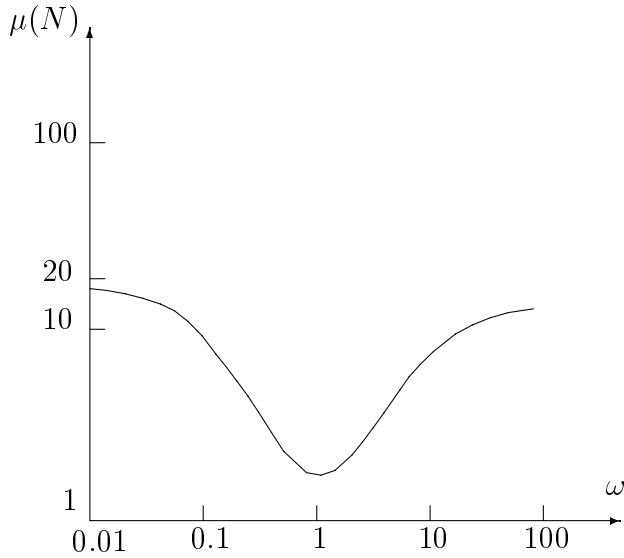
$$D = \begin{bmatrix} d_1 I_2 & 0 \\ 0 & d_2 I_2 \end{bmatrix},$$

като в (3.22) имаме равенство, тъй като  $D$  има само два блока, и (3.22) в този случай се опростява до

$$\mu(N) = \inf_d \bar{\sigma} \left( \begin{bmatrix} N_{11} & \frac{1}{d}N_{12} \\ dN_{21} & N_{22} \end{bmatrix} \right),$$

където  $d$  е положителен реален скалар. Зависимостта на  $\mu(N)$  от  $\omega$  при  $\alpha = 1$  е показана на фиг. 3.10, от която намираме че  $\|N\|_\mu = 18.5$ .

Тъй като  $\|N\|_\mu$  се изменя обратно пропорционално на  $\alpha$ , то управляващото устройство (3.21) постига робастност на качеството при условие, че  $\alpha > 18.5$ .



Фигура 3.10: Структурно сингулярно число на  $N$  при  $\alpha = 1$

### 3.4 $\mu$ -синтез и $DK$ -итерации

Структурното сингулярно число  $\mu$  е мощен инструмент за анализ на робастното качество при зададено управляващо устройство. Може обаче и да се търси управляващото устройство, което минимизира дадено  $\mu$ -условие: това е задачата за  $\mu$ -синтез.

#### $DK$ -итерации

Понастоящем не съществува пряк метод за синтез на  $\mu$ -оптимално управляващо устройство. Възможно е обаче да се използва итеративна процедура, която се нарича  $DK$ -итерация. Тя съчетава  $\mathcal{H}_\infty$ -синтез (разгледан по-нататък подробно в гл. 6) и  $\mu$ -анализ и обикновено дава добри резултати. Началната точка е горната граница (3.13) върху  $\mu$  в термините на мащабираното сингулярно число

$$\mu(N) \leq \inf_D \bar{\sigma}(DND^{-1}).$$

Идеята е да се намери управляващо устройство, което минимизира максималната стойност на тази горна граница като функция от честотата, т.е. да се реши задачата

$$\inf_K (\inf_D \|DN(K)D^{-1}\|_\infty) \quad (3.23)$$

минимизирайки алтернативно  $\|DN(K)D^{-1}\|_\infty$  по отношение на  $K$  или  $D$  (докато другото остава фиксирано). За да започнат итерациите се избира начална устойчива рационална предавателна матрица  $D(s)$  с подходяща структура. Обикновено добър начален избор на  $D$  е единичната матрица, при условие, че системата е подходящо мащабирана за качество.  $DK$ -итерацията тогава се осъществява по следния начин:

1.  **$K$ -стъпка.** Синтезира се  $\mathcal{H}_\infty$  управляващо устройство за мащабираната задача,  $\min_K \|DN(K)D^{-1}\|_\infty$  с фиксирано  $D(s)$ .
2.  **$D$ -стъпка.** Намира се  $D(j\omega)$ , което да минимизира за всяка честота  $\bar{\sigma}(DND^{-1}(j\omega))$  при фиксирано  $N$ .
3. Апроксимира се големината на всеки елемент на  $D(j\omega)$  с устойчива и минимално фазова предавателна функция  $D(s)$  и се отива на стъпка 1.

Итеративната процедура може да продължи докато се постигне удовлетворително качество,  $\|DN(K)D^{-1}\|_\infty < 1$ , или докато  $\mathcal{H}_\infty$ -нормата спре да намалява. Една основна трудност при този подход е, че въпреки изпъкналостта на всяка от отделните задачи ( $K$ -стъпка и  $D$ -стъпка), не е гарантирана съвместната изпъкналост. Ето защо итерациите могат да се сходят към локален оптимум. Практическите експерименти обаче показват, че в повечето случаи методът работи добре.

Редът на управляващото устройство, което се получава на всяка итерация е равен на броя на състоянията на обекта  $P(s)$  плюс броя на състоянията в тегловните матрици, плюс два пъти броя на състоянията в  $D(s)$ . В повечето случаи истинското  $\mu$ -оптимално управляващо устройство не е рационално и следователно е от безкрайно висок ред. Тъй като обаче ние използваме  $D(s)$  с краен ред за апроксимиране на  $D$ -мащабиращите множители, получаваме управляващо устройство от краен (но често от висок) ред. Истинското  $\mu$ -оптимално управляващо устройство би имало плоска  $\mu$ -крива (като функция на честотата), с изключение на безкрайно високите честоти, където  $\mu$  в общия случай трябва да стреми към фиксирана стойност, която не зависи от управляващото устройство (тъй като при тези честоти  $L(j\omega) = 0$  за реалните системи). С управляващо устройство от краен ред, обаче, в общия случай не е възможно (и може да не е желателно) да се разширява плоскостта на кривата в безкрайно високите честоти.

$DK$ -итерацията зависи силно от оптималните решения на Стъпки 1 и 2 и също от добрите апроксимации на Стъпка 3, които е за предпочитане да са с предавателни функции от нисък ред. Една от причините за предпочитане на апроксимация от нисък ред е, че това намалява реда на  $\mathcal{H}_\infty$  задачата, което обикновено подобрява числените свойства на  $\mathcal{H}_\infty$  оптимизацията (Стъпка 1) и също дава управляващо устройство от нисък ред. В някои случаи итеративната процедура се сходя бавно и може да е трудно да се прецени дали итерациите са сходящи. Възможно е даже да се случи, че стойността на  $\mu$  се увеличава. Това може да е резултат от числени трудности или неточности (напр. горната граница на  $\mu$  на Стъпка 2 да е по-висока от  $\mathcal{H}_\infty$ -нормата, получена на Стъпка 1) или поради лоша апроксимация на  $D$ -мащабиращите множители. В такива случаи би могло да се започне отначало, като се премащабират входовете и изходите.

На  $K$ -стъпката (Стъпка 1), където се синтезира  $\mathcal{H}_\infty$  управляващото устройство, често е желателно да се използва субоптимално управляващо устройство (напр. с  $\mathcal{H}_\infty$  норма  $\gamma$ , която е с 5% по-голяма от оптималната стойност  $\gamma_{\min}$ ).

### Апроксимиране на теглата върху качеството

Да напомним, че ако  $\mu$  за дадена честота е различно от 1, то това означава, че на тази честота можем да толерираме  $1/\mu$ -пъти повече неопределеност и все още да удовлетворяваме изискванията към качеството със запас от  $1/\mu$ . При  $\mu$ -синтеза проектантът обикновено трябва да настройва някои параметри в теглата върху качеството или неопределеността, докато максималната стойност на  $\mu$  се получи близка до 1. В някои случаи неопределеността е фиксирана и в действителност оптимизираме най-лошото качество чрез настройване на параметър в теглото върху качеството. Нека например разгледаме теглото върху качеството

$$w(s) = \frac{s/M + \omega_B}{s + \omega_B A},$$

където искаме  $M$  да остава постоянно и да намерим най-голямата достижима ширина на честотната лента  $\omega_B$ . Тогава оптимизационната задача става

$$\max |\omega_B| \quad \text{така че} \quad \mu(N) < 1, \text{ за всяка } \omega,$$

където  $N$  зависи от  $\omega_B$ . Това може да се осъществи като допълнителен външен цикъл на DK-итеративната процедура.

### Управляващи устройства с фиксиран ред

Понякога е желателно да се намери управляващо устройство от нисък ред с дадена структура. Това може да се постигне с числена оптимизация, където  $\mu$  се минимизира по отношение на параметрите на управляващото устройство. Трудността тук е, че оптимизационната задача не е изпъкнала по параметрите.

DK-итерацията може да се осъществи с функциите от  $\mu$ -Analysis and Synthesis Toolbox [7] на системата MATLAB.



## Глава 4

# СИНТЕЗ ПРИ ЗАДАДЕНИ ПОЛЮСИ

В тази глава се разглежда синтезът при зададени полюси и синтезът на наблюдатели на състоянието. В т. 4.1 се формулират най-общо различните задачи за синтез на линейни многомерни системи, описани в пространство на състоянията. Постановката на задачата за синтез при зададени полюси се дава в т. 4.2, а в т. 4.3 се разглеждат различни методи за синтез при зададени полюси. В т. 4.4 е изложено решението на по-общата задача за синтез на оптимална собствена структура. В т. 4.5 е описан синтеза на наблюдатели на състоянието от пълен и понижен ред, а в т. 4.6 се разглежда реализацията на линейни закони на управление с помощта на наблюдатели на състоянието.

### 4.1 Задачи на синтеза

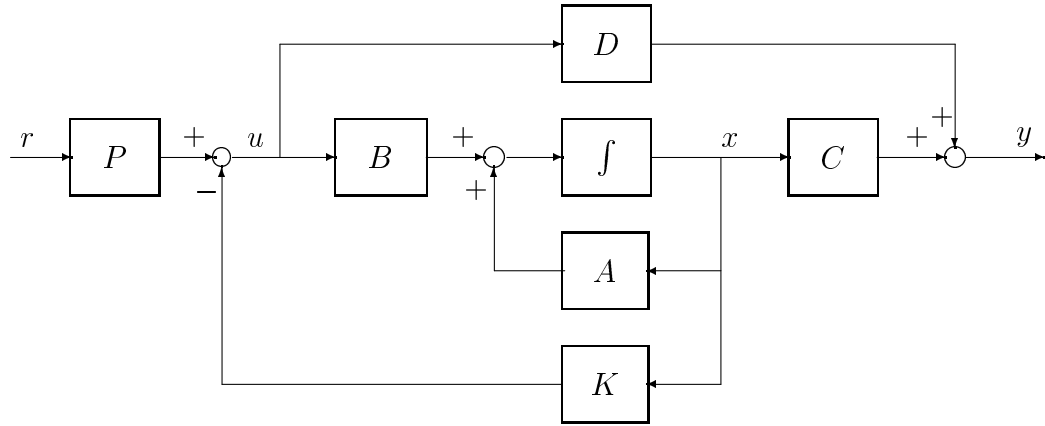
Да разгледаме линейната многомерна система

$$\begin{aligned}\dot{x} &= Ax + Bu, \\ y &= Cx + Du.\end{aligned}\tag{4.1}$$

Задачата за управление на системата (4.1) се състои във формиране на такъв вектор на управляващите въздействия  $u(t)$ , за който векторът на изхода на системата  $y(t)$  има желаното изменение във времето.

В затворените системи за управление, в които се използва принципът на обратната връзка, формирането на управляващите въздействия става въз основа на сравнение на действителното състояние на системата с желаното. Начинът, по който управлението  $u(t)$  се формира от състоянието  $x(t)$  и заданието  $r(t)$ , се нарича *закон на управление* и се записва във вида  $u(t) = u(r(t), x(t))$ . В линейните системи се използват закони на управление, при които  $u(t)$  е линейна функция на заданието  $r(t)$  и състоянието  $x(t)$ .

Синтезът на линейните системи за управление се състои в определяне на линейен закон на управление, за който затворената система трябва да удовлетворява поставените изисквания към качеството.



Фигура 4.1: Система с обратна връзка по състоянието

В случаите, когато векторът на състоянието  $x(t)$  е достъпен за измерване, линейният закон на управление се реализира като линейна обратна връзка по състоянието

$$u(t) = -Kx(t) + Pr(t), \quad (4.2)$$

където  $K$  се нарича *матрица на обратната връзка по състоянието*, а  $P$  – *матрица на правата връзка*. Елементите на  $K$  са коефициенти на обратните връзки по компонентите на вектора на състоянието.

Системата (4.1) със закон на управление (4.2) е показана на фиг. 4.1.

Уравненията на затворената система се дават от

$$\begin{aligned} \dot{x} &= (A - BK)x + BPr, \\ y &= (C - DK)x + DPr. \end{aligned} \quad (4.3)$$

Както ще бъде показано по-нататък, доминиращо влияние върху поведението на затворената система има матрицата на обратната връзка  $K$ . Матрицата на правата връзка  $P$  влияе върху установения режим на работа на системата, но не оказва влияние върху нейната динамика.

В някои случаи се разглежда задачата за регулиране ( $r(t) = 0$ ), при която представлява интерес само преходния процес, дължащ се на ненулеви начални условия на системата. В този случай вместо уравненията (4.3) се използват уравненията

$$\begin{aligned} \dot{x} &= (A - BK)x, \\ y &= (C - DK)x. \end{aligned} \quad (4.4)$$

От тези уравнения следва, че поведението на затворената система се определя от матрицата на състоянието на затворената система  $A - BK$ . За асимптотическа устойчивост на системата е необходимо матрицата  $K$  да се избере така, че собствените стойности на  $A - BK$  да лежат в лявата комплексна полуравнина.

Тази задача се нарича *задача за синтез на регулатор на състоянието*.



Въпреки че предположението за достъпност на вектора на състоянието е не-реалистично за голяма част от физическите системи, методите, разработени за синтез на регулатори на състоянието, имат голямо значение и в по-общият случай, когато е достъпен за измерване само част от вектора на състоянието.

Трябва да се отбележи, че обратната връзка по състоянието  $u(t) = -Kx(t)$  не променя управляемостта на системата. Ако системата (4.1) е напълно управляема, то такава е и системата (4.4) за всяка матрица на обратната връзка  $K$ . По-общият резултат, който може да се докаже е, че индексите на управляемост на системата остават непроменени при осъществяване на обратна връзка по състоянието. Наблюдаемостта на системата, обаче, може да се промени от обратна връзка по състоянието.

#### Пример 4.1 Системата

$$\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] = \left[ \begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 0 & 0 \\ \hline 1 & 0 & 0 \end{array} \right]$$

е управляема и наблюдаема. С обратна връзка по състоянието

$$u = -Kx = [-1 \quad -1]x,$$

системата става

$$\left[ \begin{array}{c|c} A - BK & B \\ \hline C - DK & D \end{array} \right] = \left[ \begin{array}{cc|c} 0 & 0 & 1 \\ 1 & 0 & 0 \\ \hline 1 & 0 & 0 \end{array} \right]$$

и не е напълно наблюдаема.

На практика обратната връзка по състоянието често не може да се съществи непосредствено поради следните причини:

1. Обикновено се разполага само с част от вектора на състоянието под формата на изходния вектор. Това се дължи на обстоятелството, че някои от вътрешните координати на реалните системи не са достъпни за измерване.
2. Дори и всички компоненти на вектора на състоянието да са достъпни, измерването на някои от тях е съпроводено с нежелано усложняване на системата или с въвеждане на допълнителни шумове. Така например, при системи с голяма размерност управлението по състоянието изисква прилагането на голям брой обратни връзки, което затруднява техническата реализация на системата.

При решаването на задачата за синтез при непълна информация за състоянието обикновено се използва подхода на реконструкция на вектора на състоянието, при който се търси подходяща апроксимация на вектора на състоянието  $\hat{x}(t) \approx x(t)$ , която да се използва в закона за управление вместо самия вектор на състоянието.

В този случай задачата за синтез се декомпозира в две фази. На първата фаза се синтезира закон на управление, като се предполага, че векторът на състоянието е достъпен за измерване. Поради това могат да се използват методите

за синтез на обратна връзка по състоянието. На втората фаза се синтезира динамична система, която реконструира вектора на състоянието. Тази система е обикновено от  $n$ -ти или  $n - r$ -ти ред и при детерминирани системи се нарича *наблюдател на състоянието* (*наблюдател на Луенбергер*), а при стохастичните системи – *филтър на Калман*.

### Определяне на матрицата на правата връзка

При условие, че матрицата на обратната връзка  $K$  е вече известна, определянето на матрицата на правата връзка  $P$  може да стане въз основа на алгебричните уравнения, описващи установения режим на работа на системата.

От уравненията на затворената система (4.3) при  $\dot{x} = 0$  и  $r = \text{const}$  за състоянието в установен режим  $x_\infty$  получаваме

$$(A - BK)x_\infty + BPr = 0,$$

откъдето

$$x_\infty = -(A - BK)^{-1}BPr.$$

(Тъй като затворената система е асимптотически устойчива, матрицата  $A - BK$  е неособена.)

Заместването на установеното състояние в уравнението на изхода дава

$$y_\infty = (C - DK)x_\infty + DPr = (-(C - DK)(A - BK)^{-1}B + D)Pr.$$

При *задачата за следене* е необходимо векторът на изхода в установен режим  $y_\infty$  да е равен на вектора на заданието  $r$ , откъдето получаваме

$$P = (-(C - DK)(A - BK)^{-1}B + D)^{-1},$$

при условие, че разглежданата матрица е неособена. По такъв начин матрицата на правата връзка може да се намери от условието за равенство на изхода и заданието в установен режим.

## 4.2 Постановка на синтеза при зададени полюси

Разглежда се системата

$$\dot{x} = Ax + Bu \tag{4.5}$$

и законът за управление

$$u(t) = -Kx(t) + Pr(t). \tag{4.6}$$

*Синтезът при зададени полюси* се състои в определяне на матрицата на обратната връзка  $K$ , за която полюсите на затворената система

$$\dot{x} = (A - BK)x + BPr \tag{4.7}$$

са разположени на желаните места в комплексната равнина. Изборът на тези места трябва да е съобразен с изискванията към динамиката на затворената

система. Например, от условието за асимптотическа устойчивост следва, че всички полюси на затворената система трябва да се намират в лявата комплексна полуравнина. Видът на преходните процеси обикновено се определя от полюсите, които се намират най-близо до имагинерната ос. Поради това тези полюси се наричат *доминиращи*.

**Пример 4.2** Дадена е системата

$$\dot{x}(t) = Ax(t) + Bu(t) \quad (4.8)$$

където  $x(t) \in \mathbb{R}^5$ ,  $u(t) \in \mathbb{R}^2$  и

$$A = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 \\ 0.1 & 0 & 0 & 0 & 0 \\ 0 & -0.2 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0.4 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 10 & 0 \\ 0 & 5 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

На фиг. 4.2 са показани преходните процеси за първата компонента на вектора на състоянието на затворената система за начално състояние

$$x_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

и реални кратни полюси

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \sigma,$$

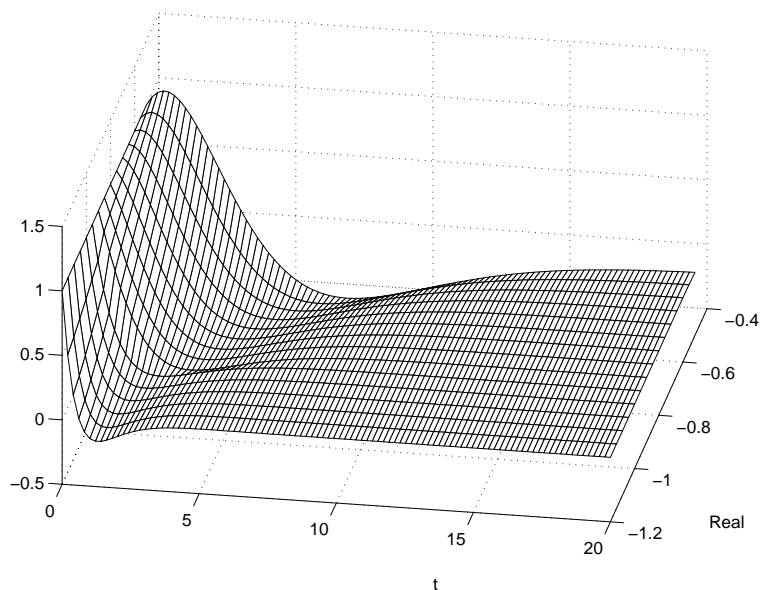
където  $\sigma$  се изменя от  $-0.5$  до  $-1.2$ . Вижда се, че намаляването на реалната част на полюсите води до ускоряване на преходните процеси. Това обаче води и до повишаване на коефициентите на обратните връзки (елементите на матрицата  $K$ .)

Влиянието на имагинерната част на доминиращите полюси върху преходните процеси на затворената система е показано на фиг. 4.3 за същото начално условие и полюси

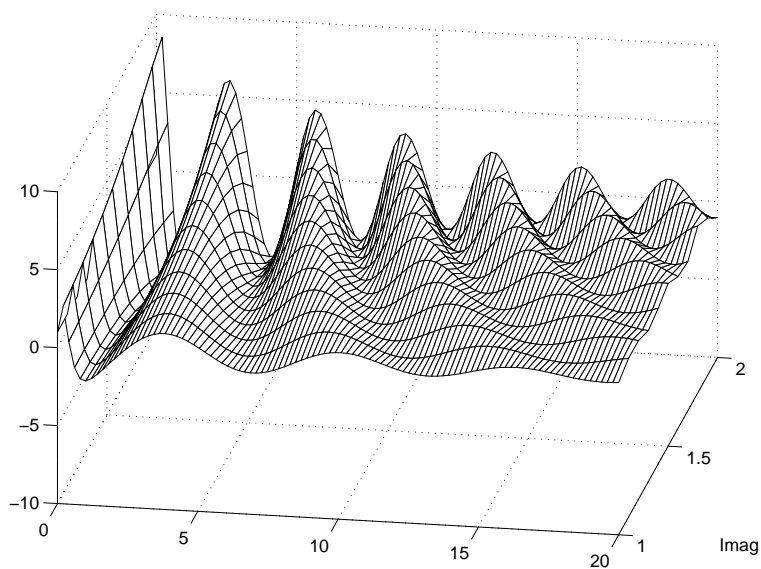
$$\lambda_1 = -0.1 + j\omega, \quad \lambda_2 = -0.1 - j\omega, \quad \lambda_3 = -3, \quad \lambda_4 = -4, \quad \lambda_5 = -5,$$

където  $\omega$  се изменя от 1 до 2. От фигурата се вижда, че с увеличаване на имагинерната част на двойката доминиращи комплексно спрегнати полюси амплитудата и честотата на колебанията нарастват.

Разполагането на полюсите на затворената система води до задаване на характера на свободните движения (т.е. на модите) на системата. (Поради тази причина понякога се използва термина *модално управление*.) Амплитудите на



Фигура 4.2: Преходни процеси при полюси с различна реална част



Фигура 4.3: Преходни процеси при полюси с различна имажинерна част

тези движения в различните компоненти на състоянието, а оттам и в изхода, точността на системата в установен режим и т.н. зависят от свойствата изобщо на матрицата на състоянието на затворената система  $A - BK$ , докато синтезът при зададени полюси има за цел да се фиксират само нулите на характеристичния полином на  $A - BK$ . Ето защо са разработени и методи, които дават възможност да се разполагат на зададени места не само полюсите, а и нулите на

предавателната матрица, да се извършва синтез на оптимална собствена структура (собствени стойности и собствени вектори) на  $A - BK$  и т.н. Някои от тези възможности са разгледани по-нататък в т. 4.4.

Синтезът при зададени полюси се основава на следния фундаментален резултат.

**Теорема 4.1** *Нека  $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$  е произволно множество от  $n$  комплексни числа  $\lambda_i$ , в което всяко  $\lambda_i$  с  $\text{Im}(\lambda_i) \neq 0$  участва в спрегната двойка. Тогава матричната двойка  $(A, B)$  е напълно управляема, ако и само ако за всеки избор на  $\Lambda$  съществува реална матрица  $K$ , такава че  $A - BK$  има  $\Lambda$  за множество от своите собствени стойности.*

Този резултат показва, че полюсите на затворената система (4.7) могат да се разположат с подходящ избор на матрицата  $K$  на произволни места в комплексната равнина (при ограничението, че комплексните полюси трябва да образуват спрегнати двойки), ако и само ако дадената система е напълно управляема. От това следва важният извод, че напълно управляемата система винаги може да се направи устойчива чрез обратна връзка по състоянието.

Нека сега разгледаме случая, когато системата

$$\dot{x}(t) = Ax(t) + Bu(t)$$

не е напълно управляема. Прилагайки неособено преобразуване  $x(t) = S\tilde{x}(t)$ , получаваме

$$\dot{\tilde{x}}(t) = \tilde{A}\tilde{x}(t) + \tilde{B}u(t),$$

където

$$\tilde{A} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}$$

и двойката  $(A_{11}, B_1)$  е управляема. Нека обратната връзка по състоянието е взета като

$$u(t) = -\tilde{K}\tilde{x}(t) = -[K_1, K_2]x(t),$$

където разделянето на матрицата  $\tilde{K}$  съответствува на това на  $\tilde{A}$ . Преобразуваната затворена система се описва от

$$\dot{\tilde{x}}(t) = \tilde{F}\tilde{x}(t),$$

където

$$\tilde{F} = \begin{bmatrix} A_{11} - B_1K_1 & A_{12} - B_1K_2 \\ 0 & A_{22} \end{bmatrix}$$

е горна блок-триъгълна матрица.

Тъй като  $\det(\lambda I - \tilde{F}) = \det(\lambda I - A_{11} + B_1K_1) \det(\lambda I - A_{22})$  то собствените стойности на матрицата  $\tilde{F}$  са тези на  $A_{11} - B_1K_1$  и  $A_{22}$ . Следователно, обратната връзка по състоянието влияе само върху управляемата част на системата. Управляемите полюси могат да се поместят на желаните места, но неуправляемите

полюси не се променят от обратната връзка по състоянието. Ако системата е стабилизируема, матрицата  $A_{22}$  е устойчива и следователно може да се намери обратна връзка по състоянието, за която затворената система е асимптотически устойчива. Матрицата  $K_2$  не влияе върху полюсите на затворената система и може да се избере произволно.

Задачата за синтез при зададени полюси на линейната дискретна система

$$x_{k+1} = Ax_k + Bu_k \quad (4.9)$$

е напълно аналогична на задачата за непрекъснати системи. Теорема 4.1 е в сила и за дискретния случай. Единствената разлика е, че желаните полюси трябва да лежат вътре в единичната окръжност, което осигурява асимптотическата устойчивост на затворената система.

Интересна особеност на дискретния случай е, че преходният процес на затворената система може да се направи да затихне за  $n$  стъпки, т.е. възможно е да се намери обратна връзка по състоянието, такава че  $x_n = 0$  за всяко  $x_0$ . Дискретни системи с това свойство се наричат *системи с безкрайна степен на устойчивост*. Затворената система е с безкрайна степен на устойчивост ако и само ако нейните полюси са равни на нула. Съгласно теоремата на Кейли–Хамилтън в този случай имаме, че

$$(A - BK)^n = 0$$

т.е.  $A - BK$  е нилпотентна матрица и

$$x_n = (A - BK)^n x_0 = 0.$$

В следващата точка е показано, че състоянието на многовходова дискретна система може да се направи да затихне за  $p < n$  стъпки, където  $p$  е индексът на управляемост на системата.

### 4.3 Методи за синтез при зададени полюси

В тази точка се разглеждат някои от известните методи за синтез при зададени полюси.

Да разгледаме най-напред едноходовата система

$$\dot{x}(t) = Ax(t) + bu(t), \quad (4.10)$$

където  $b$  е  $n$ -мерен вектор-колона и  $u(t)$  е скалярна функция. Предполага се, че системата е напълно управляема.

Линейната обратна връзка в случая има вида

$$u(t) = -kx(t), \quad (4.11)$$

където  $k$  е  $n$ -мерен вектор-ред. Желаното положение на полюсите на затворената система може да се зададе чрез коефициентите на характеристичния полином на затворената система

$$d(\lambda) = \lambda^n + d_n \lambda^{n-1} + \dots + d_1. \quad (4.12)$$

При отбелязаните по-горе ограничения върху полюсите тези коефициенти са реални числа.

Използвайки неособеното преобразуване

$$x(t) = S\tilde{x}(t)$$

системата (4.10) се преобразува в каноничната управляема форма

$$\dot{\tilde{x}}(t) = A_c\tilde{x}(t) + b_c u(t) \quad (4.13)$$

$$A_c = S^{-1}AS = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a_1 & -a_2 & -a_3 & \dots & -a_n \end{bmatrix}, \quad b_c = S^{-1}b = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Числата  $a_1, a_2, \dots, a_n$  са коефициентите на характеристичния полином на отворената система

$$\det(\lambda I - A) = \lambda^n + a_n \lambda^{n-1} + \dots + a_1.$$

Обратната връзка по състоянието (4.11) добива вида

$$u(t) = -k_c \tilde{x}(t), \quad (4.14)$$

където

$$k_c = kS = [k_1, k_2, \dots, k_n].$$

Системата (4.13) заедно с обратната връзка по състоянието (4.14) се описва от

$$\dot{\tilde{x}}(t) = (A_c - b_c k_c) \tilde{x}(t),$$

където

$$A_c - b_c k_c = \begin{bmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \\ -a_1 - k_1 & -a_2 - k_2 & \dots & -a_n - k_n \end{bmatrix}.$$

Тъй като матрицата  $A_c - b_c k_c$  е в съпровождаща форма, нейният характеристичен полином е

$$\det(\lambda I - A_c + b_c k_c) = \lambda^n + (a_n + k_n) \lambda^{n-1} + \dots + a_1 + k_1. \quad (4.15)$$

Изравнявайки изразите (4.15) и (4.12), получаваме

$$k_c = d - a,$$

където

$$d = [d_1, d_2, \dots, d_n], \quad a = [a_1, a_2, \dots, a_n].$$

По този начин матрицата на обратната връзка, която помещава полюсите на желаните места, се дава от

$$k = (d - a)S^{-1}.$$

Полученият резултат показва, че при едноходовите системи решението за матрицата на обратната връзка е единствено.

В многоходовия случай

$$\dot{x} = Ax + Bu \quad (4.16)$$

където  $B$  е  $n \times m$  матрица ( $1 < m < n$ ), обратната връзка

$$u(t) = -Kx(t)$$

която разполага полюсите на затворената система на желаните места, не е единствена. Това се дължи на факта, че матрицата на обратната връзка съдържа  $nm$  параметри, които удовлетворяват  $n$  условия, съответстващи на  $n$ -те коефициенти на желания характеристичен полином.

Най-простият начин за намиране на обратната връзка в многоходовия случай е да се приведе системата (4.16) в еквивалентна едноходова система, използвайки *диадична матрица на обратната връзка*

$$K = qk, \quad (4.17)$$

където  $q$  е  $m$ -мерен вектор-колона и  $k$  е  $n$ -мерен вектор-ред. Тъй като в този случай рангът на матрицата  $K$  е равен на 1, съответстващата обратна връзка е известна още като *обратна връзка от единичен ранг*. Посредством тази обратна връзка синтезът при зададени полюси на многомерната система се свежда до синтеза при зададени полюси на еквивалентната едноходова система

$$\dot{x} = Ax + bw(t), \quad (4.18)$$

където  $b = Bq$ .

За дадено  $q$  матрицата  $k$  може да се определи от желаното разположение на полюсите при условие, че системата (4.18) е напълно управляема. Това условие е еквивалентно на изискването матрицата  $A$  да е *циклична*, т.е. да съществува вектор  $s$ , за който

$$\text{rank}[s, As, \dots, A^{n-1}s] = n.$$

Ако матрицата  $A$  не е циклична, трябва предварително да се използва произволна обратна връзка  $u(t) = -K_0x(t)$ , за която матрицата  $A - BK_0$  има различни собствени стойности и следователно е циклична матрица.

Използването на диадична обратна връзка води до загуба на свобода при синтеза, тъй като възможностите на многоходовата система изкуствено се свеждат до тези на едноходовата. Появяват се и трудности поради необходимостта матрицата на състоянието на еквивалентната едноходова система да е циклична. При този подход матрицата на обратната връзка се получава от единичен ранг, което в някои случаи е нежелателно.



Матрица на обратната връзка от пълен ранг може да се получи с помощта на каноничната форма на Луенбергер (т.1.4).

Нека чрез неособеното преобразуване  $\tilde{x}(t) = Tx(t)$  системата (4.16) е приведена в каноничната форма

$$\dot{\tilde{x}}(t) = A_c \tilde{x}(t) + B_c u(t), \quad (4.19)$$

където матриците  $A_c = TAT^{-1}$  и  $B_c = TB$  имат вида (1.4), съответно (1.5).

Законът за управление на системата (4.19) се взема като

$$u(t) = -K_c \tilde{x}(t),$$

където  $K_c$  е  $m \times n$  матрица.

Тъй като матрицата на състоянието на затворената преобразувана система

$$F_c = \left[ \begin{array}{cccc|cccc|c|c|cccc} 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \dots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & & 0 & 0 & \dots & 0 \\ \times & \times & \dots & \times & \times & \times & \dots & \times & & \times & \times & \dots & \times \\ - & - & - & - & - & - & - & - & - & - & - & - & - \\ 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \dots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & & 0 & 0 & \dots & 0 \\ \times & \times & \dots & \times & \times & \times & \dots & \times & & \times & \times & \dots & \times \\ - & - & - & - & - & - & - & - & - & - & - & - & - \\ & & & \vdots & & & \dots & & \ddots & & & & \vdots \\ - & - & - & - & - & - & - & - & - & - & - & - & - \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \dots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & & 0 & 0 & \dots & 1 \\ \times & \times & \dots & \times & \times & \times & \dots & \times & & \times & \times & \dots & \times \end{array} \right],$$

има нетривиални редове само в позициите  $\sigma_1, \sigma_2, \dots, \sigma_m$ ,  $\sigma_i = \sum_{j=1}^i p_j$ , желаното множество от собствени стойности на  $F_c$  може да се постигне, ако редовете на матрицата  $K_c$  се изберат така, че поддиагоналните или наддиагоналните блокове на  $F_c$  се нулират, а най-долните редове на диагоналните блокове се променят по подходящ начин. Тогава матрицата  $F_c$  добива вида (ако се нулират поддиагоналните блокове)

$$F_c = \begin{bmatrix} F_{11} & F_{12} & \dots & F_{1m} \\ & F_{22} & \dots & F_{2m} \\ & & \ddots & \vdots \\ 0 & & & F_{mm} \end{bmatrix},$$

където всяка от матриците  $F_{ii}$ ,  $i = 1, 2, \dots, m$  е в съпровождаща форма. Характеристичният полином на  $F_c$ ,

$$\det(\lambda I - F_c) = \prod_{i=1}^m \det(\lambda I - F_{ii}),$$

е произведение от характеристичните полиноми на диагоналните блокове. Коефициентите на тези полиноми са елементите в последните редове на диагоналните блокове, взети с обратни знаци. Тези елементи се получават от желаните собствени стойности на диагоналните блокове.

Нека от редовете на матриците  $A_c, F_c, B_c$  в позициите  $\sigma_1, \sigma_2, \dots, \sigma_m$  са конструирани  $m \times m$  матрицата  $\hat{A}$ ,  $m \times n$  матрицата  $\hat{F}$  и  $m \times m$  матрицата  $\hat{B}$ , съответно. Отбележете, че матрицата  $\hat{B}$  е в горна триъгълна форма с единици върху диагонала,

$$\hat{B} = \begin{bmatrix} 1 & \times & \times & \dots & \times \\ & 1 & \times & \dots & \times \\ & & \ddots & & \times \\ & & & 1 & \times \\ 0 & & & & 1 \end{bmatrix}$$

и следователно е неособена матрица. Тогава уравненението

$$B_c K_c = A_c - F_c$$

се свежда до

$$\hat{B} \hat{K} = \hat{A} - \hat{F}$$

така че

$$\hat{K} = \hat{B}^{-1}(\hat{A} - \hat{F}).$$

Матрицата  $\hat{K}$ , получена по този начин, е неединствена, тъй като желаните собствени стойности на матрицата на състоянието на затворената система могат да се разпределят по различен начин между блоковете на  $F_c$ .

Матрицата на обратната връзка на първоначалната система се определя от

$$K = K_c T.$$

В общия случай матрицата  $K$  е от пълен ранг.

**Пример 4.3** Нека

$$A = \begin{bmatrix} -3 & 11 & -5 & 4 & -26 \\ -5 & 14 & 2 & -4 & -33 \\ 4 & -7 & 6 & -1 & 5 \\ 3 & -7 & 9 & -6 & 9 \\ -3 & 7 & 1 & -3 & -14 \end{bmatrix}, \quad B = \begin{bmatrix} -3 & 7 \\ -8 & 19 \\ -4 & 10 \\ -4 & 10 \\ -3 & 7 \end{bmatrix}$$

и желаните полюси на затворената система са

$$-1 + j, -1 - j, -2 + j, -2 - j, -3.$$

Използвайки неособено преобразуване с матрица

$$T = \begin{bmatrix} 77 & -102 & 1 & 50 & 127 \\ 52 & -49 & -6 & 34 & 41 \\ 3 & -4 & 0 & 2 & 5 \\ 2 & -2 & 0 & 1 & 2 \\ 1 & 1 & -3 & 4 & -5 \end{bmatrix},$$

получаваме

$$A_c = TAT^{-1} = \left[ \begin{array}{cc|cc} 0 & 1 & 0 & 0 & 0 \\ -5 & 7 & 125 & -155 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -2 & 3 & 49 & -66 & -10 \end{array} \right], \quad B = \left[ \begin{array}{cc} 0 & 1 \\ 1 & 0 \\ \hline 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{array} \right]$$

така че  $p_1 = 2$  и  $p_2 = 3$ .

Представяйки характеристичният полином на затворената система в съответствие с индексите на Кронекер,

$$(\lambda^2 + 2\lambda + 2)(\lambda^3 + 7\lambda^2 + 17\lambda + 15),$$

получаваме

$$F_c = \left[ \begin{array}{cc|cc} 0 & 1 & 0 & 0 & 0 \\ -2 & -2 & 125 & -155 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -15 & -17 & -7 \end{array} \right].$$

Вземайки вторите и петите редове на  $A_c$ ,  $F_c$  и  $B_c$  намираме

$$\hat{A} = \left[ \begin{array}{ccccc} -5 & 7 & 125 & -155 & 0 \\ -2 & 3 & 49 & -66 & -10 \end{array} \right],$$

$$\hat{F} = \left[ \begin{array}{ccccc} -2 & -2 & 125 & -155 & 0 \\ 0 & 0 & -15 & -17 & -7 \end{array} \right],$$

$$\hat{B} = \left[ \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right].$$

Следователно,

$$\hat{K} = K_c T^{-1} = \left[ \begin{array}{ccccc} -3 & 9 & 0 & 0 & 0 \\ -2 & 3 & 64 & -49 & -3 \end{array} \right]$$

и

$$K = \hat{K}T = \left[ \begin{array}{ccccc} 237 & -135 & -57 & 156 & -12 \\ 93 & -104 & -11 & 69 & 106 \end{array} \right].$$

Друго решение се получава, ако вземем

$$F_c = \left[ \begin{array}{cc|cc} 0 & 1 & 0 & 0 & 0 \\ -2 & -2 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -2 & 3 & -15 & -17 & -7 \end{array} \right],$$

така че

$$\hat{F} = \begin{bmatrix} -2 & -2 & 0 & 0 & 0 \\ -2 & 3 & -15 & -17 & -7 \end{bmatrix},$$

$$\hat{K} = \begin{bmatrix} -3 & 9 & 125 & -155 & 0 \\ 0 & 0 & 64 & -49 & -3 \end{bmatrix}$$

и

$$K = \begin{bmatrix} 302 & -325 & -57 & 251 & 303 \\ 91 & -161 & 9 & 67 & 237 \end{bmatrix}.$$

Аналогично могат да се зададат и други структури на  $F_c$ , всяка от които ще води до различна матрица на обратната връзка. Очевидно е, че достижимите структури на матрицата  $F_c$  зависят от индексите на Кронекер за дадената система. В тази връзка възниква следната по-обща задача: Възможно ли е с обратна връзка по състоянието да се направи матрицата  $A - BK$  подобна на матрица  $F$ , която се характеризира с определена структура, например зададени степени на нейните инвариантни полиноми? (Ще напомним, че когато матрицата  $F$  е в рационална канонична форма,

$$F = \begin{bmatrix} F_1 & & & \\ & F_2 & & \\ & & \ddots & \\ & & & F_h \end{bmatrix},$$

където  $F_1, F_2, \dots, F_h$  са съпровождащи матрици, степените на инвариантните полиноми на  $F$  са равни на размерностите на тези матрици.)

Отговорът на този въпрос се дава от следната теорема.

**Теорема 4.2** (Фундаментална теорема за линейната обратна връзка.) *Нека  $f_1, f_2, \dots, f_h$  са степените на инвариантните полиноми на матрицата  $F$ . Тогава съществува матрица  $K$ , такава че  $A - BK$  е подобна на  $F$  ако и само ако*

$$\sum_{i=1}^s (f_i - p_i) \geq 0 \quad \text{за всички } s = 1, 2, \dots, \min \{h, m\} \quad (4.20)$$

където  $p_1, p_2, \dots, p_m$  са индексите на управляемост на двойката  $(A, B)$ .

**Пример 4.4** Дадена е система с  $n = 3$ ,  $m = 2$ . Съгласно (4.20) достижимите рационални канонични форми на матрицата на затворената система са

$$f_1 = 3, f_2 = 0 \qquad f_1 = 2, f_3 = 1$$

$$F = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \times & \times & \times \end{bmatrix}, \quad F = \begin{bmatrix} 0 & 1 & | & 0 \\ \times & \times & | & 0 \\ - & - & - & - \\ 0 & 0 & | & \times \end{bmatrix}.$$

Синтезът при зададени полюси на дискретната система

$$x_{k+1} = Ax_k + Bu_k$$

може да се осъществи по същия начин, както в непрекъснатия случай, имайки предвид, че желаните полюси трябва да се намират вътре в единичната окръжност. Използвайки каноничната форма на Луенбергер, в многовходовия случай преобразуваната матрица на затворената система може да се избере така, че елементите в нетривиалните редове да са равни на нула,

$$F_c = \left[ \begin{array}{cccc|cccc|ccc|cccc} 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & & & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \dots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & & & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & & & 0 & 0 & \dots & 0 \\ \hline 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & & & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \dots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & & & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & & & 0 & 0 & \dots & 0 \\ \hline & & & \vdots & & & & \dots & & & & & & \vdots \\ \hline 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & & & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \dots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & & & 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & & & 0 & 0 & \dots & 0 \end{array} \right].$$

Това задаване на  $F_c$  отговаря на нулеви желани полюси, т.е. синтезираната система ще притежава безкрайна степен на устойчивост. Интересното в този случай е, че матрицата на състоянието на затворената система ще удовлетворява съотношението

$$(A - BK)^p = 0,$$

което означава, че състоянието на системата ще затихва за  $p$  стъпки, т.е.  $x_p = 0$ . В дадения случай характеристичният полином на матрицата  $A - BK$  е  $s^n$ , а нейният минимален полином е  $s^p$  и на  $n$ -кратната нулева собствена стойност съответствуват  $m$  линейно независими собствени вектори. По такъв начин състоянието на многовходова дискретна система може да се направи да затихне за по-малко от  $n$  стъпки.

## 4.4 Синтез на оптимална собствена структура

Както бе отбелязано в т. 4.2, свободата, която съществува при синтеза при зададени полюси, може да се използва за удовлетворяване на различни изисквания към качеството на затворената система. В тази точка е разгледан метод за синтез на оптимална собствена структура на многовходови системи, при който собствените вектори се определят така, че да се екстремизира някакъв показател на качеството на системата. При този метод свободата в синтеза се параметризира

директно в определени елементи на собствените вектори или обобщените собствени вектори и тези елементи се определят итеративно в процеса на решаването на различни оптимизационни задачи, възникващи при синтеза на многовходови непрекъснати и дискретни системи.

Да разгледаме многовходовата система

$$\begin{aligned}\dot{x}(t) &= Ax(t) + B_1 w(t) + B_2 u(t) \\ z(t) &= C_1 x(t) + D_{12} u(t)\end{aligned}\quad (4.21)$$

където  $x(t)$ ,  $u(t)$ ,  $w(t)$  и  $z(t)$  са векторите на състоянието, управлението, смущението и управляемия изход, съответно, и  $A$ ,  $B_1$ ,  $B_2$ ,  $C_1$ ,  $D_{12}$  са матрици със съответните размерности (предполагаме, че  $\text{rank}(B_2) = m > 1$ ).

Нека  $(m_1, \dots, m_p)$  е множеството на спрегнатите Кронекерови индекси на двойката  $(A, B_2)$ :

$$m_1 = m = \text{rank}(B_2)$$

$$m_i = \text{rank}(P_i) - \text{rank}(P_{i-1}); \quad i = 2, \dots, p$$

където  $p$  е индексът на управляемост на  $(A, B_2)$  и

$$P_i = [B_2 \ AB_2 \ \dots \ A^{i-1} B_2].$$

Предполага се, че векторът на състоянието  $x(t)$  е достъпен за измерване.

Прилагайки обратната връзка по състоянието  $u(t) = -Kx(t)$ , получаваме затворената система

$$\begin{aligned}\dot{x}(t) &= A_{cl}x(t) + B_{cl}w(t) \\ z(t) &= C_{cl}x(t)\end{aligned}\quad (4.22)$$

където  $A_{cl} = A - B_2 K$ ,  $B_{cl} = B_1$ ,  $C_{cl} = C_1 - D_{12} K$ .

Нека

$$\lambda = \{\lambda_1, \dots, \lambda_n\}$$

е множеството (симетрично относно реалната ос) на зададените полюси на (4.22), т.е. собствените стойности на матрицата на затворената система  $A - B_2 K$ . Тогава съществува  $m \times n$  матрица  $K$ , такава че

$$\text{spect}(A - B_2 K) = \lambda, \quad (4.23)$$

където със  $\text{spect}(A)$  е означен спектърът (множеството от собствените стойности) на  $A$ .

Множеството  $\mathcal{K}_\lambda$  на всички матрици  $K$  със свойството (4.23) се нарича *общо решение на задачата за синтез при зададени полюси* за системата (4.21).

За дадено  $\lambda$ , да означим с  $\mathcal{V}_\lambda$  множеството на всички матрици  $V$ , чийто колони са собствените или обобщените собствени вектори на  $A - B_2 K$  за  $K \in \mathcal{K}_\lambda$ .

Задачата за намиране на  $K$  за дадено  $\lambda$  и  $\mathcal{V}_\lambda$  се нарича *задача за синтез при зададена собствена структура* за системата (4.21) и е частен случай на общата задача за синтез при зададени полюси. В т. 7.10 е описан числен метод за решаване на тази задача, при който се използва параметризация на множеството

$\mathcal{K}_\lambda$  чрез елементи на собствените вектори и обобщените собствени вектори на матрицата  $A - B_2K$ .

Свободата в множеството  $\mathcal{K}_\lambda$  може да се използва за оптимизиране на някакъв показател на качеството, специфициращ поведението на затворената система. Това може да се извърши посредством някаква оптимизационна техника, изменяща по итеративен път свободните елементи на собствените вектори и матрицата на обратната връзка, така че да се намери екстремума на съответната целева функция.

Да разгледаме приложението на синтеза при зададена собствена структура към следните оптимизационни задачи, възникващи при синтеза на многомерни системи.

- **Синтез при зададени полюси с  $H_2$  оптимизация**

В този случай оптимизационната задача се състои в намиране на матрицата  $K$ , която разполага полюсите на затворената система на желаните места и в същото време минимизира 2-нормата на предавателната матрица на затворената система

$$H(s) = C_{cl}(sI - A_{cl})^{-1}B_{cl}$$

от  $w$  към  $z$ .

Имаме, че

$$\|H\|_2^2 = \text{trace}[C_{cl}P_cC_{cl}^T] = \text{trace}[B_{cl}^TP_oB_{cl}]$$

където матриците  $P_o$ ,  $P_c$  са решенията на матричните уравнения на Ляпунов

$$\begin{aligned} A_{cl}P_c + P_cA_{cl}^T + B_{cl}B_{cl}^T &= 0 \\ A_{cl}^TP_o + P_oA_{cl} + C_{cl}^TC_{cl} &= 0. \end{aligned}$$

- **Синтез при зададени полюси с  $H_\infty$  оптимизация**

Целта на тази техника на синтез е да се осигури желаната динамика на затворената система, задавайки полюсите на затворената система и минимизирайки в същото време  $H_\infty$  нормата на предавателната матрица на затворената система от  $w$  към  $z$ , за да се подтисне смущаващото въздействие.

- **Робастен синтез при зададени полюси**

Свободата в собствената структура на матрицата на затворената система  $A_{cl}$  може да се използва за минимизиране на числото на обусловеност  $\text{cond}_2(V)$  или  $\text{cond}_F(V)$  на матрицата от собствените вектори  $V$  на затворената система  $A_{cl}$ , което води до минимизиране на цялостната чувствителност на полюсите на затворената система. Тази целева функция работи

само ако желаният спектър позволява пълен набор от линейно независими собствени вектори на  $A_{cl}$ . (Програмната реализация на съответния метод се осъществява с функцията `place` от MATLAB.)

Трябва да се подчертае, че е възможно също така синтезът при зададена собствена структура да се приложи и с други целеви функции, в това число комбинации на някои от целевите функции представени по-горе (например, синтез при зададени полюси със смесена  $H_2/H_\infty$  оптимизация).

Определянето на матрицата на обратната връзка при решаване на разгледащите задачи може да се осъществи с алгоритъма DEIGEN, изложен в т. 7.10, който дава възможност чрез задаване на свободните елементи на собствените вектори да се получи определено решение за матрицата  $K$ . Тези свободни елементи могат да се подберат с помощта на оптимизационна процедура по такъв начин, че да се екстремизира съответния показател на качеството на затворената система.

**Пример 4.5** Да разгледаме непрекъсната система от пети ред с два входа и два изхода с матрици

$$A = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 \\ 0.1 & 0 & 0 & 0 & 0 \\ 0 & -0.2 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0.4 & -1 \end{bmatrix}, B_1 = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ -1 & 1 \\ 0 & 1 \\ 1 & -1 \end{bmatrix},$$

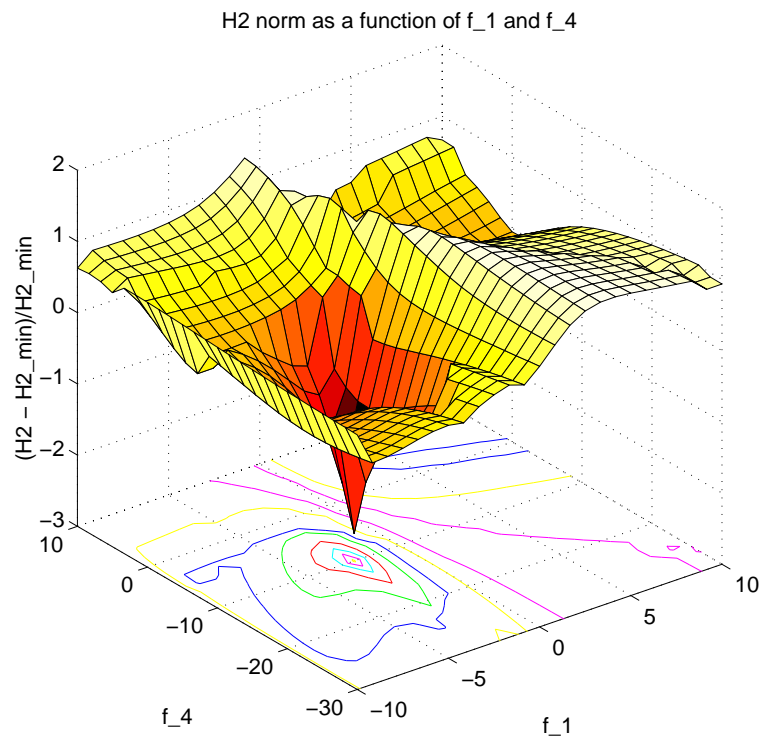
$$B_2 = \begin{bmatrix} 10 & 0 \\ 0 & 5 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, C_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, D_{12} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Полюсите на затворената система се избират като

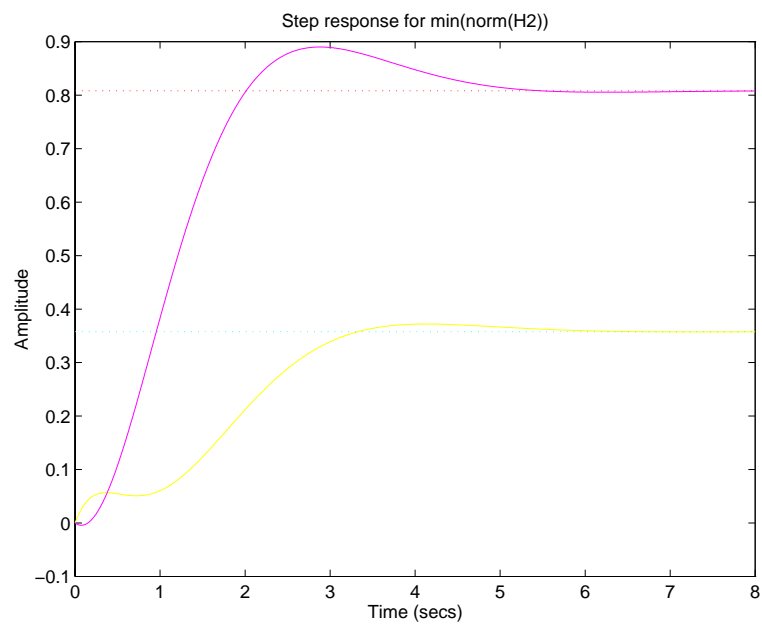
$$\lambda = \{-1.0 + j1.0, -1.0 - j1.0, -1.5, -2.0 + j0.5, -2.0 - j0.5\}.$$

За всяка от трите задачи, дадени по-горе и едно и също разположение на полюсите, свободните елементи на собствените вектори и матрицата на обратната връзка (в този случай те са 8) се определят итеративно, така че да се екстремизира съответната целева функция. Оптимизацията се извършва чрез глобалния оптимизационен алгоритъм за непрекъснато симулирано откаляване [14].  $H_\infty$  нормата на затворената система се определя чрез функцията `normhinf` от Robust Control Toolbox [10]. Във всички случаи при оптимизацията се използват допуск от  $10^{-3}$  и едни и същи начални стойности на свободните параметри  $f_1, f_2, \dots, f_8$  (генерирани по случаен начин чрез функцията `rand` от MATLAB). Трябва да се отбележи, че стойността на допускателна не се отразява на точността на определяне на  $K$  от желаните полюси и зададените елементи на собствените вектори.

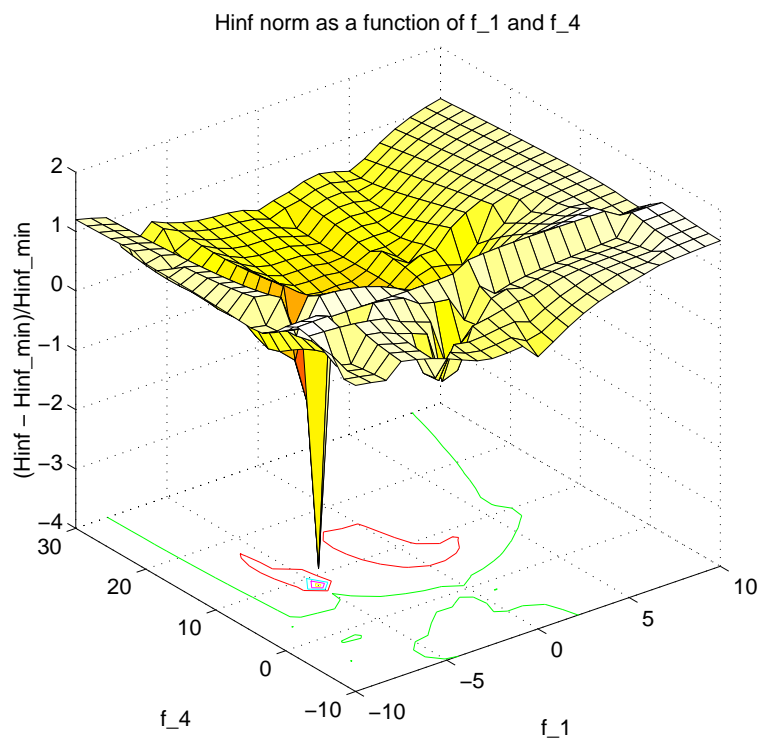




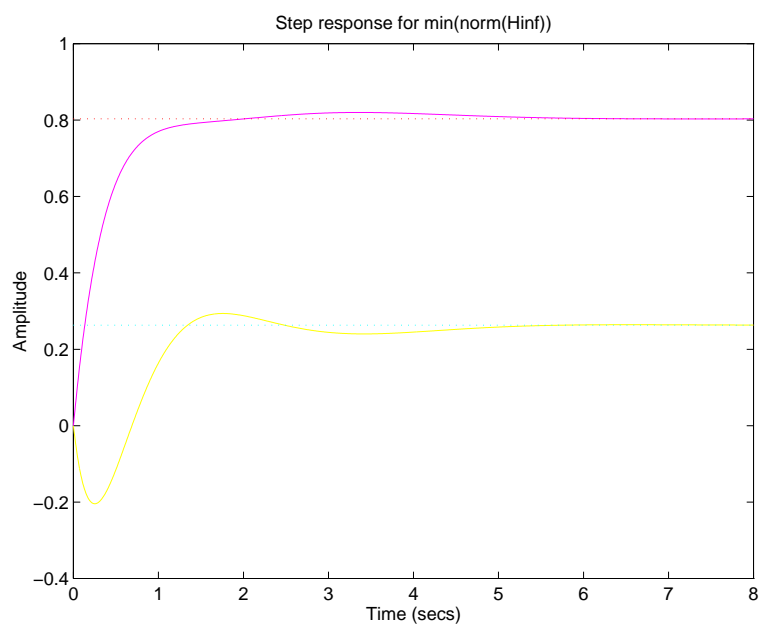
Фигура 4.4:  $\|H\|_2$  като функция на параметрите  $f_1$  и  $f_4$  около минимума



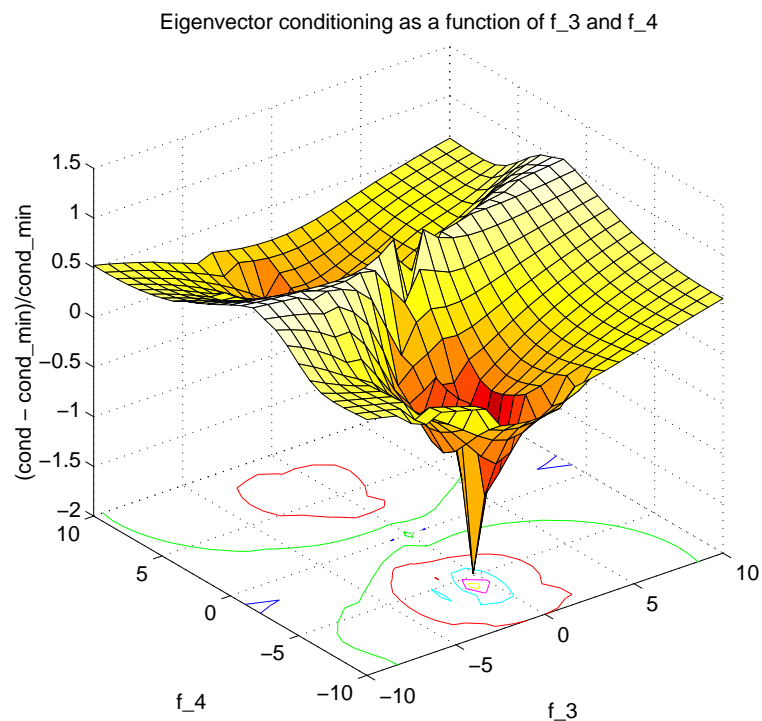
Фигура 4.5: Преходни процеси при минимално  $\|H\|_2$



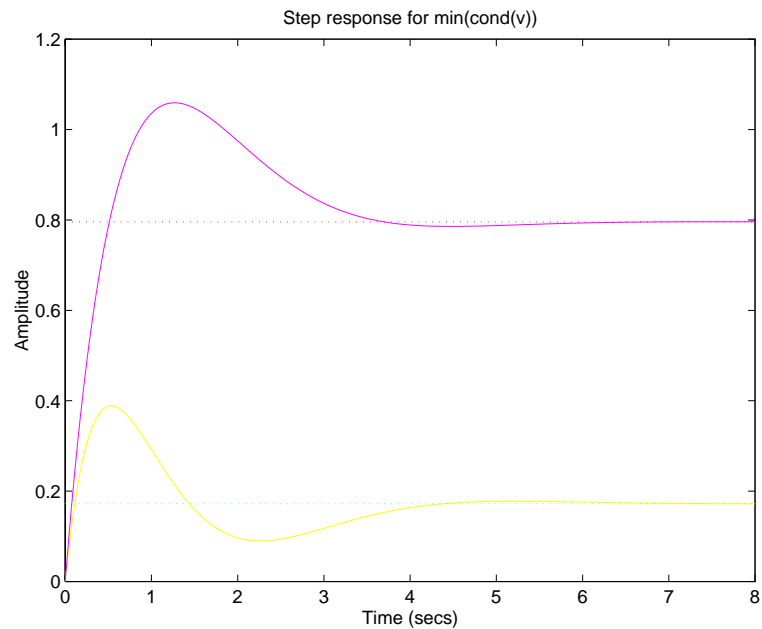
Фигура 4.6:  $\|H\|_{\infty}$  като функция на параметрите  $f_1$  и  $f_4$  около минимума



Фигура 4.7: Преходни процеси при минимално  $\|H\|_{\infty}$



Фигура 4.8: Обусловеност на матрицата от собствените вектори като функция на параметрите  $f_3$  и  $f_4$  около минимума



Фигура 4.9: Преходни процеси при минимално  $\text{cond}_2(V)$

За оптималното решение на съответната задача се изчисляват и стойностите на останалите целеви функции, за да се покаже разликата между свойствата на трите затворени системи, оптимални по отношение на различни целеви функции. Резултатите са представени в Таблица 4.1.

Таблица 4.1.

Оптимални регулатори на състоянието

Целева функция	$\ H\ _2$	$\ H\ _\infty$	$\text{cond}_2(V)$
$\min\ H\ _2$	$1.07 \times 10^0$	$1.27 \times 10^0$	$1.94 \times 10^2$
$\min\ H\ _\infty$	$1.65 \times 10^0$	$1.17 \times 10^0$	$1.52 \times 10^2$
$\min \text{cond}_2(V)$	$2.05 \times 10^0$	$1.79 \times 10^0$	$2.57 \times 10^1$

На фиг. 4.4–4.9 са показани формите на целевите функции около техните екстремуми при варирането на някои от свободните елементи на собствените вектори, както и преходните процеси на съответните оптимални затворени системи при единични стъпаловидни смущаващи въздействия. Видът на целевите функции потвърждава необходимостта от търсене на глобален екстремум.

## 4.5 Синтез на наблюдатели на състоянието

Системата от  $\ell$ -ти ред

$$\dot{z}(t) = Fz(t) + Gy(t) + Hu(t), \quad z(0) = z_0 \quad (4.24)$$

се нарича *наблюдател на състоянието* (или *наблюдател на Луенбергер*) на системата

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \quad x(0) = x_0 \\ y(t) &= Cx(t) \end{aligned} \quad (4.25)$$

ако за всяко начално състояние  $x_0$  на (4.25) съществува начално състояние  $z_0$  на (4.24), такова че

$$z(t) = Tx(t) \quad t \geq 0 \quad (4.26)$$

за всяко  $u(t)$ ,  $t \geq 0$  и някаква постоянна  $\ell \times n$  матрица  $T$ .

Входове на наблюдателя са изходът  $y(t)$  и управлението  $u(t)$  на системата (4.25), а негов изход е векторът  $z(t)$ , за който се предполага, че е достъпен за измерване. Зависимостта (4.26) дава възможност посредством измерване на  $z(t)$  да се получи апроксимация на вектора на състоянието  $x(t)$ .

При синтеза на наблюдатели на състоянието основно значение има следният резултат.

**Теорема 4.3** *Системата (4.24) е наблюдател на състоянието на системата (4.25), ако съществуват матрици  $G$  и  $T$ , които удовлетворяват линейното матрично уравнение*

$$TA - FT = GC \quad (4.27)$$

и матрицата  $H$  е определена от

$$H = TB. \quad (4.28)$$

За да се удовлетвори условието (4.26) е необходимо началното условие на наблюдателя да се избере като

$$z_0 = Tx_0. \quad (4.29)$$

Доказателството на теоремата се извършва лесно, замествайки (4.26) в (4.24), вземайки предвид (4.25).

Уравнението (4.27) е частен случай на уравнението на Силвестър. За дадена матрица  $G$  това уравнение има единствено решение за  $T$ , ако и само ако матриците  $A$  и  $F$  нямат общи собствени стойности.

Ако условието (4.29) не е изпълнено, вместо (4.26) съществува зависимостта

$$z(t) = Tx(t) + e(t),$$

където векторът на грешката  $e(t)$  е решение на хомогенното диференциално уравнение

$$\dot{e}(t) = Fe(t), \quad e(0) = z_0 - Tx_0.$$

### Наблюдатели на състоянието от пълен ред

Синтезът на наблюдатели на състоянието е удобно да се разгледа най-напред за т.н. *наблюдатели от пълен ред* (или *идентични наблюдатели*). Системата от  $n$ -ти ред

$$\dot{\hat{x}}(t) = F\hat{x}(t) + Gy(t) + Hu(t), \quad \hat{x}(0) = \hat{x}_0 \quad (4.30)$$

е идентичен наблюдател на (4.25), ако условието  $\hat{x}_0 = x_0$  води до  $\hat{x}(t) = x(t)$ ,  $t \geq 0$ .

В съответствие с (4.27), (4.28) и (4.29) може да се покаже, че (4.30) е идентичен наблюдател на (4.25), ако и само ако

$$F = A - KC, \quad G = K, \quad H = B \quad (4.31)$$

и  $K$  е матрица, която се определя по-нататък.

Идентичният наблюдател се представя във вида

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K[y(t) - C\hat{x}(t)].$$

От това уравнение следва, че наблюдателят от пълен ред представлява модел на системата (4.25) с допълнително управление, което е пропорционално на изходната грешка  $y(t) - \hat{y}(t)$ , където  $\hat{y}(t) = C\hat{x}(t)$  е реконструираният изход на системата.

Наблюдателят от пълен ред може да се представи и във вида

$$\dot{\hat{x}}(t) = (A - KC)\hat{x}(t) + Bu(t) + Ky(t) \quad (4.32)$$

откъдето се вижда, че устойчивостта на наблюдателя се определя от собствените стойности на матрицата  $A - KC$ .

Нека (4.30) е идентичен наблюдател на състоянието на системата (4.25). Грешката

$$e(t) = x(t) - \hat{x}(t),$$

с която се реконструира състоянието на (4.25), удовлетворява хомогенното диференциално уравнение

$$\dot{e}(t) = (A - KC)e(t), \quad e(0) = x_0 - \hat{x}_0. \quad (4.33)$$

Грешката на реконструирането ще затихва асимптотически, т.е.  $e(t) \rightarrow 0$  при  $t \rightarrow \infty$ , ако и само ако наблюдателят е асимптотически устойчив. По такъв начин устойчивостта на наблюдателя на състоянието и асимптотическото поведение на грешката на реконструирането се определят от свойствата на матрицата  $A - KC$ . Ето защо задачата за синтез на наблюдател от пълен ред се състои в намиране на такава матрица  $K$ , която да обезпечи желаната динамика на (4.33). Условиата за съществуване на тази матрица се дават от следната теорема.

**Теорема 4.4** *Собствените стойности на матрицата  $A - KC$  могат да се разположат на произволни симетрични места в комплексната равнина посредством подходящ избор на матрицата  $K$ , ако и само ако двойката  $(C, A)$  е напълно наблюдаема.*

Доказателството на тази теорема се извършва, като се разгледа двойката  $(A^T, C^T)$ , която е напълно управляема, ако двойката  $(C, A)$  е напълно наблюдаема. В съответствие с Теорема 4.2 винаги е възможно да се намери матрица  $K^T$ , която помества собствените стойности на  $A^T - C^T K^T$  на желаните места в комплексната равнина.

Ако двойката  $(C, A)$  не е напълно наблюдаема, всички непреместваеми собствени стойности отговарят на ненаблюдаемите моди на системата.

Нека разгледаме едновходовата едноизходна система

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), \\ y(t) &= cx(t), \end{aligned}$$

където двойката  $(c, A)$  е напълно наблюдаема. Използвайки неособеното преобразуване

$$x(t) = S\tilde{x}(t)$$

тази система се преобразува в каноничната наблюдаема форма

$$\begin{aligned} \dot{\tilde{x}}(t) &= A_o\tilde{x}(t) + b_o u(t), \\ y(t) &= c_o\tilde{x}(t), \end{aligned}$$

където

$$A_o = S^{-1}AS = \begin{bmatrix} 0 & 0 & \dots & 0 & -a_1 \\ 1 & 0 & \dots & 0 & -a_2 \\ 0 & 1 & \dots & 0 & -a_3 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -a_n \end{bmatrix},$$

$$b_o = S^{-1}b,$$

$$c_o = cS = [0, 0, \dots, 1].$$

Всичките  $(n)$  собствени стойности на  $A_o - k_o c_o$ ,  $k_o = S^{-1}k$ , могат да се разположат на произволни места посредством  $n$ -мерния вектор

$$k_o = [k_1, k_2, \dots, k_n]^T,$$

тъй като

$$A_o - k_o c_o = \begin{bmatrix} 0 & 0 & \dots & 0 & -a_1 - k_1 \\ 1 & 0 & \dots & 0 & -a_2 - k_2 \\ 0 & 1 & \dots & 0 & -a_3 - k_3 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -a_n - k_n \end{bmatrix}$$

и

$$\det(\lambda I - A_o + k_o c_o) = \lambda^n + (a_n + k_n)\lambda^{n-1} + \dots + a_1 + k_1.$$

Синтезът наблюдатели от пълен ред за многоизходна система може да се извърши по аналогичен начин, привеждайки системата (4.25) в наблюдаема форма на Луенбергер

$$\begin{aligned} \dot{\tilde{x}}(t) &= A\tilde{x}(t) + Bu(t), \\ y(t) &= C\tilde{x}(t), \end{aligned}$$

където

$$A_o = \left[ \begin{array}{cccc|cccc|c|cccc} 0 & \dots & 0 & \times & 0 & \dots & 0 & \times & & 0 & \dots & 0 & \times \\ 1 & \dots & 0 & \times & 0 & \dots & 0 & \times & & 0 & \dots & 0 & \times \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \dots & \vdots & & \vdots & \vdots \\ 0 & \dots & 1 & \times & 0 & \dots & 0 & \times & & 0 & \dots & 0 & \times \\ \hline 0 & \dots & 0 & \times & 0 & \dots & 0 & \times & & 0 & \dots & 0 & \times \\ 0 & \dots & 0 & \times & 1 & \dots & 0 & \times & & 0 & \dots & 0 & \times \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \dots & \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & \times & 0 & \dots & 1 & \times & & 0 & \dots & 0 & \times \\ \hline & & \vdots & & & & \dots & & \ddots & & & \vdots & \\ \hline 0 & \dots & 0 & \times & 0 & \dots & 0 & \times & & 0 & \dots & 0 & \times \\ 0 & \dots & 0 & \times & 0 & \dots & 0 & \times & & 1 & \dots & 0 & \times \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \dots & \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & \times & 0 & \dots & 0 & \times & & 0 & \dots & 1 & \times \end{array} \right],$$

$$C_o = \left[ \begin{array}{cccc|cccc|c|cccc} 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & \times & 0 & \dots & 0 & 1 & & 0 & \dots & 0 & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \dots & \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & \times & 0 & \dots & 0 & \times & & 0 & \dots & 0 & 1 \end{array} \right].$$

В случая матрицата  $F_o = A_o - K_o C_o$  има нетривиални колони само в позициите  $\sigma_1, \sigma_2, \dots, \sigma_r$ ,  $\sigma_i = \sum_{j=1}^i q_j$  и желаното множество от собствени стойности на  $F_o$  може да се постигне, ако колоните на матрицата  $K_o$  се изберат така, че поддиагоналните или наддиагоналните блокове на  $F_o$  се нулират, а последните колони на диагоналните блокове се променят по подходящ начин. Тогава матрицата  $F_o$  добива вида (ако се нулират поддиагоналните блокове)

$$F_o = \begin{bmatrix} F_{11} & F_{12} & \dots & F_{1r} \\ & F_{22} & \dots & F_{2r} \\ & & \ddots & \vdots \\ 0 & & & F_{rr} \end{bmatrix},$$

където елементите в последните колони на диагоналните блокове (коэффициентите на характеристичните полиноми, взети с обратен знак) се определят от желаните собствени стойности на тези блокове.

Полюсите на наблюдателя от пълен ред (собствените стойности на матрицата  $A - KC$ ) теоретически могат да се разположат произволно наляво в комплексната равнина, което да доведе до извънредно бърза сходимост на грешката  $e(t)$ . Това обаче води и до големи коефициенти на усилване, което може да причини нежелани нелинейни ефекти, както и повишена чувствителност към шумовете. От друга страна, бавната сходимост на грешката може да доведе до неудовлетворителни преходни процеси, особено ако началните условия на  $x(t)$  и  $\hat{x}(t)$  се различават значително. Ето защо полюсите на наблюдателя трябва да се разполагат така, че той да е малко "по-бърз" от системата, чието състояние трябва да се реконструира.

Приложението на наблюдателя от пълен ред включва конструиране на  $n$ -мерна динамична система, чийто  $n$  компоненти на състоянието  $\hat{x}_i(t)$  апроксимират динамичното поведение на съответните компоненти  $x_i(t)$  на състоянието на дадената система. Трябва да се отбележи, че  $r$  изходни променливи  $y(t)$  са достъпни за измерване, като при  $\text{rank}(C) = r$  всяка отделна компонента  $y_i(t)$  представлява известна линейна комбинация на променливите на състоянието. Ето защо конструирането на  $n$ -мерен наблюдател не е необходимо. Възможно е да се използва наблюдател с понижена размерност  $n - r$ , за да се реконструират само тези линейни комбинации на състоянието, които не присъстват в  $Cx(t) = y(t)$ .

### Наблюдатели на състоянието от понижен ред

Наблюдателят от понижен ред има вида (4.24) с  $\ell = n - r$ . Ако условията (4.27) и (4.28) се удовлетворяват, за оценка на състоянието може да се използва вектора

$$\hat{x}(t) = \begin{bmatrix} T \\ C \end{bmatrix}^{-1} \begin{bmatrix} z(t) \\ y(t) \end{bmatrix},$$



при условие, че матрицата  $\begin{bmatrix} T \\ C \end{bmatrix}$  е неособена. Възможно е да се покаже, че това условие винаги може да се изпълни посредством подходящ избор на матриците  $F$  и  $G$ , ако системата е напълно наблюдаема.

Ако условието (4.29) е удовлетворено, то  $\hat{x}(t) = x(t)$  и състоянието на системата се реконструира точно. В общият случай ( $z_0 \neq Tx_0$ ) имаме, че

$$\hat{x}(t) = x(t) - \begin{bmatrix} T \\ C \end{bmatrix}^{-1} \begin{bmatrix} e(t) \\ 0 \end{bmatrix}, \quad (4.34)$$

където  $(n - r)$ -мерният вектор  $e(t)$  е решение на

$$\dot{e}(t) = Fe(t), \quad e(0) = z_0 - Tx_0.$$

Желаното асимптотично поведение на  $e(t)$  може да се постигне посредством подходящ избор на собствените стойности на  $F$  (полусите на наблюдателя).

По-долу се разглежда един от възможните методи за синтез на наблюдатели от понижен ред. В този метод се задават матриците  $F$  и  $G$ , като за асимптотическа устойчивост на наблюдателя е необходимо собствените стойности на  $F$  да имат отрицателни реални части. Обикновено матрицата  $F$  се избира като (квази)диагонална, което дава възможност лесно да се променят полусите на наблюдателя. При зададени  $F$  и  $G$  уравнението (4.27) може да се реши по различни начини. От изчислителна гледна точка е най-подходящ методът на Бартелс-Стюърт, разгледан в т.7.8. За да се осигури единствено решение на (4.27), матрицата  $F$  трябва да се избере така, че нейните собствени стойности да са различни от собствените стойности на  $A$ . След това трябва да се провери дали матрицата  $\begin{bmatrix} T \\ C \end{bmatrix}$  е неособена. Ако това условие не е изпълнено, трябва да се зададат други матрици  $F$  и  $G$ , за които процедурата се повтаря.

**Пример 4.6** Разглежда се системата (4.25) с матрици

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 2 & 0 \\ -3 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}.$$

Собствените стойности на  $A$  са

$$\lambda_1 = 2, \quad \lambda_2 = 3, \quad \lambda_3 = -2.$$

Тъй като  $\text{rank} C = 2$ , наблюдателят от понижен ред ще бъде от първи ред. При зададени матрици

$$F = -5 \quad \text{и} \quad G = [9, -10],$$

получаваме

$$T = [1, 1, 2],$$

така че

$$H = [1, 5], \quad \begin{bmatrix} T \\ C \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}.$$

От (4.34) следва, че

$$\begin{aligned} x_1(t) &= \hat{x}_1(t) + e(t), \\ x_2(t) &= \hat{x}_2(t), \\ x_3(t) &= \hat{x}_3(t), \end{aligned}$$

където грешката  $e(t) = e^{-5t}e_0$  е решение на

$$e(t) = -5e(t), \quad e_0 = z_0 - x_1(0) - x_2(0) - 2x_3(0).$$

## 4.6 Реализация на закони за управление с наблюдатели

Основното приложение на наблюдателя на Луенбергер е да се реализира приблизително обратна връзка по състоянието, използвайки оценката  $\hat{x}(t)$  вместо недостъпният вектор на състоянието  $x(t)$ .

Нека е даден законът на управление по състоянието

$$u(t) = -K_r x(t) + Pr(t), \quad (4.35)$$

където матрицата  $K_r$  е определена така, че затворената система

$$\dot{x}(t) = (A - BK_r)x(t) + BPr(t) \quad (4.36)$$

да има желаната динамика (напр. желано разположение на полюсите),  $P$  е матрицата на правата връзка и  $r(t)$  е външен вход.

Използвайки оценката  $\hat{x}(t)$ , получена с помощта на наблюдател на състоянието от пълен ред

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K_o[y(t) - C\hat{x}(t)], \quad (4.37)$$

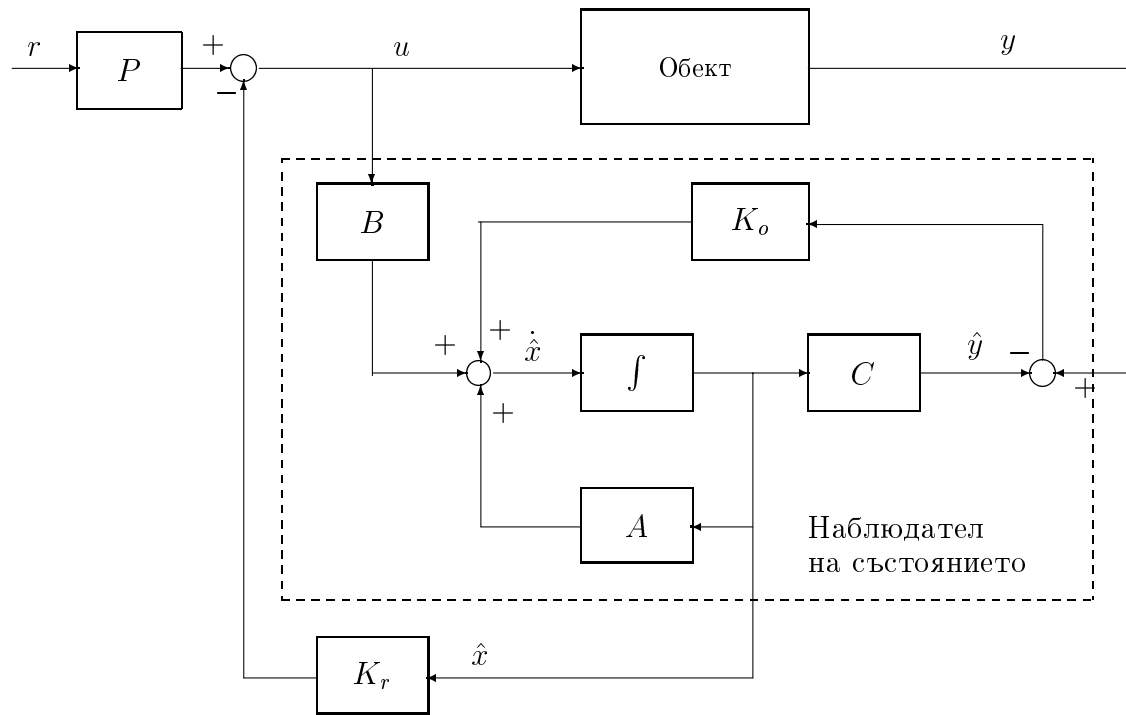
законът за управление (4.35) се реализира във вида

$$u(t) = -K_r \hat{x}(t) + Pr(t). \quad (4.38)$$

Структурната схема на затворената система с наблюдател на състоянието е показана на фиг. 4.10.

Динамиката на разширената система от  $2n$ -ти ред се описва с уравнението

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\hat{x}}(t) \end{bmatrix} = \begin{bmatrix} A & -BK_r \\ K_o C & A - BK_r - K_o C \end{bmatrix} \begin{bmatrix} x(t) \\ \hat{x}(t) \end{bmatrix} + \begin{bmatrix} BP \\ BP \end{bmatrix} r(t). \quad (4.39)$$



Фигура 4.10: Реализация на закон за управление с наблюдател от пълен ред

При условие, че са изпълнени зависимостите (4.31), чрез неособеното преобразуване

$$\begin{bmatrix} x(t) \\ e(t) \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ -I_n & I_n \end{bmatrix} \begin{bmatrix} x(t) \\ \hat{x}(t) \end{bmatrix},$$

системата (4.39) се привежда в

$$\begin{bmatrix} \dot{x}(t) \\ \dot{e}(t) \end{bmatrix} = \begin{bmatrix} A - BK_r & -BK_r \\ 0 & A - K_o C \end{bmatrix} \begin{bmatrix} x(t) \\ e(t) \end{bmatrix} + \begin{bmatrix} BP \\ 0 \end{bmatrix} r(t). \quad (4.40)$$

От (4.40) се вижда, че  $n$ -мерната подсистема за грешката  $e(t)$  е неуправляема, вследствие на това, че  $e(t)$  трябва да се стреми асимптотически към нулата, независимо от входа  $r(t)$ . Също така

$$\det(\lambda I - \begin{bmatrix} A - BK_r & -BK_r \\ 0 & A - K_o C \end{bmatrix}) = \det(\lambda I - A + BK_r) \det(\lambda I - A + K_o C),$$

т.е. полюсите на разширената система са собствените стойности на  $A - BK_r$  и  $A - K_o C$ . Ето защо наблюдателят не променя полюсите на затворената система с обратна връзка по състоянието, а само добавя своите. Ако полюсите на наблюдателя са разположени наляво от собствените стойности на  $A - BK_r$ , динамиката на системата (4.40) по отношение на вектора  $x(t)$  ще се различава незначително от динамиката на системата с обратна връзка по състоянието (4.36).

По такъв начин за напълно управляема и напълно наблюдаема система от  $n$ -ти ред може да се синтезира допълнителна динамична система от  $n$ -ти ред,

така че всичките  $2n$  полюси на разширената система да са разположени на предварително зададени места. По-общият резултат е, че стабилизируема и детектируема система винаги може да се направи устойчива с помощта на обратна връзка, реализирана посредством наблюдател на състоянието.

Уравненията (4.37), (4.38) описват управляващо устройство във вида

$$u(s) = -K(s)y(s) + Pr(s),$$

където

$$K(s) = \left[ \frac{A - BK_r - K_oC}{K_r} \mid \frac{K_o}{0} \right].$$

**Пример 4.7** Дадена е непрекъснатата система от 5-ти ред

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t),\end{aligned}$$

за която

$$A = \begin{bmatrix} 0 & 0 & 1.1320 & 0 & -1.0000 \\ 0 & -0.0538 & -0.1712 & 0 & 0.0705 \\ 0 & 0 & 0 & 1.0000 & 0 \\ 0 & 0.0485 & 0 & -0.8556 & -1.0130 \\ 0 & -0.2909 & 0 & 1.0532 & -0.6859 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & 0 & 0 \\ -0.1200 & 1.0000 & 0 \\ 0 & 0 & 0 \\ 4.4190 & 0 & -1.6650 \\ 1.5750 & 0 & -0.0732 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

За системата се синтезира закон на управление по състоянието (4.35), който разполага полюсите на затворената система в  $-5+6j$ ,  $-5-6j$ ,  $-6$ ,  $-7$ ,  $-8$  и наблюдател на състоянието от пълен ред (4.37), чийто полюси са  $-10$ ,  $-11$ ,  $-12$ ,  $-13$  и  $-14$ . Синтезът се извършва с функцията `polsyn`, показана в табл. 4.2. `polsyn` използва функцията `place` от MATLAB за робастен синтез при зададени полюси. Като резултат се получава (до втория знак)

$$K_r = \begin{bmatrix} -20.65 & -0.28 & -4.36 & 0.40 & 7.25 \\ -0.03 & 7.84 & -0.57 & -0.12 & 0.06 \\ -183.09 & -0.18 & -62.84 & -6.81 & 40.90 \end{bmatrix},$$

$$K_o = \begin{bmatrix} 22.73 & -2.55 & 1.05 \\ -0.80 & 12.15 & -0.15 \\ 1.93 & 0.18 & 23.52 \\ 32.27 & -1.44 & 126.26 \\ -117.67 & 28.19 & 13.03 \end{bmatrix}, \quad P = \begin{bmatrix} -20.65 & -0.07 & 4.37 \\ -0.03 & 7.92 & -0.35 \\ -183.09 & 0.41 & -15.84 \end{bmatrix}.$$

Таблица 4.2: MATLAB функция за синтез при зададени полюси на обратна връзка по състоянието и наблюдател от пълен ред

---

```
function [Kr,Ko,P,At,Bt,Ct] = polsyn(A,B,C,wr,wo)
%
% Синтез при зададени полюси на обратна връзка по състоянието и
% наблюдател от пълен ред
%
% ВХОДНИ ДАННИ:
% A,B,C:      Описание в пространство на състоянията на системата.
% wr:         Вектор, съдържащ реалните и имагинерните части на
%             желаните полюси на системата.
% wo:         Вектор, съдържащ реалните и имагинерните части на
%             полюсите на наблюдателя.
%
% ИЗХОДНИ ДАННИ:
% Kr:         Матрица на обратната връзка по състоянието,
%             разполагаща полюсите на желаните места.
% Ko:         Матрица на усилването на наблюдателя.
% P:         Матрица на правата връзка, осигуряваща равенство на
%             изхода и заданието в установен режим.
% At,Bt,Ct:   Матрици на разширената система, включваща обратната
%             връзка по оценката на състоянието и наблюдателя.
%
%
% Определяне на матрицата на обратната връзка
%
Kr = place(A,B,wr);
%
% Определяне на матрицата на правата връзка
%
P = inv(-C*((A-B*Kr)\B));
%
% Определяне на матрицата на усилването на наблюдателя
%
Ko = place(A',C',wo);
Ko = Ko';
%
% Изчисляване на матриците на затворената система
%
At = [ A      -B*Kr
      Ko*C  A-B*Kr-Ko*C];
Bt = [B*P
      V*P];
```

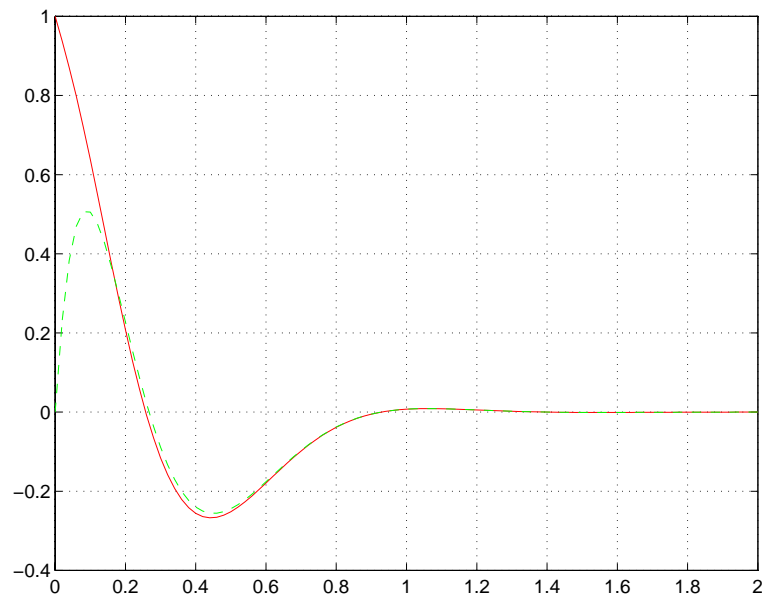
$$Ct = [C \ 0 * C];$$


---

Преходните процеси на разширената система за  $x_1(t)$  и  $\hat{x}_1(t)$  при начални условия

$$x_0(t) = \begin{bmatrix} 1 \\ 3 \\ -1 \\ -3 \end{bmatrix} \quad \text{и} \quad \hat{x}_0(t) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

са показани на фиг. 4.11. Вижда се, че при избраните полюси динамиката на наблюдателя е на порядък по-бърза от динамиката на системата, чието състояние се оценява.



Фигура 4.11: Преходни процеси в система с наблюдател на състоянието (с непрекъснатата линия е означено  $x_1(t)$ , а с прекъснатата –  $\hat{x}_1(t)$ )

Нека сега законът за управление (4.35) е реализиран с помощта на наблюдател на състоянието от понижен ред (4.24) във вида

$$u(t) = -K_r \hat{x}(t) + Pr(t) = -Ez(t) - Dy(t) + Pr(t),$$

където

$$[E \ D] = K_r \begin{bmatrix} T \\ C \end{bmatrix}^{-1}.$$

Динамиката на разширената система от  $(2n - r)$ -ти ред се описва с уравнението

$$\begin{bmatrix} \dot{x}(t) \\ \dot{z}(t) \end{bmatrix} = \begin{bmatrix} A - BDC & -BE \\ GC - HDC & F - HE \end{bmatrix} \begin{bmatrix} x(t) \\ z(t) \end{bmatrix} + \begin{bmatrix} BP \\ BP \end{bmatrix} r(t). \quad (4.41)$$

При условие, че са изпълнени зависимостите (4.27) и (4.28), чрез неособеното преобразуване

$$\begin{bmatrix} x(t) \\ e(t) \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ -T & I_{n-r} \end{bmatrix} \begin{bmatrix} x(t) \\ z(t) \end{bmatrix},$$

системата (4.41) се привежда в

$$\begin{bmatrix} \dot{x}(t) \\ \dot{e}(t) \end{bmatrix} = \begin{bmatrix} A - BK_r & -BE \\ 0 & F \end{bmatrix} \begin{bmatrix} x(t) \\ e(t) \end{bmatrix} + \begin{bmatrix} BP \\ 0 \end{bmatrix} r(t).$$

От това уравнение следва, че подобно на случая, при който се използва наблюдател на състоянието от пълен ред, полюсите на разширената система са собствените стойности на  $A - BK$  и  $F$ .

## 4.7 Синтез на дискретни наблюдатели

Наблюдателят на състоянието на дискретна система

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k, \\ y_k &= Cx_k \end{aligned}$$

се взема във вида

$$z_{k+1} = Fz_k + Gy_k + Hu_k,$$

където матриците  $F, G$  и  $H$  отново удовлетворяват (4.27) и (4.28). Собствените стойности на  $F$  трябва да лежат вътре в единичната окръжност, което гарантира сходимостта на грешката на оценяването. За разлика от непрекъснатия случай, грешката на оценяването може да се нулира за краен брой стъпки, избирайки собствените стойности на  $F$  равни на нула. Наблюдатели с такова свойство се наричат *наблюдатели с безкрайна степен на устойчивост*. Използвайки наблюдаемата канонична форма може да се покаже, че наблюдателят от пълен ред на едноизходна система ще получи точна оценка на вектора на състоянието за  $n$  стъпки. Използвайки каноничната наблюдаема форма на Луенбергер, в многоизходния случай матрицата на наблюдателя от пълен ред  $A - KC$  може да се направи подобна на матрицата

$$F_o = \left[ \begin{array}{cccc|cccc|c|cccc} 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & & 0 & \dots & 0 & 0 \\ 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & & 0 & \dots & 0 & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \dots & \vdots & & \vdots & \vdots \\ 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & & 0 & \dots & 0 & 0 \\ \hline 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & & 0 & \dots & 0 & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \dots & \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 & & 0 & \dots & 0 & 0 \\ \hline & & \vdots & & & & \dots & & \ddots & & & \vdots & \\ \hline 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & & 1 & \dots & 0 & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \dots & \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & & 0 & \dots & 1 & 0 \end{array} \right],$$

от което следва, че  $(A - KC)^q = 0$ , т.е. точна оценка на състоянието ще се получи за  $q$  стъпки, където  $q$  е индексът на наблюдаемост на системата.

Използвайки резултатите от т. 4.3 за синтез на дискретни системи със зададени полюси, получаваме следната теорема.

**Теорема 4.5** *Ако дискретната многомерна система е напълно управляема и напълно наблюдаема, то съществува закон на управление, реализиран чрез наблюдател от пълен ред, за който състоянието на затворената система ще затихва за  $p+q$  стъпки, където  $p$  е индексът на управляемост, а  $q$  – индексът на наблюдаемост на системата.*



## Глава 5

# $\mathcal{H}_2$ ОПТИМАЛНО УПРАВЛЕНИЕ

В тази глава се разглежда оптималното управление на линейни стационарни системи с квадратичен критерий на качеството. Оптималното управление, основаващо се на теорията на оптималната филтрация, разработена от Винер и Колмогоров през четиридесетте години на двадесети век, достигна етапа на своята зрялост през шестдесетте години с това, което сега се нарича линейно–квадратично гаусово (LQG) управление. Това управление бе използвано успешно при решаването на редица практически задачи, но то не получи голямо разпространение при управлението на промишлени процеси. Причината за това е, че в повечето случаи не са известни точните модели на обектите, а предположенията за смущения във вид на бели шумове обикновено не са в сила. В резултат на това синтезът, основаващ се на LQG управление, не е достатъчно робастен, за да се използва на практика. Поради тази причина през осемдесетте години по-широко разпространение получиха методите, основани на  $H_\infty$  оптималното управление.

### 5.1 Линейно–квадратична оптимизация

В стандартното LQG управление се предполага, че динамиката на обекта е линейна и известна и че шумът на измерването и смущаващите сигнали (шумът на процеса) са стохастични процеси с известни статистически характеристики. По такъв начин имаме, че моделът на обекта е

$$\dot{x} = Ax + Bu + w_d, \quad (5.1)$$

$$y = Cx + w_n \quad (5.2)$$

където  $w_d$  и  $w_n$  са смущението (шумът в процеса) и шумът на измерването, съответно, които обикновено се предполагат, че са некорелирани гаусови стохастични процеси с нулеви средни стойности с постоянни матрици на спектралната плътност  $W$  и  $V$ , съответно. По такъв начин  $w_d$  и  $w_n$  са бели шумове с матрици

на ковариациите

$$\mathcal{E} \{w_d(t)w_d(\tau)^T\} = W\delta(t-\tau), \quad (5.3)$$

$$\mathcal{E} \{w_n(t)w_n(\tau)^T\} = V\delta(t-\tau) \quad (5.4)$$

и

$$\mathcal{E} \{w_d(t)w_n(\tau)^T\} = 0, \quad \mathcal{E} \{w_n(t)w_d(\tau)^T\} = 0, \quad (5.5)$$

където  $\mathcal{E}$  е операторът за математическо очакване и  $\delta(t-\tau)$  е делта-функцията.

Задачата за LQG-управление е да се намери оптималното управление  $u(t)$  което минимизира

$$J = \mathcal{E} \left\{ \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [x^T Q x + u^T R u] dt \right\}, \quad (5.6)$$

където  $Q$  и  $R$  са подходящо избрани постоянни тегловни матрици, такива че  $Q = Q^T \geq 0$  и  $R = R^T > 0$ . Наименованието LQG (от английското Linear Quadratic Gaussian) произлиза от използването на линеен модел, интегрално-квадратичен критерий и гаусов бял шум за моделиране на смущаващите сигнали и шума.

Решението на LQG задачата, известно като Теорема за разделянето, е изненадващо просто и елегантно. То се състои първо от намиране на оптималното управление на детерминистичната задача за линейно-квадратичен регулатор (LQR-от английското Linear Quadratic Regulator), т.е. задачата, формулирана по-горе без сигналите  $w_d$  и  $w_n$ . Решението на тази задача може да се запише във вида на закон за управление по състоянието

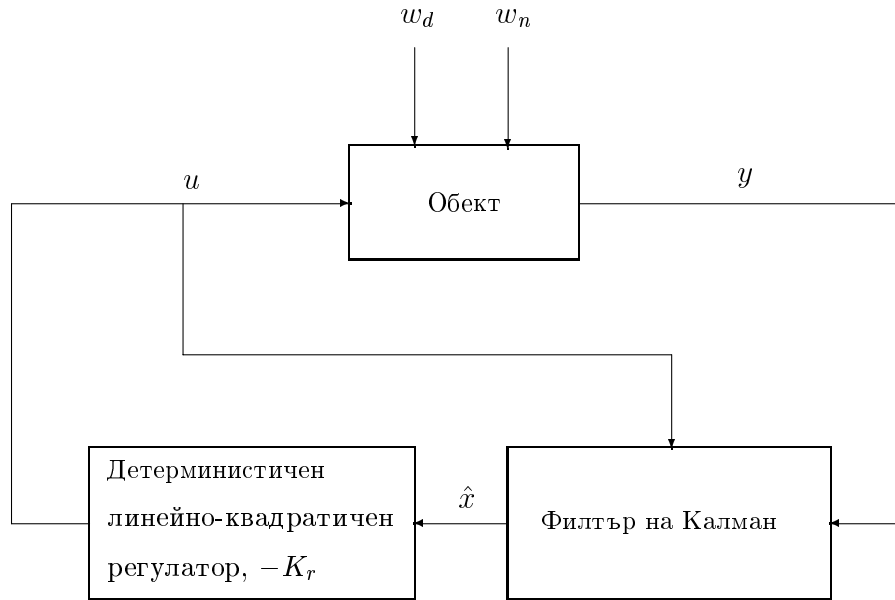
$$u(t) = -K_r x(t) \quad (5.7)$$

където  $K_r$  е постоянна матрица, която лесно може да се изчисли и очевидно е независима от матриците  $W$  и  $V$ , които характеризират статистическите характеристики на шумовете. Следващата стъпка е да се намери оптимална оценка  $\hat{x}$  на  $x$ , така че да се минимизира  $\mathcal{E} \{[x - \hat{x}]^T [x - \hat{x}]\}$ . Оптималната оценка на състоянието се получава с т.н. *филтър на Калман* и е независима от матриците  $Q$  и  $R$ . Търсеното решение на LQG задачата тогава се намира, замествайки  $x$  с  $\hat{x}$ , което дава  $u(t) = -K_r \hat{x}(t)$ . Следователно, LQG задачата и нейното решение могат да се разделят на две части, така както е илюстрирано на фиг. 5.1.

По-долу се дават уравненията, които са необходими за намиране на оптималната матрица на обратната връзка по състоянието  $K_r$  и на филтъра на Калман.

**Оптимална матрица на обратната връзка по състоянието.** LQR-задачата, в която всички състояния са известни, е детерминистична задача, която се формулира по следния начин. Дадена е системата

$$\dot{x} = Ax + Bu \quad (5.8)$$



Фигура 5.1: Теорема за разделянето

с ненулево начално състояние  $x(0)$ , да се намери входния сигнал  $u(t)$ , който привежда системата в нулевото състояние ( $x = 0$ ) по оптимален начин, т.е. минимизирайки квадратичния критерий

$$J_r = \int_0^\infty [x(t)^T Q x(t) + u(t)^T R u(t)] dt. \quad (5.9)$$

Оптималното решение (за всяко начално състояние) е

$$u(t) = -K_r x(t), \quad (5.10)$$

където

$$K_r = R^{-1} B^T X \quad (5.11)$$

и  $X = X^T \geq 0$  е единственото положително полуопределено решение на матричното алгебрично уравнение на Рикати

$$A^T X + X A - X B R^{-1} B^T X + Q = 0. \quad (5.12)$$

**Филтър на Калман.** Филтърът на Калман има структурата на обикновен оценител на състоянието, така както е показан на фиг. 5.2, с

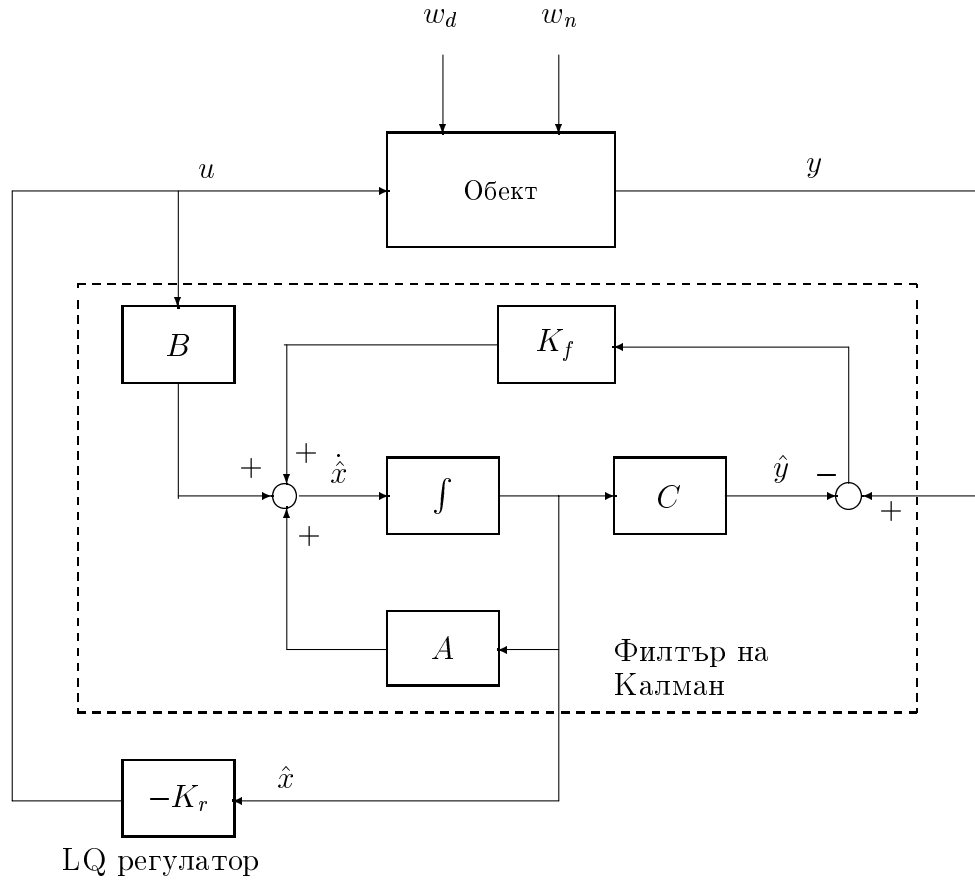
$$\dot{\hat{x}} = A \hat{x} + B u + K_f (y - C \hat{x}). \quad (5.13)$$

Оптималният избор на  $K_f$ , който минимизира  $E \{[x - \hat{x}]^T [x - \hat{x}]\}$ , се дава от

$$K_f = Y C^T V^{-1}, \quad (5.14)$$

където  $Y = Y^T \geq 0$  е единственото положително полуопределено решение на алгебричното уравнение на Рикати

$$Y A^T + A Y - Y C^T V^{-1} C Y + W = 0. \quad (5.15)$$



Фигура 5.2: LQG регулатор и зашумен обект

**LQG: Комбинирана оптимална оценка на състоянието и оптимална обратна връзка по състоянието.** Задачата за LQG управление е да се минимизира  $J$  в (5.6). Структурата на LQG управляващото устройство е илюстрирана на фиг. 5.2; неговата предавателна матрица от  $y$  към  $u$  (т.е. предполагайки положителна обратна връзка), се дава от

$$\begin{aligned}
 K_{LQG}(s) &\stackrel{\text{def}}{=} \left[ \begin{array}{c|c} \frac{A - BK_r - K_f C}{-K_r} & K_f \\ \hline & 0 \end{array} \right] \\
 &= \left[ \begin{array}{c|c} \frac{A - BR^{-1}B^T X - YC^T V^{-1}C}{-R^{-1}B^T X} & YC^T V^{-1} \\ \hline & 0 \end{array} \right]. \quad (5.16)
 \end{aligned}$$

То има същият ред (брой на полюсите), както обекта.

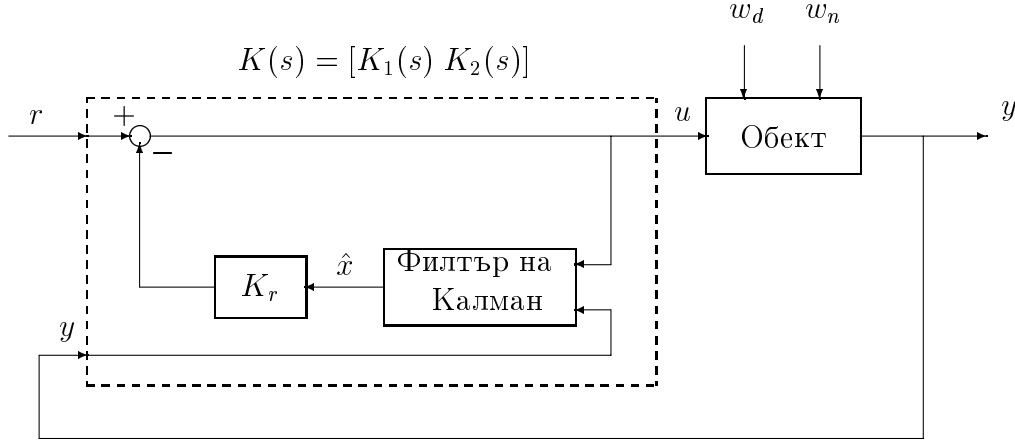
Трябва да се отбележи, че оптималните матрици  $K_f$  и  $K_r$  съществуват и LQG-управляемата система е вътрешно устойчива, при условие, че системите с реализации в пространство на състоянията  $(A, B, Q^{\frac{1}{2}})$  и  $(A, W^{\frac{1}{2}}, C)$  са стабилизируеми и детектируеми.

За системата и LQG управляващото устройство от фиг. 5.2 може да се пока-

же, че динамиката на затворената система се описва от

$$\frac{d}{dt} \begin{bmatrix} x \\ x - \hat{x} \end{bmatrix} = \begin{bmatrix} A - BK_r & BK_r \\ 0 & A - K_f C \end{bmatrix} \begin{bmatrix} x \\ x - \hat{x} \end{bmatrix} + \begin{bmatrix} I & 0 \\ I & -K_f \end{bmatrix} \begin{bmatrix} w_d \\ w_n \end{bmatrix}.$$

Това показва, че множеството от полюсите на затворената система е обединение на полюсите на детерминираната LQR система (собствените стойности на  $A - BK_r$ ) и полюсите на филтъра на Калман (собствените стойности на  $A - K_f C$ ). Точно това е, което би трябвало да се очаква от Теоремата на разделянето.



Фигура 5.3: Модифицирано LQG управляващо устройство със задаващ сигнал

На фиг. 5.3 в конфигурацията LQG-управляващо устройство – обект е въведен задаващ вход  $r$  и управляващото устройство е модифицирано, за да се покаже неговата двустепенна структура. Може да се покаже, че предавателната матрица на управляващото устройство  $K(s)$ , свързваща  $[r^T y^T]$  с  $u$ , е

$$\begin{aligned} K(s) &= [K_1(s) \ K_2(s)] \\ &= -K_r(sI - A + BK_r + K_f C)^{-1} [B \ K_f] + [I \ 0]. \end{aligned}$$

Ако LQG управляващото устройство е устойчиво, за което няма гаранция, то  $K_1(s)$  може да се приложи като пред-филтър, а  $K_2(s)$  – като управляващо устройство в обратната връзка. Тъй като, обаче, редът (броят на състоянията) на  $K_1(s)$ ,  $K_2(s)$  и  $K_{LQG}(s)$  е един и същ, се получава усложняване при прилагането на  $K_1(s)$  и  $K_2(s)$  поотделно, поради което това не се препоръчва.

В Таблица 5.1 е дадена програма за MATLAB, с помощта на която може да се синтезира LQG оптимално управляващо устройство.

## 5.2 Робастни свойства

Възможно е да се покаже, че LQR-оптималната система притежава много добри свойства по отношение на робастността. Ако тегловната матрица  $R$  е избрана като диагонална, то функцията на чувствителността  $S = (I + K_r(sI - A)^{-1}B)^{-1}$  удовлетворява

$$\bar{\sigma}(S(j\omega)) \leq 1, \forall \omega.$$

Таблица 5.1: MATLAB функция за синтез на LQG оптимално управляващо устройство

---

```

% Използва Robust Control Toolbox
function [Ak,Bk,Ck,Dk] = lqg(A,B,C,D,Q,R,W,V)
%
% Синтез на LQG оптимално управляващо устройство
%
% ВХОДНИ ДАННИ:
% A,B,C,D:      Описание в пространство на състоянията на обекта.
% Q,R           Тегловни матрици на състоянието и управлението.
% V,W           Матрици на спектралните плътности на шумовете в
%              обекта и измерването.
%
% ИЗХОДНИ ДАННИ:
% Ak,Bk,Ck,Dk:  Реализация в пространство на състоянията на полученото
%              оптимално управляващо устройство.
%
% Изчисляване на матрицата на оптималния регулатор
%
Kx = lqr(A,B,Q,R);
%
% Изчисляване на матрицата на оптималния оценител
%
Vnoise = eye(size(A));
Ke = lqe(A,Vnoise,C,W,V);
%
% Изчисляване на реализацията в пространство на състоянията на
% управляващото устройство
%
[Ak,Bk,Ck,Dk] = reg(A,B,C,D,Kx,Ke);

```

---

Оттук може да се покаже, че системата ще има запас по амплитуда равен на безкрайност и запас по фаза  $60^\circ$  по всеки входен канал на обекта. За едновходова система това означава, че ходографът на Найквист на предавателната функция на отворената система  $K_r(sI - A)^{-1}B$  винаги ще лежи извън единичната окръжност с център в  $-1$ .

**Пример 5.1** Да разгледаме процес от първи ред  $P(s) = 1/(s-a)$  с реализация в пространство на състоянията

$$\dot{x}(t) = ax(t) + u(t), \quad y(t) = x(t)$$

така че състоянието се измерва директно. За ненулево начално състояние, квадратичният функционал, който трябва да се минимизира, е

$$J_r = \int_0^\infty (x^2 + Ru^2)dt.$$

Алгебричното уравнение на Рикати (5.12) в случая добива вида ( $A = a$ ,  $B = 1$ ,  $Q = 1$ )

$$aX + Xa - XR^{-1}X + 1 = 0 \quad \Leftrightarrow \quad X^2 - 2aRX - R = 0$$

което, тъй като  $X \geq 0$ , дава  $X = aR + \sqrt{(aR)^2 + R}$ . Оптималното управление се дава от  $u = -K_r x$ , където от (5.11)

$$K_r = X/R = a + \sqrt{a^2 + 1/R}$$

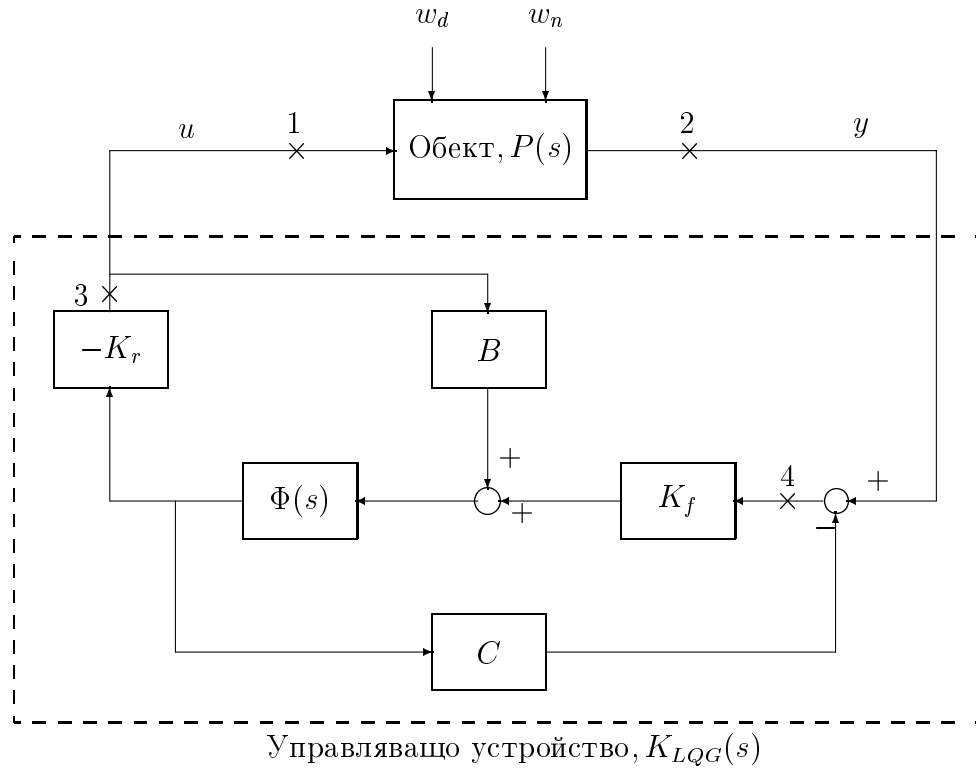
и получаваме затворената система

$$\dot{x} = ax + u = -(\sqrt{a^2 + 1/R})x.$$

Полюсът на затворената система се намира в  $s = -\sqrt{a^2 + 1/R} < 0$ . По такъв начин, ходографът на корените за полюса на оптималната затворена система по отношение на  $R$  започва от  $s = -|a|$  за  $R = \infty$  (безкрайно голямо тегло върху входа) и се движи към  $-\infty$  по реалната ос при  $R$  приближаващо към нулата. Обърнете внимание, че ходографът на корените е идентичен за устойчив ( $a < 0$ ) и неустойчив ( $a > 0$ ) обект  $P(s)$  с една и съща стойност на  $|a|$ . В частност, за  $a > 0$  се вижда, че минималната входна енергия, необходима да се стабилизира обекта (съответстваща на  $R = \infty$ ) се получава с вход  $u = -2|a|x$ , който премества полюса от  $-a$  в неговото огледално изображение  $s = -a$ .

За  $R$  малко ("икономично управление") спрягащата честота на предавателната функция на отворената система  $L = PK_r = K_r/(s - a)$  се дава приблизително от  $\omega_c \approx \sqrt{1/R}$ . Отбележете също, че  $L(j\omega)$  има наклон от -20 dB/dec във високите честоти, което е общо свойство на LQR синтеза. Освен това, ходографът на Найквист на  $L(j\omega)$  избягва единичната окръжност с център в критичната точка -1, т.е.  $|S(j\omega)| = 1/|1 + L(j\omega)| \leq 1$  при всички честоти. Това е очевидно за устойчив обект с  $a < 0$ , тъй като  $K_r > 0$  и тогава фазата на  $L(j\omega)$  се изменя от  $0^\circ$  (при нулева честота) до  $90^\circ$  (при безкрайно голяма честота). Изненадващото е, че това също е вярно за неустойчив обект с  $a > 0$ , въпреки че фазата на  $L(j\omega)$  се изменя от  $-180^\circ$  до  $-90^\circ$ .

Нека разгледаме сега филтъра на Калман показан на фиг 5.2, който също представлява динамична система. Използвайки дуалността може да се покаже, че ако матрицата на спектралната плътност се избере като диагонална, то на входа към оптималната матрица  $K_f$  ще има безкрайно голям запас по амплитуда и запас по фаза, равен на  $60^\circ$ . Ето защо за едновходова система ходографът на Найквист на предавателната функция на отворения филтър  $C(sI - A)^{-1}K_f$  ще лежи извън единичната окръжност с център в -1.



Фигура 5.4: LQG-управляем обект

Независимо от много добрите свойства на LQR-оптималната система и филтъра на Калман, взети поотделно, LQG-оптималната система не притежава гарантиран запас на устойчивост. За да се обясни този факт, нека разгледаме LQG-системата, показана на фиг. 5.4. Предавателните матрици, отговарящи на отбелязаните точки от 1 до 4 съответно са:

$$\begin{aligned}
 L_1(s) &= K_r[\Phi(s)^{-1} + BK_r + K_fC]^{-1}K_fC\Phi(s)B \\
 &= -K_{LQG}(s)G(s) \\
 L_2(s) &= -P(s)K_{LQG}(s) \\
 L_3(s) &= K_r\Phi(s)B \text{ (предавателна матрица на регулатора)} \\
 L_4(s) &= C\Phi(s)K_f \text{ (предавателна матрица на филтъра)}
 \end{aligned}$$

където

$$\Phi(s) = (sI - A)^{-1},$$

$K_{LQG}(s)$  е както в (5.16) и  $P(s) = C\Phi(s)B$  е моделът на обекта.

В точките 3 и 4 имаме гарантираните робастни свойства на LQR системата и филтъра на Калман, съответно. Но в действителният вход и изход на обекта (точки 1 и 2), където сме най-заинтересовани от постигане на добър запас на устойчивост, имаме сложни предавателни матрици, които в общия случай не дават гаранция за удовлетворителни робастни свойства. Отбележете също, че



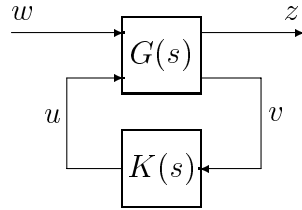
точките 3 и 4 в действителност са вътре в LQG управляващото устройство, което трябва да бъде реализирано (най-често като програмно осигуряване), така че имаме добър запас на устойчивост, където той в действителност не е необходим и липса на гаранции, където трябва.

Робастността на конкретна LQG система е възможно е да се подобри, отслабвайки изискванията по отношение на оптималността на регулатора по състоянието или на филтъра. За тази цел се използва подхода на т.н. *възстановяване на предавателните свойства на LQG-контура* (LQG/LTR). Идеята на този подход е да се синтезира закон за управление по състоянието или матрица на филтъра по такъв начин, че LQG законът за управление да апроксимира свойствата на контура на обикновеното LQR управление.

### 5.3 Стандартна задача за $\mathcal{H}_2$ оптимално управление

В тази точка се разглежда стандартната задача за  $\mathcal{H}_2$  оптимизация на многомерни системи за управление.

Дадена е системата, показана на фиг. 5.5 и разгледана преди това в гл. 3.



Фигура 5.5: Стандартна конфигурация на система за управление

Тази система се описва с уравненията

$$\begin{bmatrix} z(s) \\ v(s) \end{bmatrix} = \begin{bmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{bmatrix} \begin{bmatrix} w(s) \\ u(s) \end{bmatrix} \stackrel{\text{def}}{=} G(s) \begin{bmatrix} w(s) \\ u(s) \end{bmatrix}$$

Предполага се, че обобщеният обект  $G(s)$  има реализация в пространство на състоянията във вида

$$\begin{aligned} \dot{x} &= Ax + B_1 w + B_2 u \\ z &= C_1 x + D_{11} w + D_{12} u \\ v &= C_2 x + D_{21} w + D_{22} u \end{aligned} \tag{5.17}$$

т.е.

$$G(s) = \left[ \begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{array} \right].$$

В описанието на системата  $u$  са управляващите променливи,  $v$  са измерваните променливи,  $w$  са външните сигнали, такива като смущенията  $w_d$  и заданията  $r$

и  $z$  са т.н. "грешки", които трябва да се минимизират в някъкъв смисъл, за да се изпълнят изискванията към качеството.

Стандартната задача за  $\mathcal{H}_2$  оптимално управление се формулира по следния начин.

**Оптимално  $\mathcal{H}_2$  управление:** *да се намери управляващо устройство с правилна, реално-рационална предавателна матрица  $K$ , което стабилизира вътрешно  $G$  и минимизира  $\mathcal{H}_2$  нормата на предавателната матрица  $T_{zw}$  от  $w$  към  $z$ .*

От уравненията на системата следва, че

$$z = G_{11}w + G_{12}u, \quad v = G_{21}w + G_{22}u.$$

Използвайки  $u = Kv$ , можем да изключим  $u$  и  $v$ , като в резултат се получава

$$z = [G_{11} + G_{12}K(I - G_{22}K)^{-1}G_{21}]w. \quad (5.18)$$

Изразът, заключен в скоби в уравнение (5.18), е всъщност предавателната матрица  $T_{zw}$ , която представлява долно ЛДП  $F_\ell(G, K)$ . По такъв начин (5.18) може да се препише като

$$z = F_\ell(G, K)w. \quad (5.19)$$

По такъв начин  $\mathcal{H}_2$  оптимизацията се състои в минимизирането на  $\mathcal{H}_2$  нормата на  $F_\ell(G, K)$ . Решението на задачата за намиране на  $H_2$  оптимално управляващо устройство в пространство на състоянията включва решаването на две матрични уравнения на Рикати и размерността на управляващото устройство се получава равна на размерността на обобщения обект  $G$ .

При решаването на задачата за  $H_2$  оптимизация обикновено се правят следните допускания

(A1)  $(A, B_1)$  е стабилизируема и  $(A, C_1)$  е детектируема.

(A2)  $(A, B_2)$  е стабилизируема и  $(A, C_2)$  е детектируема.

(A3)  $D_{11} = 0$  и  $D_{22} = 0$ .

Предположенията (A1) и (A2) се изискват за съществуването на стабилизиращи управляващи устройства  $K$ . Предположението (A3) опростява значително изразите, с които се изчислява оптималното управляващо устройство. За опростяване на решението също така понякога се предполага, че  $D_{12}$  и  $D_{21}$  удовлетворяват

(A4)  $D_{12}^T [C_1 \ D_{12}] = [0 \ I]$ .

(A5)  $\begin{bmatrix} B_1 \\ D_{21} \end{bmatrix} D_{21}^T = \begin{bmatrix} 0 \\ I \end{bmatrix}$ .

Това може да се осигури без загуба на общност чрез мащабиране на сигналите  $u$  и  $v$  и унитарни преобразувания на  $w$  и  $z$ .

Горните предположения на пръв поглед са силно ограничаващи, но в действителност повечето добре поставени задачи ги удовлетворяват.

За общата конфигурация на системата за управление, показана на фиг. 5.5 и описана с уравнения (5.17), при предположенията от (A1) до (A5), съществува стабилизиращо управляващо устройство  $K_{opt}(s)$ , което минимизира  $\|F_\ell(G, K_{opt})\|_2$  ако и само ако

(1)  $X_2 \geq 0$  е решение на алгебричното уравнение на Рикати

$$A^T X_2 + X_2 A + C_1^T C_1 - X_2 B_2 B_2^T X_2 = 0 \quad (5.20)$$

такова че  $\operatorname{Re} \lambda_i[A - B_2 B_2^T X_2] < 0, \forall i$ ; и

(2)  $Y_2 \geq 0$  е решение на алгебричното уравнение на Рикати

$$A Y_2 + Y_2 A^T + B_1 B_1^T - Y_2 C_2^T C_2 Y_2 = 0 \quad (5.21)$$

такова че  $\operatorname{Re} \lambda_i[A - Y_2 C_2^T C_2] < 0, \forall i$ .

Единственото оптимално управляващо устройство се дава от

$$K_{opt}(s) \stackrel{\text{def}}{=} \left[ \begin{array}{c|c} \hat{A}_2 & -L_2 \\ \hline F_2 & 0 \end{array} \right], \quad (5.22)$$

където

$$F_2 = -B_2^T X_2, \quad L_2 = -Y_2 C_2^T, \quad (5.23)$$

и

$$\hat{A}_2 = A + B_2 F_2 + L_2 C_2. \quad (5.24)$$

При това

$$\min \|F_\ell(G, K)\|_2 = \|G_c B_1\|_2^2 + \|F_2 G_f\|_2^2 = \|G_c L_2\|_2^2 + \|C_1 G_f\|_2^2$$

където

$$G_c(s) = \left[ \begin{array}{c|c} A_{F_2} & I \\ \hline C_{1F_2} & 0 \end{array} \right], \quad G_f(s) = \left[ \begin{array}{c|c} A_{L_2} & B_{1L_2} \\ \hline I & 0 \end{array} \right]$$

и

$$A_{F_2} = A + B_2 F_2, \quad C_{1F_2} = C_1 + D_{12} F_2,$$

$$A_{L_2} = A + L_2 C_2, \quad B_{1L_2} = B_1 + L_2 D_{21},$$

$$\hat{A}_2 = A - B_2 F_2 + L_2 C_2.$$

Управляващото устройство има същият брой състояния както обобщения обект  $G(s)$ . То може да се раздели на оценител на състоянието (наблюдател) във вида

$$\dot{\hat{x}} = A\hat{x} + B_2u + L_2(C_2\hat{x} - v) \quad (5.25)$$

и обратна връзка по състоянието

$$u = F_2\hat{x}. \quad (5.26)$$

В случая, когато има пълна информация за състоянието, т.е.  $v = x$  и  $C_2 = I$ ,  $D_{21} = 0$ , управляващото устройство се свежда до обратната връзка по състоянието  $u = K_{opt}x$ , където  $K_{opt} = F_2$ . В този случай за предавателната матрица на затворената система се получава

$$T_{zw} = \left[ \begin{array}{c|c} A_{F_2} & B_1 \\ \hline C_{1F_2} & 0 \end{array} \right].$$

**Пример 5.2** Дадена е непрекъснатата система

$$\begin{aligned} \dot{x}(t) &= Ax(t) + B_1w(t) + B_2u(t) \\ z(t) &= C_1x(t) + D_{12}u(t) \end{aligned}$$

където  $x(t) \in \mathbb{R}^5$ ,  $u(t) \in \mathbb{R}^2$ ,  $w(t) \in \mathbb{R}^2$ ,  $z(t) \in \mathbb{R}^2$  и

$$\begin{aligned} A &= \begin{bmatrix} -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 \\ 0.1 & 0 & 0 & 0 & 0 \\ 0 & -0.2 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0.4 & -1 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ -1 & 1 \\ 0 & 1 \\ 1 & -1 \end{bmatrix}, \\ B_2 &= \begin{bmatrix} 10 & 0 \\ 0 & 5 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad C_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad D_{12} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \end{aligned}$$

Оптималният  $\mathcal{H}_2$  регулатор в случая представлява обратна връзка по състоянието, чиято матрица се дава от (до четвъртия знак)

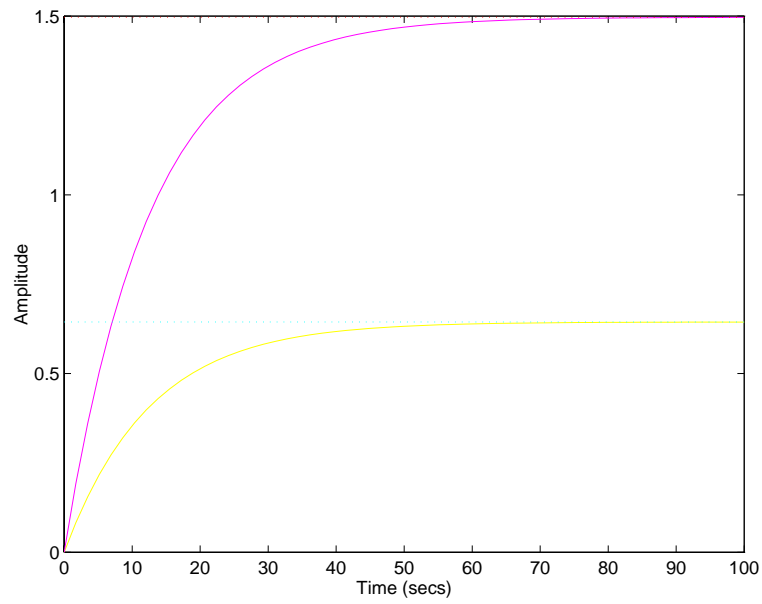
$$K_{opt} = \begin{bmatrix} -0.0135 & 0.0116 & -0.1452 & -0.0046 & -0.0125 \\ 0.0058 & -0.0050 & 0.0625 & 0.0020 & 0.0054 \end{bmatrix}.$$

За минималната стойност на  $\mathcal{H}_2$  нормата на  $T_{zw}$  се получава

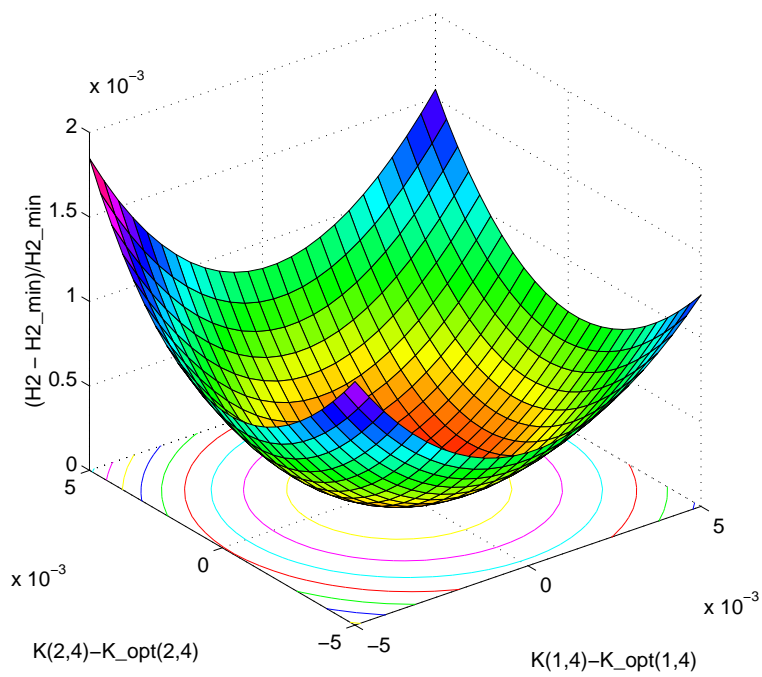
$$\min \|T_{zw}\|_2 = 0.4469.$$

Преходните процеси на затворената система, получени при единично въздействие по първото смущение, са показани на фиг. 5.6.

На фиг 5.7 е показана чувствителността на  $\mathcal{H}_2$  нормата на  $T_{zw}$  към смущения в елементите  $k_{14}$  и  $k_{24}$  на матрицата на обратната връзка  $K_{opt}$ . С  $H_2$  е означена 2-нормата на  $T_{zw}$  за смутената матрица на обратната връзка, а с  $H_{2min}$  – минималната стойност на  $\|T_{zw}\|_2$ .



Фигура 5.6: Преходни процеси при единично въздействие по първото смущение



Фигура 5.7: Чувствителност на  $\mathcal{H}_2$  нормата към смущения в елементите  $k_{14}$  и  $k_{24}$

## 5.4 LQG: специално $\mathcal{H}_2$ управляващо устройство

Важен специален случай на  $\mathcal{H}_2$  оптималното управление е LQG-задачата, разгледана в т. 5.1. За стохастичната система

$$\dot{x} = Ax + Bu + w_d, \quad (5.27)$$

$$y = Cx + w_n \quad (5.28)$$

където

$$\mathcal{E} \left\{ \begin{bmatrix} w_d(t) \\ w_n(t) \end{bmatrix} [w_d(\tau)^T \ w_n(\tau)^T] \right\} = \begin{bmatrix} W & 0 \\ 0 & V \end{bmatrix} \delta(t - \tau), \quad (5.29)$$

LQG-задачата е да се намери  $u = K(s)y$ , такова че

$$J = \mathcal{E} \left\{ \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [x^T Q x + u^T R u] dt \right\} \quad (5.30)$$

се минимизира с  $Q = Q^T \geq 0$  и  $R = R^T$ .

Тази задача може да се формулира като стандартна задача за  $H_2$ -оптимизация по следния начин. Да дефинираме сигнала на грешката  $z$  като

$$z = \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} \quad (5.31)$$

и да представим стохастичните входи  $w_d$ ,  $w_n$  като

$$\begin{bmatrix} w_d \\ w_n \end{bmatrix} = \begin{bmatrix} W^{\frac{1}{2}} & 0 \\ 0 & V^{\frac{1}{2}} \end{bmatrix} w, \quad (5.32)$$

където  $w$  е бял шум с единична интензивност. Тогава квадратичният функционал в LQG задачата е

$$J = E \left\{ \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T z(t)^T z(t) dt \right\} = \|F_\ell(G, K)\|_2^2 \quad (5.33)$$

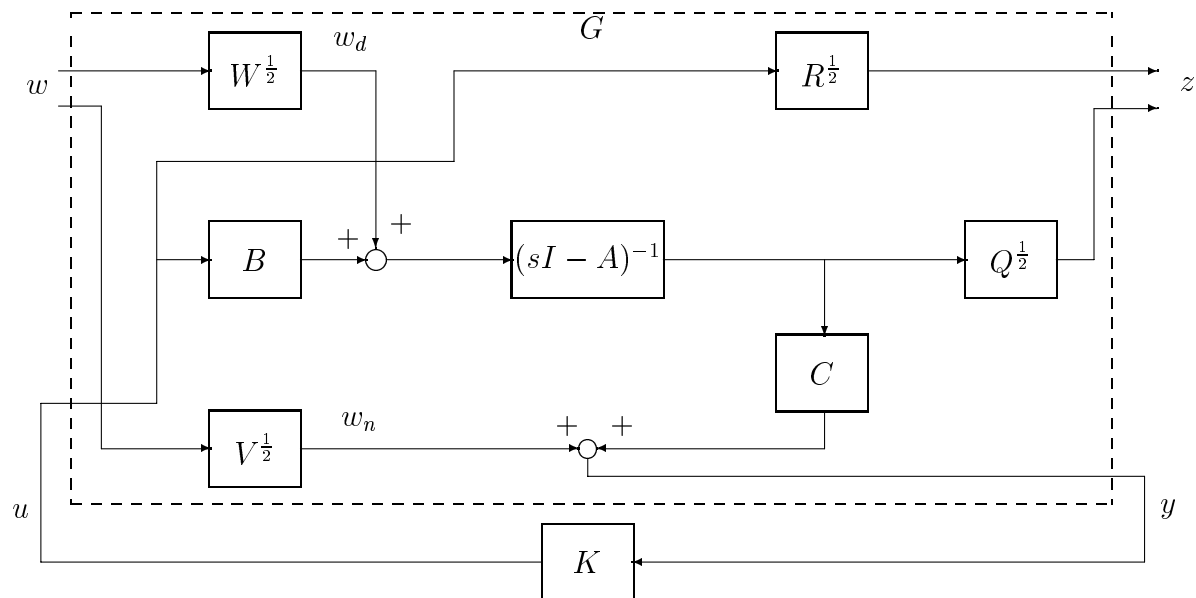
където

$$z(s) = F_\ell(G, K)w(s) \quad (5.34)$$

и обобщеният обект  $G$  се дава от

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{21} \end{bmatrix} = \left[ \begin{array}{c|cc|c} A & W^{\frac{1}{2}} & 0 & B \\ \hline Q^{\frac{1}{2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & R^{\frac{1}{2}} \\ \hline - & - & - & - \\ C & 0 & V^{\frac{1}{2}} & 0 \end{array} \right]. \quad (5.35)$$

Горната формулировка на LQG задачата е илюстрирана в рамките на общата постановка на фиг. 5.8. При стандартните предположения за LQG задачата, приложението на формулите за  $\mathcal{H}_2$ -оптимизация, дадени в предишната точка, дава обичайното LQG оптимално управляващо устройство, както в (5.16).



Фигура 5.8: LQG-управляем объект



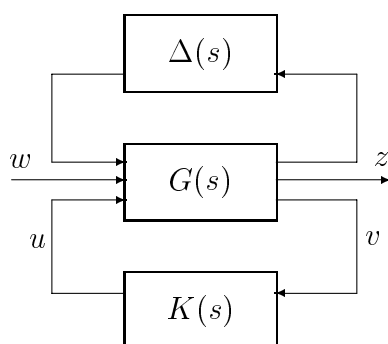


## Глава 6

# $\mathcal{H}_\infty$ ОПТИМАЛНО УПРАВЛЕНИЕ

В тази глава се разглеждат теорията и приложенията на  $\mathcal{H}_\infty$  оптималното управление. В т. 6.1 се формулират стандартните задачи за оптимално и субоптимално  $\mathcal{H}_\infty$  управление на многомерни системи. В т. 6.2 се разглеждат няколко важни задачи за синтез на многомерни системи, които могат да се формулират като задачи за  $\mathcal{H}_\infty$  оптимизация. В т. 6.3 се привеждат решението на стандартната задача за  $\mathcal{H}_\infty$  субоптимално управление. И накрая, в т. 6.4 се разглежда пример за синтез на многомерна система при смесен критерий за качество и робастност.

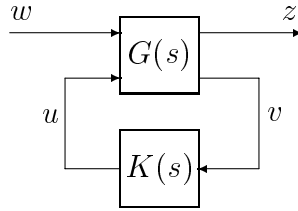
### 6.1 Определение за $\mathcal{H}_\infty$ оптимално управление



Фигура 6.1: Стандартно представяне на неточно известен обект с обратна връзка

В гл. 3 бе показано, че може да се получи точно представяне на неопределеността в модела на обекта във вида, даден на фиг. 6.1. В това представяне "външният вход" ( $w$ ) е вектор от всички сигнали, подадени към системата, а "грешката" ( $z$ ) е вектор от всички сигнали, които са необходими, за да се характеризира поведението на затворената система. И двата вектора могат да

съдържат елементи, които са абстрактни, в смисъл че те могат да се дефинирани математически, но не представят сигнали, които действително съществуват в някаква точка на системата.  $(u)$  е векторът на управляващите сигнали, а  $(v)$  е векторът на измерваните изходи.



Фигура 6.2: Както фиг. 6.1, но с премахнато смущение  $\Delta$

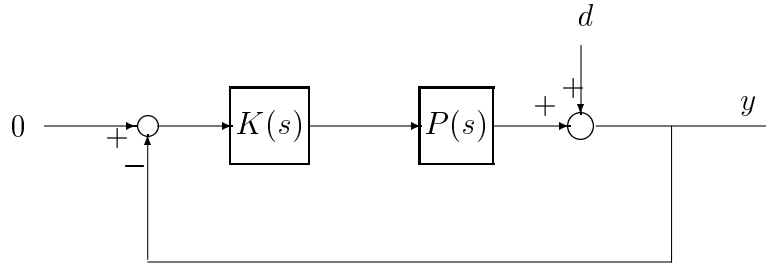
В тази глава ще се използва представянето, показано на фиг. 6.2, което е същото, както на фиг. 6.1, но без да е дадено адитивното смущение  $\Delta(s)$ . Предавателната матрица  $G(s)$  е изведена от номиналния модел на обекта, както в гл. 3. Тя обаче може да съдържа тегловни функции, които зависят от задачата за синтез, която се решава. Предполага се, че предавателните матрици на обекта  $G$  и управляващото устройство  $K$  са реално-рационални и правилни, а техните реализации в пространство на състоянията са стабилизируеми и детектируеми. Управляващото устройство се нарича *допустимо*, ако то осигурява вътрешната устойчивост на системата. Нека при тези допускания с  $T_{zw}$  означим предавателната матрица между външния вход  $w$  и грешката  $z$ . Стандартната задача за  $\mathcal{H}_\infty$  оптимално управление се формулира по следния начин.

**Оптимално  $\mathcal{H}_\infty$  управление:** *да се намерят всички допустими управляващи устройства  $K(s)$ , такива че  $\|T_{zw}\|_\infty$  е минимална.*

Трябва да се подчертае, че така дефинираните  $\mathcal{H}_\infty$  управляващи устройства в общия случай не са единствени за многомерните системи. Освен това, намирането на оптимално  $\mathcal{H}_\infty$  управляващо устройство често пъти е трудно от теоретическа и числена гледна точка. Това очевидно е в контраст със стандартната  $\mathcal{H}_2$  теория, в която оптималното управляващо устройство е единствено и може да се получи чрез решаване на две уравнения на Рикати без итерации. Познаването на достижимата оптимална (минимална)  $\mathcal{H}_\infty$  норма теоретически може да е полезно, тъй като тя задава границата, която можем да постигнем. На практика, обаче, често не е необходимо, а понякога дори и нежелателно, да се синтезира оптимално управляващо устройство и обикновено е много по-лесно да се намерят управляващи устройства, които са много близки в смисъл на нормата до оптималните. Тези управляващи устройства се наричат *субоптимални управляващи устройства*. Субоптималното управляващо устройство може също да има и някои други положителни свойства в сравнение с оптималното, например по-тясна честотна лента.

**Субоптимално  $\mathcal{H}_\infty$  управление:** За дадено  $\gamma > 0$ , да се намерят всички допустими управляващи устройства, ако такива съществуват, така че  $\|T_{zw}\| < \gamma$ .

Нека разгледаме най-напред един относително прост пример за  $\mathcal{H}_\infty$  оптимално управление. Даден е стандартния контур с обратна връзка, показан на фиг. 6.3.



Фигура 6.3: Конфигурация на обратната връзка за примера

За да се опрости максимално постановката на задачата се приема, че всички сигнали са скалярни и че  $P$  е устойчива. Предполагаме, че не е известно какви смущения  $d$  се очакват, но знаем, че техният спектър се съдържа основно в честотния диапазон  $0 \leq \omega \leq \omega_b$ . Задачата за синтез е да се намери управляващо устройство  $K$ , което минимизира най-лошата реакция  $y$ , дължаща се на някакво смущение  $d$  и стабилизира контура. Това може да се постигне чрез решаване на задачата

$$\text{да се минимизира } \|S\|_\infty$$

или еквивалентно

$$\text{да се минимизира } \sup_{\omega} |S(j\omega)|$$

където  $S = (I + PK)^{-1}$  е изходната чувствителност. В действителност тази постановка на задачата може да се подобри чрез ограничаване на минимизацията над честотния диапазон, в който възниква смущението, тъй като е известно, че намаляването на чувствителността извън този диапазон ще доведе до излишно усиление на шумовете и лош запас по устойчивост. Това може да се постигне чрез намиране на (устойчива, минимално фазова) предавателна функция  $W(s)$ , имаща свойствата

$$|W(j\omega)| \approx 1, \quad 0 \leq \omega \leq \omega_b$$

и

$$|W(j\omega)| \ll 1, \quad \omega > \omega_b$$

и след това решаване на задачата

$$\text{да се минимизира } \sup_{\omega} |W(j\omega)S(j\omega)|.$$

Този прост пример разкрива много от свойствата, които възникват в по-общите  $\mathcal{H}_\infty$  задачи. Преди всичко, ако запишем  $J = WS$  за функцията, чиято стойност

се минимизира и  $J^*$  за тази функция, когато се използва оптималното управляващо устройство, то може да се покаже, че  $J^*$  е константа. Това не е вярно в общия случай, но е вярно, че  $|J^*(j\omega)|$  е константа, което означава, че  $J^*$  има безкрайно широка честотна лента. Това означава, че изборът на тегловната функция  $W$  е решаващ при формулирането на задачи, които имат физически смисъл. Свойства, като стръмността на наклона на  $|W(j\omega)|$  ще определят стръмността, с която  $|S(j\omega)|$  ще нараства (над  $\omega_b$ ) тъй като  $|S|$  ще има същата форма, както  $|W^{-1}|$ .

Представяват интерес и някои от свойствата, които са характерни специално за този пример. Първо, оптималното решение в действителност не е реализуемо, тъй като то би изисквало управляващо устройство, чиято предавателна функция не е правилна. Това може да се избегне чрез използване на по-усложнена критериална функция – например, чрез наказване на енергията за управление и чувствителността. Второ, задачата в действителност няма решение, ако обекта няма нули в дясната комплексна полуравнина, тъй като (претеглената) чувствителност тогава може да се направи безкрайно малка, но управляващите въздействия ще са безкрайно големи. Това показва, че в този пример задачата е прекалено опростена, поради което тя става нереалистична.

## 6.2 Задачи за $\mathcal{H}_\infty$ оптимално управление

Стандартната задача за  $\mathcal{H}_\infty$  оптимизация може да се представи по следния начин.

Нека в съответствие с фиг. 6.2 уравненията на системата са записани като

$$\begin{bmatrix} z(s) \\ v(s) \end{bmatrix} = \begin{bmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{bmatrix} \begin{bmatrix} w(s) \\ u(s) \end{bmatrix} \stackrel{\text{def}}{=} G(s) \begin{bmatrix} w(s) \\ u(s) \end{bmatrix}$$

Реализацията в пространство на състоянията на обобщеният обект  $G(s)$  се взема във вида

$$G(s) = \left[ \begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{array} \right].$$

Предавателната матрица от  $w$  към  $z$  се дава от линейното дробно преобразуване

$$z = F_\ell(G, K)w, \quad (6.1)$$

където

$$F_\ell(G, K) = G_{11} + G_{12}K(I - G_{22}K)^{-1}G_{21}. \quad (6.2)$$

Посредством подходящо дефиниране на  $w$  и  $z$  (или еквивалентно на  $G$ ) е възможно редица практически задачи за синтез да се представят във вида

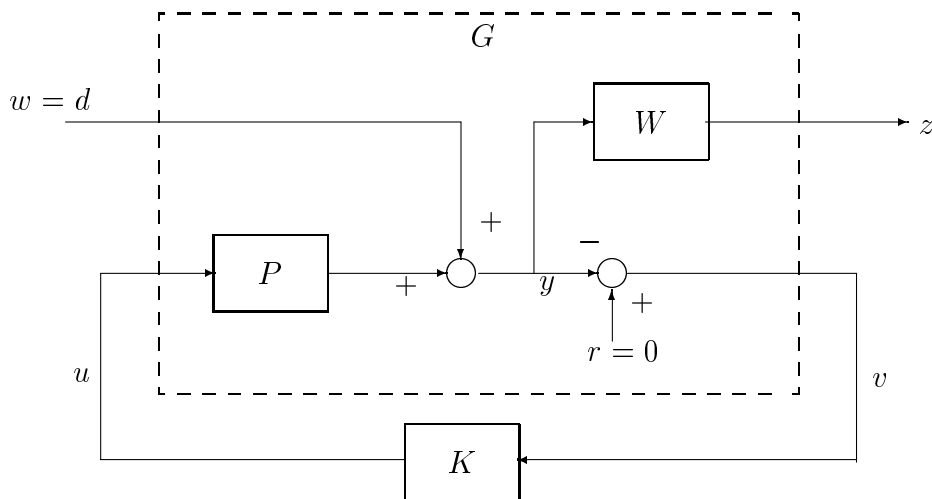
$$\text{да се минимизира } \|F_\ell(G, K)\|_\infty$$

където минимизацията е над множеството на всички допустими управляващи устройства  $K(s)$ .

### Минимизация на чувствителността

Нека най-напред представим примерът за минимизиране на чувствителността, даден в предишната точка, в стандартната форма, показана на фиг. 6.2. За целта смущението  $d$  се разглежда като единствен външен вход, т.е.  $w = d$  и като грешка се дефинира сигналът  $z = Wy$ , така както е показано на фиг. 6.4. Лесно се вижда, че  $y = (I + PK)^{-1}d = Sw$  и  $z = WSw$ , както се изисква. В случая искаме да минимизираме

$$F_\ell(G, K) = WS.$$



Фигура 6.4: Минимизация на чувствителността в стандартна форма

Имаме, че

$$G_{11} = W, \quad G_{12} = WP, \quad G_{21} = -I, \quad G_{22} = -P.$$

Трябва да се отбележи, че тази формулировка на задачата остава в сила, ако желаем да се минимизира  $\|WS\|_\infty$  за многомерен обект.

Ако  $P$  и  $W$  имат реализации в пространство на състоянията

$$P = \left[ \begin{array}{c|c} A_P & B_P \\ \hline C_P & D_P \end{array} \right], \quad W = \left[ \begin{array}{c|c} A_W & B_W \\ \hline C_W & D_W \end{array} \right],$$

то обобщеният обект  $G(s)$  има реализация в пространство на състоянията с

$$A = \left[ \begin{array}{cc} A_P & 0 \\ B_W C_P & A_W \end{array} \right], \quad B_1 = \left[ \begin{array}{c} 0 \\ B_W \end{array} \right], \quad B_2 = \left[ \begin{array}{c} B_P \\ B_W D_P \end{array} \right],$$

$$C_1 = \left[ \begin{array}{cc} D_W C_P & C_W \end{array} \right], \quad D_{11} = D_W, \quad D_{12} = D_W D_P,$$

$$C_2 = \left[ \begin{array}{cc} -C_P & 0 \end{array} \right], \quad D_{21} = -I, \quad D_{22} = -D_P.$$

### Критерий за смесена чувствителност и робастност

В задачите за смесена чувствителност функцията на чувствителността  $S = (I + GK)^{-1}$  се формира заедно с една или повече матрични предавателни функции на затворената система, такива като  $KS$  или допълнителната функция на чувствителността  $T = I - S$ .

Нека най-напред разгледаме задача за регулиране, в която желаем да подтиснем смущението  $d$ , възникващо на изхода на обекта и нека шумът на измерването е сравнително незначителен. Да напомним, че  $S$  е предавателната матрица между  $d$  и изхода, а  $KS$  е предавателната матрица между  $d$  и управляващите сигнали. Включването на  $KS$  е важно за ограничаване на големината на елементите и честотната лента на пропускане на управляващото устройство и следователно на енергията за управление. Големината на  $KS$  е важна също за робастната устойчивост по отношение на неопределеността, моделирана като адитивни смущения на обекта.

Смущението  $d$  обикновено е нискочестотен сигнал и следователно то ще бъде подтиснато успешно, ако максималното сингулярно число на  $S$  се направи малко за същият нискочестотен диапазон. За да се постигне това, може да се избере нискочестотен филтър  $w_1(s)$  с честотна лента на пропускане равна на тази на смущението и да се намери стабилизиращо управляващо устройство, което минимизира  $\|w_1 S\|_\infty$ . Тази критериална функция обаче не е практично да се използва самостоятелно. В присъствието на неминимално фазови нули е много по-полезно да се минимизира

$$\left\| \begin{bmatrix} w_1 S \\ w_2 KS \end{bmatrix} \right\|_\infty,$$

където  $w_2(s)$  е скаларен високочестотен филтър със спрягаща честота равна на тази на желаната честотна лента на пропускане на затворената система.

В общия случай скаларните предавателни функции  $w_1(s)$  и  $w_2(s)$  могат да се заместят с матрици  $W_1(s)$  и  $W_2(s)$ . Това може да е полезно за системи, чийто отделни канали имат много различни ленти на пропускане, като в този случай се препоръчват диагонални тегловни матрици.

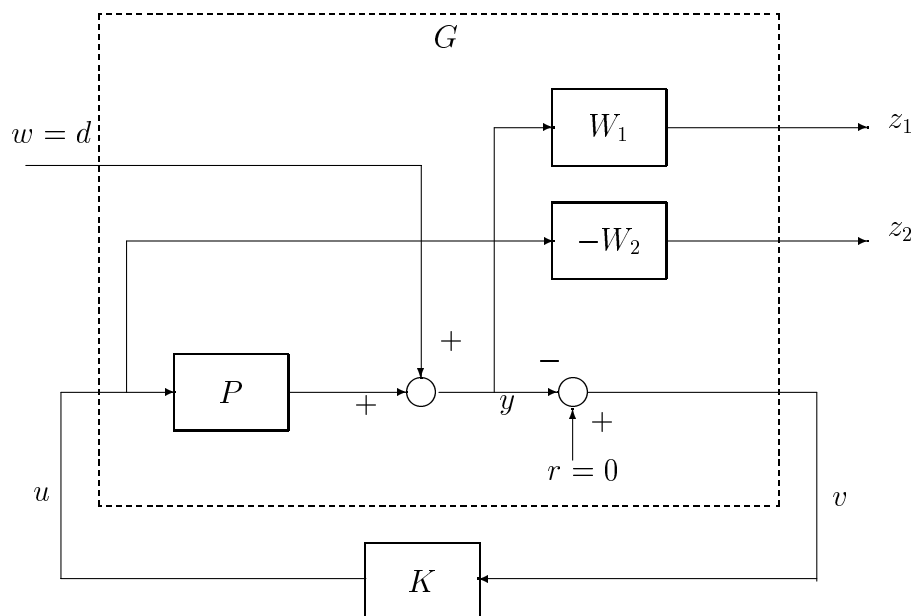
За да се види как задачата за смесена чувствителност може да се формулира в общ вид, смущението  $d$  може да се приеме като единствен външен вход и като сигнал на грешката се дефинира  $z = [z_1^T z_2^T]^T$ , където  $z_1 = W_1 y$  и  $z_2 = -W_2 u$ , така както е показано на фиг. 6.5.

От фигурата се вижда, че  $z_1 = W_1 S w$  и  $z_2 = W_2 K S w$ , както се изискваше. Елементите на обобщения обект  $G$  се дават от

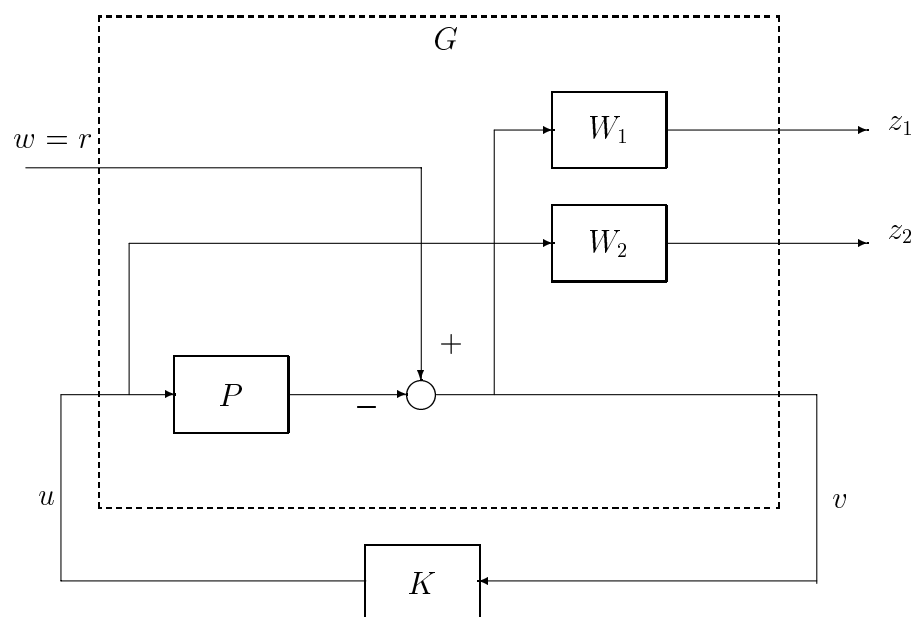
$$G_{11} = \begin{bmatrix} W_1 \\ 0 \end{bmatrix}, \quad G_{12} = \begin{bmatrix} W_1 P \\ -W_2 P \end{bmatrix}, \quad G_{21} = -I, \quad G_{22} = -P,$$

където разделянето на матрицата  $G$  на блокове е такова, че

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ v \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix},$$



Фигура 6.5: Минимизация на смесената  $S/KS$  чувствителност в стандартна форма (регулиране)



Фигура 6.6: Минимизация на смесената  $S/KS$  чувствителност в стандартна форма (слеждане)

и

$$F_\ell(G, K) = \begin{bmatrix} W_1 S \\ W_2 K S \end{bmatrix}.$$

Една друга интерпретация на минимизацията на  $S/KS$  смесената чувствителност е показана на фиг. 6.6, където се разглежда задача за следене. Външният вход е сигналът на заданието  $r$ , а сигналите на грешката са  $z_1 = W_1(r - y)$  и  $z_2 = W_2u$ . Както в задачата за регулиране от фиг. 6.5, в разглежданата задача за следене имаме, че  $z_1 = W_1Sw$  и  $z_2 = W_2KSw$ .

Нека сега да предположим, че е желателно да се получи добро подтискане на смущенията и да се поддържа устойчивостта в присъствието на неструктурирани мултипликативни изходни изменения. Това означава, че желаем да направим чувствителността  $S$  и допълнителната чувствителност  $T$  малки по големина. За съжаление,  $\mathcal{H}_\infty$  нормите на  $S$  и  $T$  не могат да се направят едновременно малки. Ако едната от тях се направи малка, другата става голяма. За да се направи компромис между тези величини се формулира задачата за смесена чувствителност, като задача за намиране на управляващо устройство  $K$ , което стабилизира затворената система и минимизира  $\|F_\ell(G, K)\|_\infty$ , където  $F_\ell(G, K)$  се дава от

$$F_\ell(G, K) = \begin{bmatrix} W_1S \\ W_3(I - S) \end{bmatrix}$$

и  $W_1, W_3$  са честотно зависими тегловни матрици, които се избират от проектанта на системата в съответствие с конкретната ситуация, т.е. те зависят от характера на смущенията и неопределеностите в обекта. Както бе отбелязано по-горе, обикновено смущенията възникват в областта на ниските честоти; следователно  $W_1$  се избира като нискочестотен филтър, за да се даде по-голямо тегло на грешката при ниски честоти. Неопределеността в обекта също е честотно зависима: колкото е по-висока честотата, толкова по-голяма е неопределеността. Ето защо  $W_3$  се избира обикновено като високочестотен филтър.

Задачата за намиране на  $K(s)$ , което стабилизира затворената система и минимизира  $\|F_\ell(G, K)\|_\infty$ , може да се представи като стандартна задача за  $\mathcal{H}_\infty$  оптимизация, така като е показано на фиг. 6.7. В случая имаме че  $w = r$  и  $v = r - y$ .

Може да се покаже, че матрицата  $F_\ell(G, K)$  е точно предавателната матрица от  $w$  към  $[z_1^T z_2^T]^T$  на затворената система. Оттук намираме, че

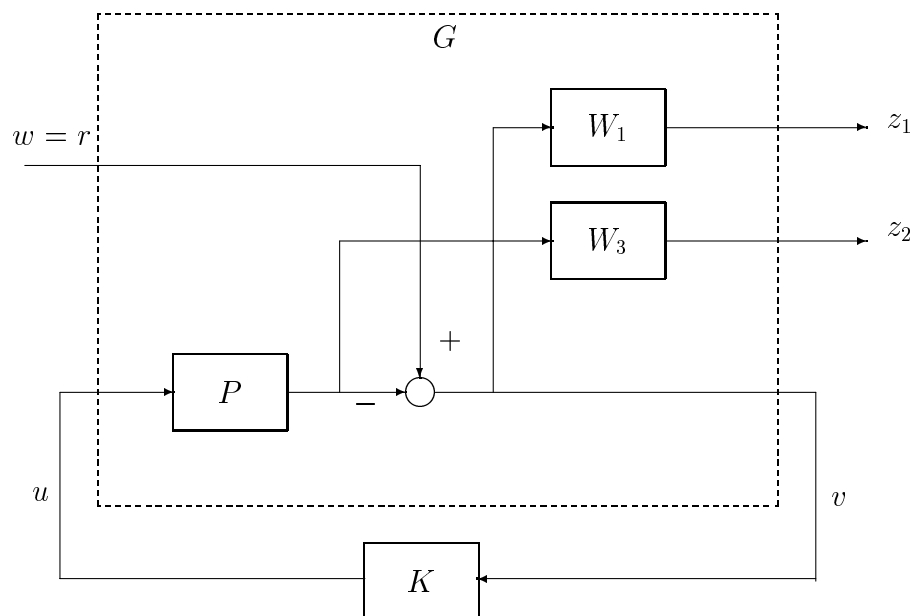
$$G_{11} = \begin{bmatrix} W_1 \\ 0 \end{bmatrix}, \quad G_{12} = \begin{bmatrix} -W_1P \\ W_2P \end{bmatrix}, \quad G_{21} = I, \quad G_{22} = -P.$$

И накрая, може да се минимизира целевата функция

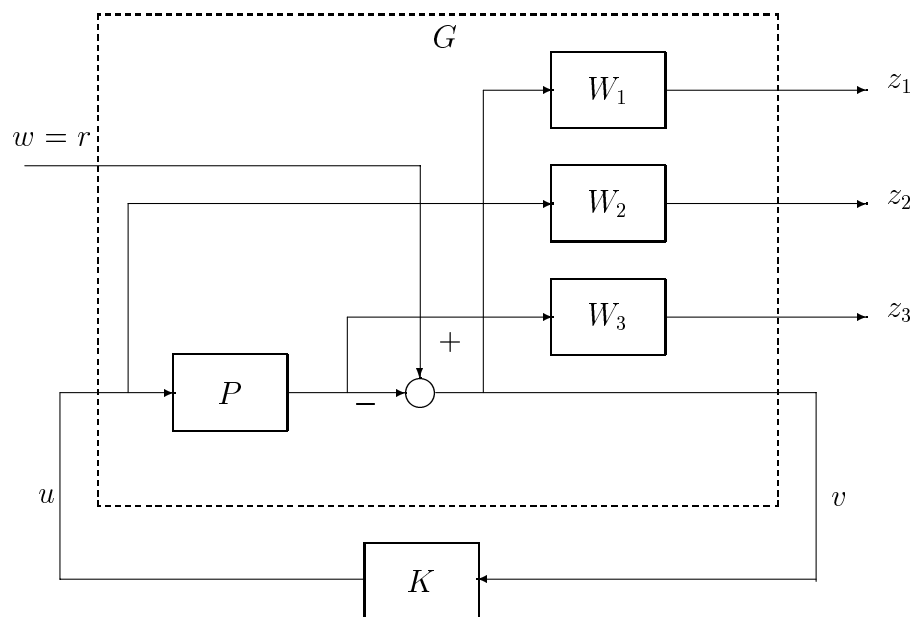
$$\left\| \begin{bmatrix} W_1S \\ W_2KS \\ W_3T \end{bmatrix} \right\|_\infty,$$

при което се постига компромис между подтискане на смущенията, големина на управляващите въздействия и робастност към мултипликативни смущения на изхода. Представянето на тази задача в стандартна форма е показано на





Фигура 6.7: Минимизация на смесената  $S/T$  чувствителност в стандартна форма



Фигура 6.8: Минимизация на смесената  $S/KS/T$  чувствителност в стандартна форма

фиг. 6.8, откъдето за предавателната матрица на обобщения обект се получава

$$G(s) = \left[ \begin{array}{c|c} W_1 & -W_1P \\ 0 & W_2 \\ 0 & W_3P \\ \hline I & -P \end{array} \right] =$$

$$\left[ \begin{array}{cccc|c|c} A_P & 0 & 0 & 0 & 0 & B_P \\ -B_{W_1}C_P & A_{W_1} & 0 & 0 & B_{W_1} & -B_{W_1}D_P \\ 0 & 0 & A_{W_2} & 0 & 0 & B_{W_2} \\ B_{W_3}C_P & 0 & 0 & A_{W_3} & 0 & B_{W_3}D_P \\ \hline -D_{W_1}C_P & C_{W_1} & 0 & 0 & D_{W_1} & -D_{W_1}D_P \\ 0 & 0 & C_{W_2} & 0 & 0 & D_{W_2} \\ D_{W_3}C_P & 0 & 0 & C_{W_3} & 0 & D_{W_3}D_P \\ \hline -C_P & 0 & 0 & 0 & I & -D_P \end{array} \right].$$

Формирането на предавателните матрици на затворената система с разгледаните целеви функции става трудно при повече от две предавателни матрици. При две матрици процедурата за синтез е сравнително лесна. Изискванията върху честотната лента на всяка от тях обикновено са допълващи се и необходимото формиране може да се постигне с прости, устойчиви, нискочестотни и високочестотни филтри. Трябва да се подчертае, че всички тегловни матрици  $W_i$  в  $H_\infty$  оптималното управление трябва да са устойчиви. Само в този случай се удовлетворяват условията за съществуване на управляващо устройство. Ако ние желаем, например, да поставим акцента върху минимизираното на  $S$  в ниските честоти чрез претегляне с член, включващ интегрално действие, ние трябва да апроксимираме  $\frac{1}{s}$  с  $\frac{1}{s+\epsilon}$ , където  $\epsilon \ll 1$ . По подобен начин би могло да се претегли  $KS$  с неправилна предавателна функция, за да се осигури малко  $K$  извън честотната лента на системата. Стандартните предположения обаче изключват използването на такова тегло. Ето защо неправилният член като например  $(1 + \tau_1 s)$  трябва да се замести с  $(1 + \tau_1 s)/(1 + \tau_2 s)$ , където  $\tau_2 \ll \tau_1$ .

### 6.3 Решение на стандартната задача

Решението на задачата за намиране на  $H_\infty$  субоптимално управляващо устройство в пространство на състоянията се получава с алгоритъма на Гловър-Дойл [11]. Както и в случая на  $H_2$ -оптимизация е необходимо да се решат две матрични уравнения на Рикати и размерността на управляващото устройство се получава равна на размерността на обобщения обект  $G$ .

Предполага се, че съгласно (6.1) реализацията в пространство на състоянията на обобщения обект се дава от

$$G(s) = \left[ \begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{array} \right].$$

Алгоритъмът получава семейство от стабилизиращи управляващи устройства  $K$ , такива че

$$\|F_\ell(G, K)\|_\infty < \gamma$$

за някаква стойност на  $\gamma$ , ако такова управляващо устройство съществува.

При решаването на задачата за  $H_\infty$ -оптимизация се правят същите допускания, както и при  $H_2$ -оптимизацията:

(A1)  $(A, B_1)$  е стабилизируема и  $(A, C_1)$  е детектируема.

(A2)  $(A, B_2)$  е стабилизируема и  $(A, C_2)$  е детектируема.

(A3)  $D_{11} = 0$  и  $D_{22} = 0$ .

Предположенията (A1) и (A2) се изискват за съществуването на стабилизиращи управляващи устройства  $K$ . Предположението (A3) опростява значително изразите, с които се изчислява субоптималното управляващо устройство. За опростяване на решението също така понякога се предполага, че  $D_{12}$  и  $D_{21}$  удовлетворяват

$$(A4) \quad D_{12}^T [C_1 \ D_{12}] = [0 \ I].$$

$$(A5) \quad \begin{bmatrix} B_1 \\ D_{21} \end{bmatrix} D_{21}^T = \begin{bmatrix} 0 \\ I \end{bmatrix}.$$

Това може да се осигури без загуба на общност чрез мащабиране на сигналите  $u$  и  $v$  и унитарни преобразувания на  $w$  и  $z$ .

Предположенията (A4) и (A5) са обичайни при  $H_2$  оптимизацията, напр. в LQG задачата, където няма смесени членове в квадратичния критерий ( $D_{12}^T C_1 = 0$ ) и шумът в обекта и шумът на измерването не са корелирани ( $B_1 D_{21}^T = 0$ ).

За общата конфигурация на системата за управление, показана на фиг. 6.2 и описана с уравнения (6.1), при предположенията от (A1) до (A5), съществува стабилизиращо управляващо устройство  $K_{sub}(s)$ , такова че  $\|F_\ell(G, K_{sub})\|_\infty < \gamma$  ако и само ако

(1)  $X_\infty \geq 0$  е решение на алгебричното уравнение на Рикати

$$A^T X_\infty + X_\infty A + C_1^T C_1 + X_\infty (\gamma^{-2} B_1 B_1^T - B_2 B_2^T) X_\infty = 0 \quad (6.3)$$

такова че  $\text{Re } \lambda_i [A + (\gamma^{-2} B_1 B_1^T - B_2 B_2^T) X_\infty] < 0, \forall i$ ; и

(2)  $Y_\infty \geq 0$  е решение на алгебричното уравнение на Рикати

$$A Y_\infty + Y_\infty A^T + B_1 B_1^T + Y_\infty (\gamma^{-2} C_1^T C_1 - C_2^T C_2) Y_\infty = 0 \quad (6.4)$$

такова че  $\text{Re } \lambda_i [A + Y_\infty (\gamma^{-2} C_1^T C_1 - C_2^T C_2)] < 0, \forall i$ ; и

(3)  $\rho(X_\infty Y_\infty) < \gamma^2$ .

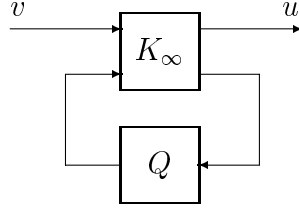
Всички такива управляващи устройства тогава се дават от долното линейно дробно преобразувание  $K_{sub} = F_\ell(K_\infty, Q)$ , където

$$K_\infty(s) \stackrel{\text{def}}{=} \left[ \begin{array}{c|cc} A_\infty & -Z_\infty L_\infty & Z_\infty B_2 \\ \hline F_\infty & 0 & I \\ \hline -C_2 & I & 0 \end{array} \right], \quad (6.5)$$

$$F_\infty = -B_2^T X_\infty, \quad L_\infty = -Y_\infty C_2^T, \quad Z_\infty = (I - \gamma^{-2} Y_\infty X_\infty)^{-1} \quad (6.6)$$

$$A_\infty = A + \gamma^{-2} B_1 B_1^T X_\infty + B_2 F_\infty + Z_\infty L_\infty C_2 \quad (6.7)$$

и  $Q(s)$  е коя да е устойчива правилна предавателна матрица, такава че  $\|Q\|_\infty < \gamma$ . Структурата на тези управляващи устройства е показана на фиг. 6.9.



Фигура 6.9: Структура на  $H_\infty$  субоптимално управляващо устройство

За  $Q(s) = 0$  получаваме

$$K_c(s) = \left[ \begin{array}{c|c} A_\infty & -Z_\infty L_\infty \\ \hline F_\infty & 0 \end{array} \right] = -Z_\infty L_\infty (sI - A_\infty)^{-1} F_\infty. \quad (6.8)$$

Това е така нареченото "централно" управляващо устройство, което има същият брой състояния както обобщения обект  $G(s)$ . Това централно управляващо устройство може да се раздели на оценител на състоянието (наблюдател) във вида

$$\dot{\hat{x}} = A\hat{x} + B_1 \underbrace{\gamma^{-2} B_1^T X_\infty \hat{x}}_{\hat{w}_{\text{най-лошо}}} + B_2 u + Z_\infty L_\infty (C_2 \hat{x} - v) \quad (6.9)$$

и обратна връзка по състоянието

$$u = F_\infty \hat{x}. \quad (6.10)$$

При сравняването на уравнението на наблюдателя (6.9) с уравнението на филтъра на Калман (5.13) се вижда, че той съдържа допълнителния член  $B_1 \hat{w}_{\text{най-лошо}}$ , където

$$\hat{w}_{\text{най-лошо}} = \gamma^{-2} B_1^T X_\infty \hat{x}$$

може да се интерпретира като оценка на най-лошото смущение (външния вход).

Ако се желае управляващо устройство, което постига  $\gamma_{\min}$  с точност до определен допуск, то може да се използва процедура на бисекция върху  $\gamma$ , докато се получи стойност, която е достатъчно точна. Разгледаният по-горе алгоритъм дава тест за всяка стойност на  $\gamma$ , за да се определи дали тя е по-малка или по-голяма от  $\gamma_{\min}$ .

Може да се покаже, че ако  $\gamma$  се увеличи до много високи стойности, то тогава централното управляващо устройство, генерирано от алгоритъма за  $H_\infty$ -оптимално управление, се сходя към LQG-управляващото устройство, което минимизира  $\|F_\ell(G, K)\|_2$ .

$H_\infty$  оптималното управляващо устройство има следните свойства:

1.  $H_\infty$  оптималната целева функция  $\|F_\ell(G, K_c)\|_\infty$  има безкрайно широка честотна лента на пропускане, т.е.  $\bar{\sigma}[F_\ell(G, K_c)] = 1$  за всички  $\omega \in \mathbb{R}$ .

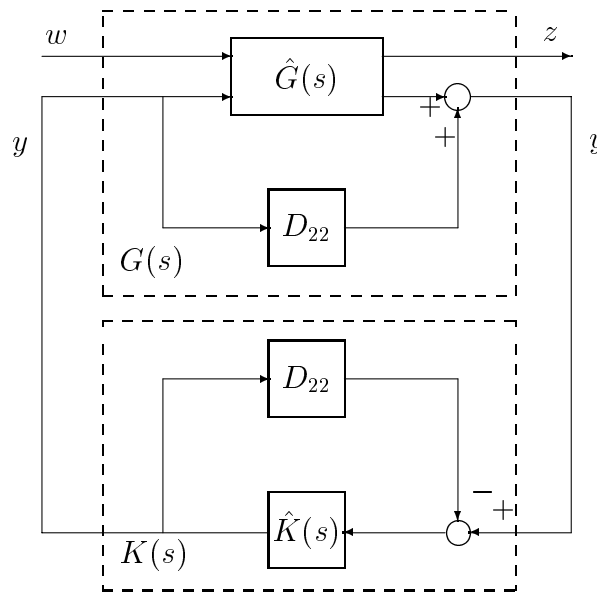
2. В задачите за смесена чувствителност  $H_\infty$  оптималното управляващо устройство винаги съкращава устойчивите полюси на обекта със своите предавателни нули.

В разгледания по-горе алгоритъм се предполага, че блоковете  $D_{11}$  и  $D_{22}$  са нулеви. Ако тези блокове не са нулеви, може да се конструира еквивалентна  $H_\infty$  задача, в която те се нулират. Така например, ако  $D_{22} \neq 0$ , т.е. обектът не е строго правилен, може да се направи следното. Алгоритъмът се прилага към

$$\hat{G} = G - \begin{bmatrix} 0 & 0 \\ 0 & D_{22} \end{bmatrix}$$

и се получава управляващо устройство  $\hat{K}$ . След това се формира управляващото устройство

$$K = \hat{K}(I + D_{22}\hat{K})^{-1}.$$



Фигура 6.10: Управляващо устройство за обект, който не е строго правилен

Фиг. 6.10 показва структурата на  $G$  и  $K$  в термините на  $\hat{G}$  и  $\hat{K}$ . Може да се покаже, че  $K$  стабилизира  $G$ , ако  $\hat{K}$  стабилизира  $\hat{G}$  и че

$$F_\ell(G, K) = F_\ell(\hat{G}, \hat{K}).$$

Единственото усложнение идва от това, че управляващо устройство  $\hat{K}(s)$ , за което  $\det(I + D_{22}\hat{K}(\infty)) = 0$ , не е допустимо.

В случая, когато е достъпна пълната информация за състоянието на обекта, т.е.  $v = x$  и  $C_2 = I, D_{12} = 0$ , субоптималното управляващо устройство се свежда

до обратна връзка по състоянието  $u = K_{sub}x$ , където  $K_{sub} = F_\infty$ . Предавателната матрица на затворената система тогава се дава от

$$T_{zw} = \left[ \begin{array}{c|c} A_{F_\infty} & B_1 \\ \hline C_{1F_\infty} & 0 \end{array} \right],$$

където

$$A_{F_\infty} = A + BF_\infty, \quad C_{1F_\infty} = C_1 + D_{12}F_\infty.$$

Фундаменталната разлика между  $\mathcal{H}_\infty$  и  $\mathcal{H}_2$  управляващите устройства в този случай е, че  $\mathcal{H}_\infty$  регулатора зависи от смущението чрез матрицата  $B_1$ , докато  $\mathcal{H}_2$  регулатора не зависи от него.

**Пример 6.1** Дадена е системата

$$\begin{aligned} \dot{x}(t) &= Ax(t) + B_1w(t) + B_2u(t), \\ z(t) &= C_1x(t) + D_{12}u(t), \end{aligned}$$

където  $x(t) \in \mathbb{R}^5$ ,  $u(t) \in \mathbb{R}^2$ ,  $w(t) \in \mathbb{R}^2$ ,  $z(t) \in \mathbb{R}^2$  и

$$\begin{aligned} A &= \begin{bmatrix} -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 \\ 0.1 & 0 & 0 & 0 & 0 \\ 0 & -0.2 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0.4 & -1 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ -1 & 1 \\ 0 & 1 \\ 1 & -1 \end{bmatrix}, \\ B_2 &= \begin{bmatrix} 10 & 0 \\ 0 & 5 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad C_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad D_{12} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \end{aligned}$$

(Системата е същата, както разгледаната в Пример 5.2.)

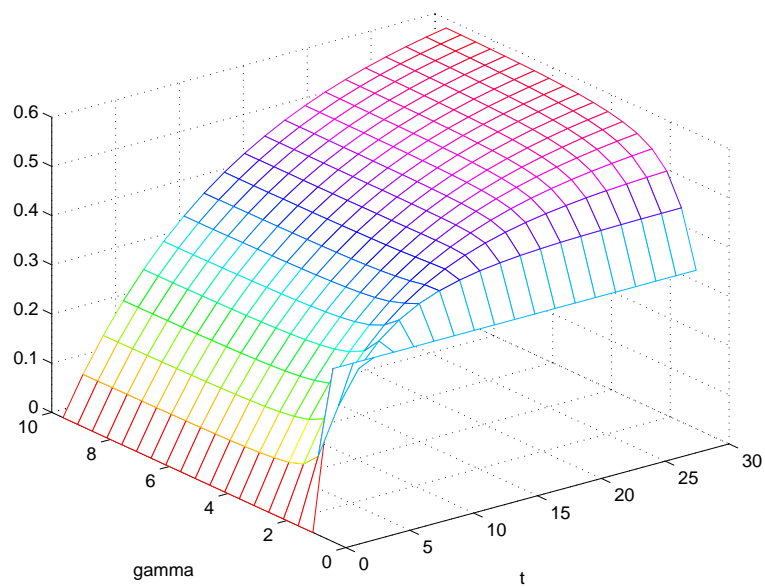
В дадения случай чрез процедура на бисекция за минималната стойност на параметъра  $\gamma$  се намира

$$\gamma_{\min} = 1.1192.$$

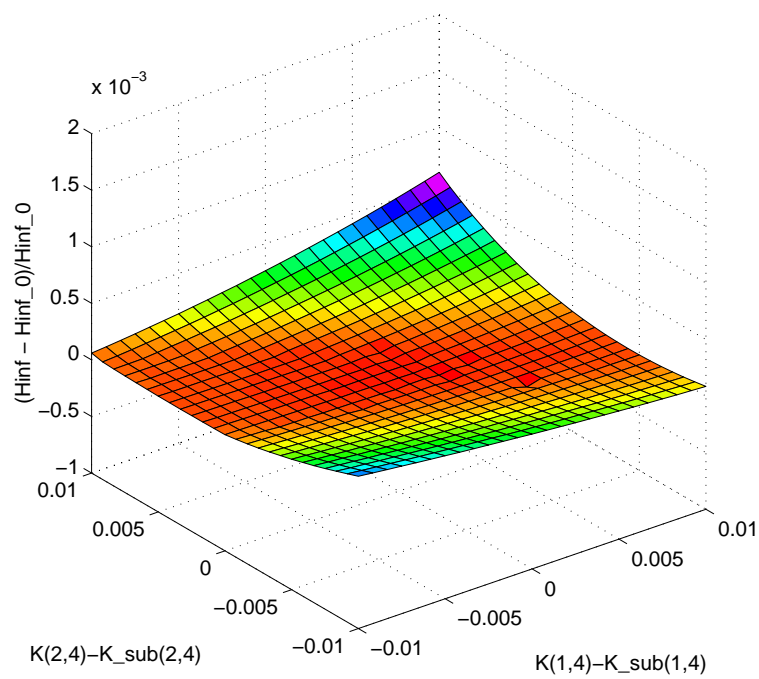
За сравнение свойствата на субоптималните регулатори е синтезирано семейство от такива регулатори за различни стойности на  $\gamma$ . Преходните процеси на съответните затворени системи, получени при единично въздействие по първото смущение, са показани на фиг. 6.11. Вижда се, че с намаляването на  $\gamma$  преходните процеси се ускоряват и установената стойност на грешката, дължаща се на смущението, също намалява. Това обаче е свързано с увеличаване на модулите на елементите в матрицата на обратната връзка.

Чувствителността на  $\mathcal{H}_\infty$  нормата на предавателната матрица на затворената система ще бъде изследвана за  $\gamma = 1.12$ . В този случай матрицата на субоптималния регулатор се дава от (до четвъртия знак)

$$K_{sub} = \begin{bmatrix} -9.9574 & 8.5711 & -107.5121 & -3.4165 & -9.2222 \\ 4.2855 & -3.6889 & 46.2718 & 1.4704 & 3.9691 \end{bmatrix}.$$



Фигура 6.11: Преходни процеси при единично въздействие по първото смущение за различни стойности на  $\gamma$



Фигура 6.12: Чувствителност на  $\mathcal{H}_\infty$  нормата към смущения в елементите  $k_{14}$  и  $k_{24}$

На фиг 6.12 е показана относителната чувствителност на  $\mathcal{H}_\infty$  нормата на  $T_{zw}$  към смущения в елементите  $k_{14}$  и  $k_{24}$  на матрицата на обратната връзка  $K_{sub}$ . С  $Hinf$  е означена  $\mathcal{H}_\infty$ -нормата на  $T_{zw}$  за смутената матрица на обратната връзка, а с  $Hinf_0$  – стойността на  $\|T_{zw}\|_\infty$  за  $\gamma = 1.12$ . Сравнението с фиг. 5.7 показва, че чувствителността на  $\mathcal{H}_\infty$ -нормата на  $T_{zw}$  близо до минимума е много по-ниска, отколкото чувствителността на  $\mathcal{H}_2$  нормата за оптималния  $\mathcal{H}_2$  регулатор.

## 6.4 Пример за синтез при смесена чувствителност

В тази точка се разглежда синтезът на  $H_\infty$  оптимално управляващо устройство за обект от пети ред с три управляващи въздействия и три изхода, чието описание в пространство на състоянията се дава с матриците

$$A = \begin{bmatrix} 0 & 0 & 1.1320 & 0 & -1.0000 \\ 0 & -0.0538 & -0.1712 & 0 & 0.0705 \\ 0 & 0 & 0 & 1.0000 & 0 \\ 0 & 0.0485 & 0 & -0.8556 & -1.0130 \\ 0 & -0.2909 & 0 & 1.0532 & -0.6859 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & 0 & 0 \\ -0.1200 & 1.0000 & 0 \\ 0 & 0 & 0 \\ 4.4190 & 0 & -1.6650 \\ 1.5750 & 0 & -0.0732 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Синтезът се извършва при следните изисквания:

1. Изискване за робастност: Максималното сингулярно число на матрицата на допълнителната чувствителност да лежи под права с наклон -20 dB/dec като при честота 100 rad/s да има стойност не по-голяма от 0 dB.

2. Изискване към качеството: Да се минимизира функцията на чувствителността.

За целта се синтезира управляващо устройство, което минимизира смесената S/T чувствителност (фиг. 6.7). В случая като целева функция се взема

$$\|F_\ell(G, K)\|_\infty = \left\| \begin{bmatrix} \frac{1}{\gamma} W_1 S \\ W_3 T \end{bmatrix} \right\|_\infty,$$

като се търси това  $\gamma$ , за което е изпълнено

$$\|F_\ell(G, K)\|_\infty \leq 1.$$



Това изискване (с точност до множител равен на  $\sqrt{2}$ ) е равносилно на

$$\sigma_i(S(j\omega)) < \gamma |W_1^{-1}(j\omega)|, \quad \sigma_i(T(j\omega)) < |W_3^{-1}(j\omega)|. \quad (6.11)$$

Получаването на минимална стойност на  $\gamma$  води до най-добро подтискане на смущенията, тъй като това отговаря на минимална стойност на  $\|S\|_\infty$ . Тегловната матрица  $W_3$  не се умножава с  $\gamma$ , за да може при малки стойности на  $\omega$  сингулярните числа на  $T$  да се запазят равни на 1, което осигурява изходът на системата в установен режим да е равен на входа (в случая се разглежда задачата за следене).

Тегловните матрици се вземат като

$$W_1 = \begin{bmatrix} \frac{10}{s+0.01} & 0 & 0 \\ 0 & \frac{10}{s+0.01} & 0 \\ 0 & 0 & \frac{10}{s+0.01} \end{bmatrix},$$

$$W_3 = \begin{bmatrix} \frac{s}{100} & 0 & 0 \\ 0 & \frac{s}{100} & 0 \\ 0 & 0 & \frac{s}{100} \end{bmatrix}.$$

Този избор на тегловните матрици съответства на поставените изисквания към синтеза, но не е единственият възможен. Обърнете внимание, че предавателната матрица  $W_3(s)$  е неправилна, т.е. тя има повече нули, отколкото полюси и не може да се реализира в пространство на състоянията. Това обаче може да се извърши за произведението  $W_3(s)G(s)$ , което е правилна предавателна матрица. Друга особеност на тази задача е, че в случая матрицата  $D_{12}$  в стандартната задача не е от пълен ранг. Ето защо вместо задачата за минимална  $S/T$  чувствителност се решава задачата за минимална  $S/KS/T$  чувствителност, като се взема  $W_2 = 10^{-3}$ . Този избор не променя съществено решаваната задача, но позволява да се получи матрица  $D_{12}$  от пълен ранг.

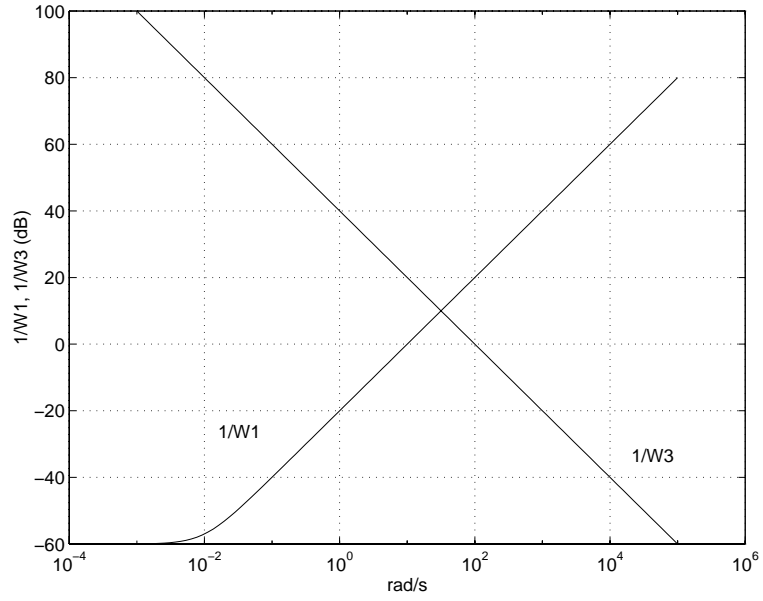
Тегловните характеристики  $W_1^{-1}$  и  $W_3^{-1}$  са показани на фиг. 6.13.

Решението на задачата за  $H_\infty$  оптимизация се получава с програмната система MATLAB посредством функцията `hfopt`, показана в Таблица 6.1. `hfopt` използва функции от Robust Control Toolbox на MATLAB. След девет итерации за оптималната стойност на параметъра  $\gamma$  се получава  $\gamma_{\min} = 0.8477$ . Тъй като обектът е от пети ред, а добавянето на тегловните матрици води до повишаване на реда на модела общо с три, оптималното управляващо устройство се получава от осми ред. Неговата предавателна матрица е

$$K_c(s) = \left[ \begin{array}{c|c} A_K & B_K \\ \hline C_K & D_K \end{array} \right],$$

където (до втория знак)

$$A_K = \begin{bmatrix} -33.87 & 2.60 & -7.59 & -1.06 & 146.58 & 2.53 & -94.23 & -1030.58 \\ 1.51 & -12.58 & 0.61 & -43.04 & -14.26 & -3.16 & -534.16 & 134.16 \\ -3.71 & 0.16 & -26.14 & -0.27 & -29.67 & -172.98 & -9.25 & -379.16 \\ 0.13 & 2.36 & 0.00 & -8.84 & -0.98 & -0.08 & -115.66 & 10.58 \\ -2.06 & 0.13 & -0.73 & -0.09 & -21.89 & -3.69 & 8.72 & 140.88 \\ 2.44 & -0.30 & -3.67 & -0.12 & -3.69 & -28.17 & -2.21 & -68.74 \\ 0.29 & 2.79 & 0.01 & -6.11 & 0.06 & 0.03 & -81.53 & 1.25 \\ 1.13 & -0.23 & 0.07 & 0.37 & 6.32 & -0.63 & 2.07 & -46.42 \end{bmatrix},$$



Фигура 6.13: Тегла върху  $S (W_1^{-1})$  и  $T (W_3^{-1})$

$$B_K = \begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 \\ 0.00 & 0.02 & 0.00 \\ 0.32 & -0.02 & 0.10 \\ 0.05 & -1.98 & -0.22 \\ -0.03 & -7.87 & 0.10 \\ -47.88 & 0.04 & -17.74 \\ 19.56 & 0.04 & -52.82 \end{bmatrix},$$

$$C_K = \begin{bmatrix} -6.88 & 10.29 & -2.74 & 36.43 & 30.34 & -5.31 & 441.04 & -270.37 \\ 6.61 & 0.13 & -20.55 & 2.73 & -68.71 & -151.53 & 52.72 & -32.19 \\ 5.10 & 27.88 & 1.24 & 103.81 & -16.83 & 7.08 & 1316.92 & 38.31 \end{bmatrix},$$

$$D_K = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Таблица 6.1: MATLAB функция за  $H_\infty$  оптимален синтез при смесена чувствителност

---

```

% Използва се Robust Control Toolbox
function [Ak,Bk,Ck,Dk,At,Bt,Ct,Dt, gamopt] = ...
    hfopt(A,B,C,D,nuw1,dnw1,nuw2,dnw2,nuw3,dnw3)
%
% Hinf синтез при смесена чувствителност
%
% ВХОДНИ ДАННИ:
% A,B,C,D:      Описание в пространство на състоянията на обекта.
% nuw1,dnw1:     Полиноми в числителя и знаменателя на тегловната
%                предавателна функция на изходната чувствителност.
% nuw2,dnw2:     Полиноми в числителя и знаменателя на тегловната
%                предавателна функция на входната чувствителност.
% nuw3,dnw3:     Полиноми в числителя и знаменателя на тегловната
%                предавателна функция на допълнителната чувствителност.
%
% ИЗХОДНИ ДАННИ:
% Ak,Bk,Ck,Dk:   Реализация в пространство на състоянията на полученото
%                оптимално управляващо устройство.
% At,Bt,Ct,Dt:   Реализация в пространство на състоянията на
%                предавателната матрица на затворената система по
%                заданието.
% gamopt         Оптимална стойност на параметъра gamma.
%
%
% Разширяване на системата с предавателните матрици на тегловните
% филтри
%
%                W1 = [nuw1;dnw1;...;nuw1;dnw1]
%                W2 = [nuw2;dnw2;...;nuw2;dnw2]
%                W3 = [nuw3;dnw3;...;nuw3;dnw3]
%
ss_p = mksys(A,B,C,D);
[rd, cd] = size(D);
for i = 1:rd;
    j = 2*i - 1;
    W1(j,:) = nuw1;
    W1(j+1,:) = dnw1;
    W3(j,:) = nuw3;
    W3(j+1,:) = dnw3;
end
for i = 1:cd;

```

```

j = 2*i - 1;
W2(j,:) = nuw2;
W2(j+1,:) = dnw2;
end
[TSS_] = augtf(ss_p,W1,W2,W3);
%
% Hinf оптимизация чрез функцията HINFOPT.M
%
[gamma,ss_cp,ss_cl] = hinfopt(TSS_,[1:rd]);
gamopt = 1/gamma;
%
% Формиране на матриците на Hinf управляващото устройство
%
[Ak,Bk,Ck,Dk] = branch(ss_cp);
%
% Изчисляване на предавателната матрица на затворената система
%
[A1,B1,C1,D1] = series(Ak,Bk,Ck,Dk,A,B,C,D);
[At,Bt,Ct,Dt] = feedbk(A1,B1,C1,D1,2);

```

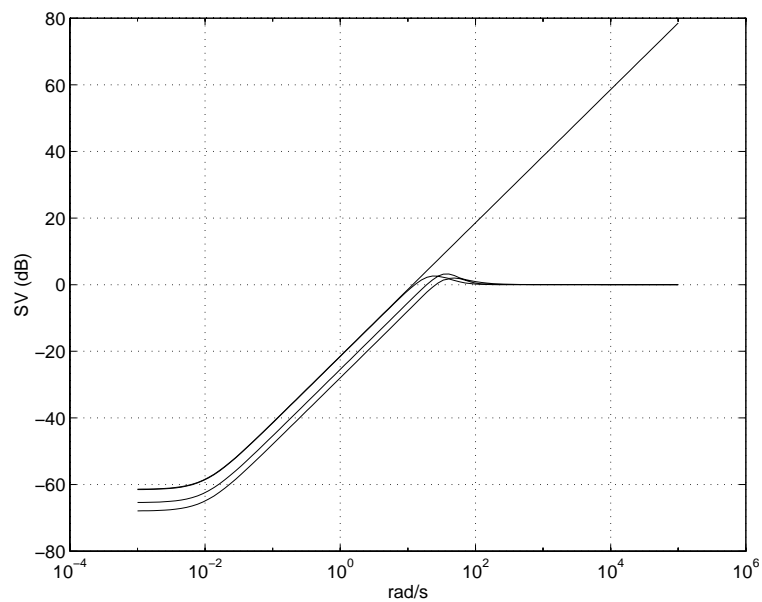
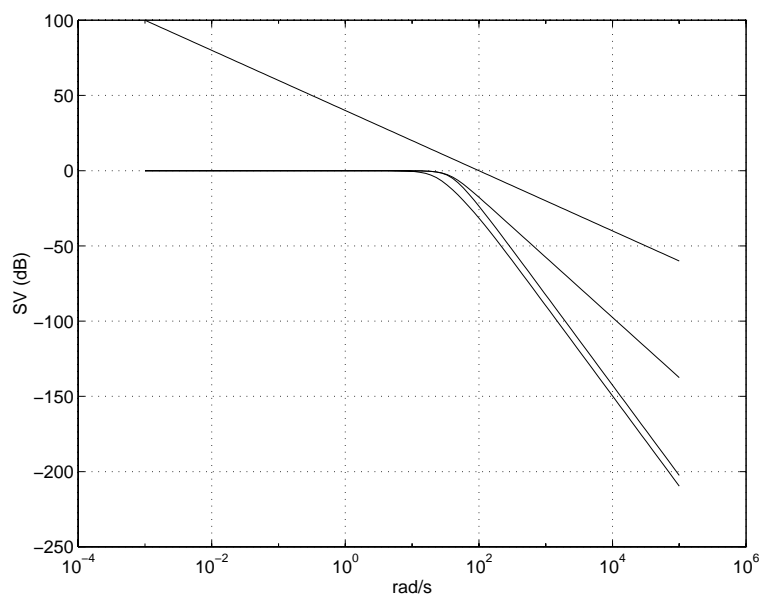
---

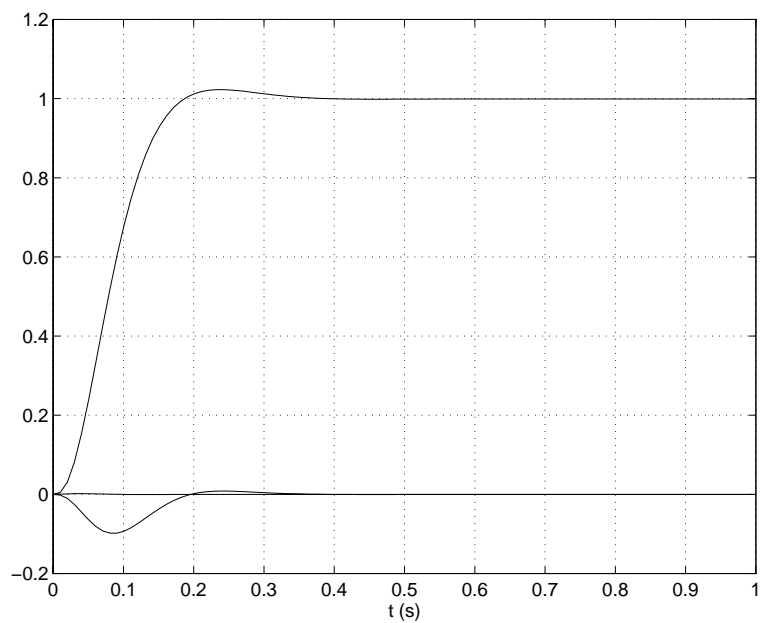
На фиг. 6.14 са показани сингулярните числа на матрицата на чувствителността за полученото оптимално управляващо устройство и тегловната характеристика  $\gamma_{\min} W_1^{-1}$ , а на фиг. 6.15 са показани сингулярните числа на матрицата на допълнителната чувствителност и тегловната характеристика  $W_3^{-1}$ . Вижда се, че изискванията към синтеза (6.12), формулирани чрез тегловните матрици  $W_1$  и  $W_3$ , се спазват както по отношение на  $S$ , така и по отношение на  $T$ .

На фиг. 6.16 са показани преходните процеси на затворената система при единично въздействие по първото задание  $r_1$ , а на фиг. 6.17 са показани преходните процеси на затворената система при единично въздействие по първото смущение  $d_1$ .

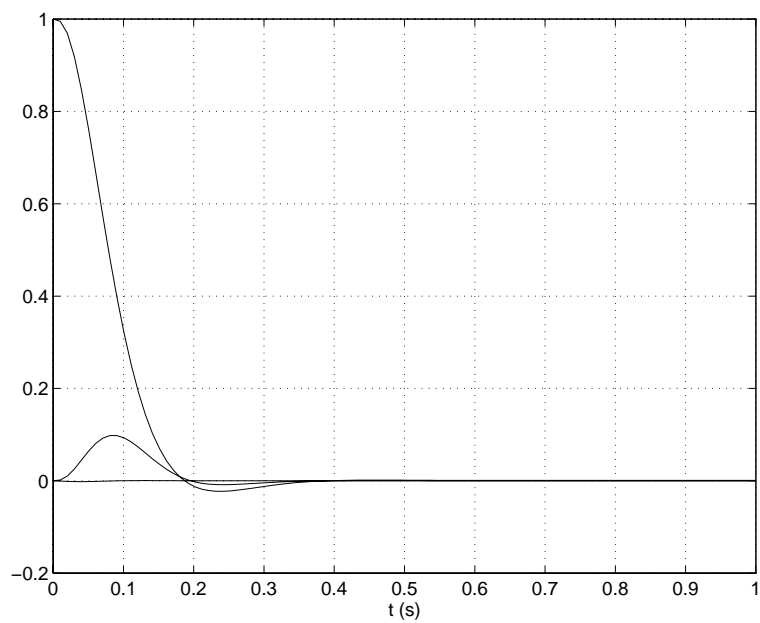
Може да се констатира, че преходните процеси се характеризират с добро бързодействие и отсъствие на големи пререгулирания. Компромис между изискванията към робастността и качеството и големината на управляващите въздействия може да се постигне с подходящ избор на тегловната матрица  $W_2$ . Важен момент при избора на тегловните матрици  $W_1$  и  $W_3$  е да се спазва изискването спрягащата честота на логаритмичната честотна характеристика на  $W_1$  да е достатъчно по-ниска от спрягащата честота на  $W_3^{-1}$ ; по-точно изисква се спазването на условието

$$\bar{\sigma}(W_1^{-1}(j\omega)) + \bar{\sigma}(W_3^{-1}(j\omega)) > 1 \text{ за всяко } \omega.$$

Фигура 6.14: Сингулярни числа на  $S$ Фигура 6.15: Сингулярни числа на  $T$



Фигура 6.16: Преходни процеси при единично въздействие по първото задание



Фигура 6.17: Преходни процеси при единично въздействие по първото смущение

## Глава 7

# ЧИСЛЕНИ МЕТОДИ

В тази глава се разглеждат числените методи, които се използват при анализа и синтеза на многомерните системи за управление. Приложението на тези методи е свързано с оценяване на обусловеността на съответната задача и с анализ на грешките от закръгляване в използвания алгоритъм. Ето защо изложението започва с представянето на числата с плаваща запетая в компютъра и анализа на грешките от закръгляване при елементарните аритметични операции. Разглеждани са най-разпространените методи за решаване на основните матрични задачи и особеностите на тяхното използване в аритметика с плаваща запетая. От голямото разнообразие на методи, намиращи приложение при анализа и синтеза на многомерни системи за управление, са изложени методи за решаване на матричните уравнения на Ляпунов, Силвестър и Рикати, преобразуване в ортогонална канонична форма и синтез при зададена собствена структура.

Подробно изложение на съвременните числени методи за решаване на линейни алгебрични задачи може да се намери в [4, 9, 15, 18]. Числени методи за анализ и синтез на линейни системи за управление са разгледани в [21].

### 7.1 Грешки от закръгляване

#### Числа с плаваща запетая

Когато се извършват в аритметика с крайна точност, каквато е аритметиката на съвременните компютри, изчислителните операции могат да се съпровождат с големи *грешки от закръгляване*. Това се дължи на факта, че реалната числова система се апроксимира в компютрите с *числова система с плаваща запетая*, която съдържа крайно, а не безкрайно множество от реални числа. Системата с плаваща запетая се характеризира с *основа  $\beta$* , *точност  $t$*  и *обхват на експонентата  $[e_{\min}, e_{\max}]$* . Тук  $\beta$  и  $t$  са положителни цели числа,  $e_{\min}$  е отрицателно цяло, а  $e_{\max}$  – положително цяло число. В тази система всяко  $t$ -значно при основа  $\beta$  число с плаваща запетая се представя в *нормализирания вид*

$$\pm 0.d_1 d_2 \dots d_t \times \beta^e \quad (7.1)$$

където целите числа при основа  $\beta$   $d_1, \dots, d_t$  (значещите цифри) удовлетворяват неравенствата

$$1 \leq d_1 \leq \beta - 1$$

$$0 \leq d_i \leq \beta - 1 \quad i = 2, 3, \dots, t$$

и  $e_{\min} \leq e_{\max}$ . Цялото число  $e$  се нарича *експонента*, а числото

$$m = \frac{d_1}{\beta} + \frac{d_2}{\beta} + \dots + \frac{d_t}{\beta} \quad \frac{1}{\beta} \leq m \leq 1$$

– *дробна част (мантиса)*.

Най-често използваните числови основи са 2, 8, 10 и 16. Параметрите  $t, e_{\min}$  и  $e_{\max}$  могат да се различават значително за различните машини.

Числото с плаваща запетая, представено със (7.1), има стойността

$$x = \pm m \cdot \beta^e.$$

Число с плаваща запетая обикновено заема една или две 32-битови думи в паметта на компютъра. Число, което заема една дума се нарича *число с единична точност*, а това което заема две думи – *число с двойна точност*. По-нататък се използва термина *работна точност* за точността, която се използва фактически при конкретни изчисления.

### Стандарт на IEEE

Тъй като подробностите на системата с плаваща запетая варират силно, полезно е да се разгледа стандартът на IEEE, който е получил широко разпространение.

**Пример 7.1 (Стандарт на IEEE за плаваща запетая).** Число с единична точност в стандарта на IEEE има следния формат.

0	1	9	10	31
$\sigma$	exp	frac		

Малките цифри над правоъгълника означават позициите на битовете в 32-битовата дума, съдържаща числото. Правоъгълникът, означен със  $\sigma$  съдържа знака на числото. Останалите два правоъгълника съдържат експонентата и мантисата. Стойността на числото е

$$(-1)^\sigma 1.\text{frac} \cdot 2^{\text{exp} - 127}, \quad 1 \leq \text{exp} \leq 254.$$



Форматът на числата с двойна точност е

0	1	12	13	63
$\sigma$	exp	frac		

Стойността на числото с двойна точност е

$$(-1)^\sigma 1.\text{frac} \cdot 2^{\text{exp} - 1023}, \quad 1 \leq \text{exp} \leq 2046.$$

Отбележете, че величините frac и exp не са същите, както величините m и e, дефинирани по-горе. Тези разлики се състоят в следното.

- Тъй като водещият бит в мантисата на нормализирано двоично число с плаваща запетая е винаги единица, не е ефективно да се отделя един бит за неговото представяне. Ето защо се запамятава само частта от мантисата, която се намира след водещия бит, като мантисата се възстановява чрез формулата  $m = (-1)^\sigma \cdot 1.\text{frac}$ .
- Числото exp се нарича *изместена експонента*, тъй като истинската стойност e на експонентата се изчислява чрез изваждане на изместване. Неизместеният обхват на експонентата е  $[-126, 127]$ , което представлява грубо обхвата на числата от  $10^{-38}$  до  $10^{38}$ . Двойната точност обхваща грубо числата от  $10^{-307}$  до  $10^{307}$ . И в двете точности граничните експоненти (т.е. -127 и 128) са запазени за специални цели.
- Нулата се представя с exp = 0 (една от резервираните експоненти) и frac = 0. Знаковият бит може да е или 0 или 1, така че системата има и двете +0 и -0.

## Закръгляване

*Закръглената стойност* на  $x$  се дефинира като числото с плаваща запетая  $\tilde{x}$ , което е най-близо до  $x$  при условието, че ако  $x$  лежи точно между две числа с плаваща запетая, то  $\tilde{x}$  ще бъде това с по-големия модул. Понякога вместо правилно закръгляване се използва *отсичане*, при което се взема най-близкото число с плаваща запетая, чийто модул е по-малък от  $|x|$ .

За да се анализира влиянието на грешките от закръгляване върху изчисленията е необходимо да се получи оценка за големината на тази грешка. Може да се покаже, че абсолютната грешка при правилното закръгляване удовлетворява

$$|\tilde{x} - x| \leq \frac{1}{2} \beta^{e-t}, \quad (7.2)$$

където дясната страна е равна на максимално възможната стойност на отхвърлените разряди. В случая на отсичане тази стойност е два пъти по-голяма.

Използвайки (7.2) и вземайки предвид, че  $\beta^{e-1} \leq |x|$ , за относителната грешка на закръглената стойност на  $x$  се получава (предполага се, че  $x \neq 0$ )

$$\frac{\tilde{x} - x}{x} \leq \frac{\beta^{e-t}}{2\beta^{e-1}} = \frac{1}{2}\beta^{1-t}. \quad (7.3)$$

Неравенството (7.3) може да се запише като

$$\tilde{x} = x(1 + \delta), \quad |\delta| \leq \frac{1}{2}\beta^{1-t}, \quad (7.4)$$

където  $\delta$  е относителната грешка от закръгляване. Много съществен е фактът, че тази грешка зависи само от параметрите на аритметиката с плаваща запетая и не зависи от големината на числото, което се апроксимира. Ето защо при анализа на грешките от закръгляване обикновено се разглеждат относителните, а не абсолютните грешки.

Нека с  $\text{fl}(x)$  означим закръглената стойност на число  $x$  в  $t$ -значна аритметика с плаваща запетая при основа  $\beta$ . Тогава от (7.4) следва, че

$$\text{fl}(x) = x(1 + \delta), \quad |\delta| \leq \varepsilon_M, \quad (7.5)$$

където числото  $\varepsilon_M$  се нарича *относителна машинна точност на използвания компютър*.

**Пример 7.2 (Стандарт на IEEE).** Относителната машинна точност за единичната точност на стандарта на IEEE е около  $10^{-7}$ . Относителната точност за двойната точност е около  $10^{-16}$ .

Нека  $\circ$  означава коя да е от аритметичните операции  $+$ ,  $-$ ,  $\times$ ,  $\div$  и нека  $\text{fl}(x \circ y)$  означава резултатът от извършване на операцията в аритметика с плаваща запетая с относителна машинна точност  $\varepsilon_M$ . При анализа на грешките от закръгляване се предполага, че съответната операция е извършена точно, след което полученият резултат е закръглен правилно. В резултат получаваме

$$\text{fl}(x \circ y) = (x \circ y)(1 + \delta), \quad \delta \leq \varepsilon_M. \quad (7.6)$$

Границата (7.6) се нарича *стандартна граница на грешките* в аритметиката с плаваща запетая. За някои компютри тази граница не е валидна, но уравнението (7.6) все още може да се използва за някакво по-голямо  $\varepsilon_M$ , при условие, че в процеса на изчисленията се използва т.н. *спомогателна цифра*.

Уравнението (7.6) може да се използва само когато

$$r \leq |x \circ y| \leq R,$$

където  $r$  и  $R$  са съответно най-малкото и най-голямото положителни числа, представими в аритметиката с плаваща запетая. Ако  $|a \circ b| > R$ , то възниква *препълване* и изчисленията се прекратяват, тъй като не съществува приемлива апроксимация на резултата. Ако  $|a \circ b| < r$  възниква *загуба на значимост* и при

повечето компютри резултатът се полага равен на нула. Трябва да се отбележи, че за разлика от препълването, загубата на значимост обикновено не води до разрушаване на изчислителния процес.

Уравнението (7.6) показва, че елементарните аритметични операции се извършват с малки относителни грешки.

*Анализът на грешките от закръгляване* на числените алгоритми се извършва използвайки изразите за грешките, възникващи при елементарните аритметични операции. При този анализ ще предполагаме, че при изпълнението на алгоритъма не възникват препълвания или загуби на значимост.

## 7.2 Обусловеност и числена устойчивост

### Обратен анализ на грешките

Един от най-простите примери, с които може да се илюстрират особеностите на анализа на грешките от закръгляване, е изчисляването на сумата на  $n$  числа

$$s_n = x_1 + x_2 + \cdots + x_n.$$

Подробностите на анализа зависят от реда, по който числата са сумирани. За определеност ще предполагаме, че сумирането е извършено по следния начин:

$$\begin{aligned} s_1 &= \text{fl}(x_1), \\ s_k &= \text{fl}(x_{k-1} + x_k), \quad k = 2, \dots, n. \end{aligned}$$

Естественят подход е да се запише, например,  $\tilde{s}_k$  за изчислените стойности на  $s_k$  и да се направи опит за изчисляване на горна граница върху  $|\tilde{s}_k - s_k|$  от граница върху  $|\tilde{s}_{k-1} - s_{k-1}|$ . Тази техника се нарича *прав анализ на грешките от закръгляване*. Въпреки че такава процедура е полезна в някои случаи, обикновено се използва друг, непряк метод, който се нарича *обратен анализ на грешките*. Идеята при този метод е грешките от закръгляване на всяка стъпка да се отнесат към първоначалните данни, като изчисленият резултат се представи като точен за променени данни. За целта се процедира на етапи.

• **Приложение на стандартната граница.** Това е най-трудната част на анализа. От стандартната граница имаме, че

$$s_2 = \text{fl}(x_1 + x_2) = (x_1 + x_2)(1 + \varepsilon_1) = x_1(1 + \varepsilon_1) + x_2(1 + \varepsilon_1),$$

където  $|\varepsilon_1| \leq \varepsilon_M$ . По подобен начин

$$\begin{aligned} s_3 &= \text{fl}(s_2 + x_3) = (s_2 + x_3)(1 + \varepsilon_2) \\ &= x_1(1 + \varepsilon_1)(1 + \varepsilon_2) + \\ &\quad x_2(1 + \varepsilon_1)(1 + \varepsilon_2) + \\ &\quad x_3(1 + \varepsilon_2). \end{aligned}$$

Продължавайки по този начин, намираме че

$$\begin{aligned}
 s_n &= \Pi(s_{n-1} + x_n) = (s_{n-1} + x_n)(1 + \varepsilon_{n-1}) \\
 &= x_1(1 + \varepsilon_1)(1 + \varepsilon_2) \cdots (1 + \varepsilon_{n-1}) + \\
 &\quad x_2(1 + \varepsilon_1)(1 + \varepsilon_2) \cdots (1 + \varepsilon_{n-1}) + \\
 &\quad x_3(1 + \varepsilon_2) \cdots (1 + \varepsilon_{n-1}) + \\
 &\quad \cdot \quad \cdot \quad \cdot \\
 &\quad x_{n-1}(1 + \varepsilon_{n-2})(1 + \varepsilon_{n-1}) + \\
 &\quad x_n(1 + \varepsilon_{n-1}),
 \end{aligned} \tag{7.7}$$

където  $|\varepsilon_k| \leq \varepsilon_M$  ( $k = 1, 2, \dots, n-1$ ).

Изразите (7.7) не са много информативни, поради което има смисъл да се въведат някои опростяващи означения. Нека величините  $\eta_k$  се дефинират от

$$\begin{aligned}
 1 + \eta_1 &= (1 + \varepsilon_1)(1 + \varepsilon_2) \cdots (1 + \varepsilon_{n-1}), \\
 1 + \eta_2 &= (1 + \varepsilon_1)(1 + \varepsilon_2) \cdots (1 + \varepsilon_{n-1}), \\
 1 + \eta_3 &= (1 + \varepsilon_2) \cdots (1 + \varepsilon_{n-1}), \\
 &\quad \cdot \quad \cdot \quad \cdot \\
 1 + \eta_{n-1} &= (1 + \varepsilon_{n-2})(1 + \varepsilon_{n-1}), \\
 1 + \eta_n &= (1 + \varepsilon_{n-1}).
 \end{aligned} \tag{7.8}$$

Тогава

$$s_n = x_1(1 + \eta_1) + x_2(1 + \eta_2) + x_3(1 + \eta_3) + \cdots + x_{n-1}(1 + \eta_{n-1}) + x_n(1 + \eta_n). \tag{7.9}$$

• **Граници от първи порядък.** Числата  $\eta_k$  в (7.9) се наричат (относителна) *обратна грешка* и е важно да имаме граници за тяхната големина. Тъй като членовете от втори и по-висок порядък са несъществени в сравнение с членовете от първи порядък, те могат да се пренебрегнат. В резултат получаваме

$$\begin{aligned}
 |\eta_1| &\lesssim (n-1)\varepsilon_M, \\
 |\eta_k| &\lesssim (n-k+1)\varepsilon_M, \quad k = 2, 3, \dots, n.
 \end{aligned} \tag{7.10}$$

• **Точни граници.** Ако предположим, че  $n\varepsilon \leq 0.1$ , то границите (7.10) могат да се запишат строго във вида

$$\begin{aligned}
 |\eta_1| &\lesssim (n-1)\varepsilon'_M, \\
 |\eta_k| &\lesssim (n-k+1)\varepsilon'_M, \quad k = 2, 3, \dots, n
 \end{aligned} \tag{7.11}$$

където

$$\varepsilon'_M \stackrel{\text{def}}{=} \frac{\varepsilon_M}{0.9} < 1.12\varepsilon_M.$$

### Обратна устойчивост

Изразът

$$s_n = x_1(1 + \eta_1) + x_2(1 + \eta_2) + x_3(1 + \eta_3) + \cdots + x_{n-1}(1 + \eta_{n-1}) + x_n(1 + \eta_n).$$

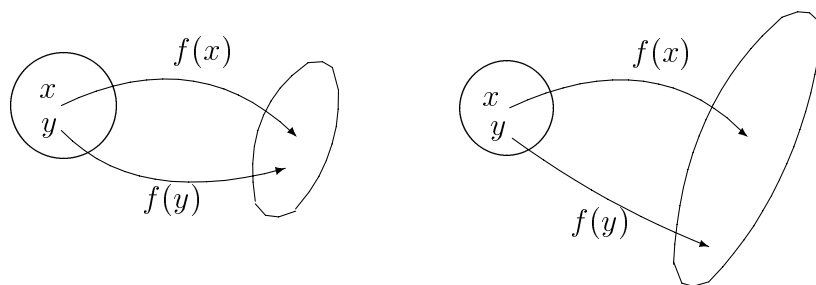
има следната интерпретация.

Когато сумата на  $n$  числа се изчислява в аритметика с плаваща запетая, резултатът е същият, както ако сме изчислили точната сума на същите  $n$  числа, променени с малки относителни грешки. Грешките са ограничени от  $n - 1$  пъти относителната машинна точност.

Ключовият момент тук е, че грешките са отнесени обратно към първоначалните данни. Алгоритми, за които това е изпълнено, се наричат *обратно устойчиви*, или в контекста на матричните изчисления, просто *устойчиви*.

Обратната устойчивост дава възможност да се погледне по различен начин на качеството на изчислените решения. На практика, данните на които се основава изчислението не са точни, а са "замърсени" с грешки. Тези грешки обикновено са много по-големи от относителната машинна точност, така че техният ефект върху изчисленият резултат е много по-голям, отколкото ефекта на грешките от закръгляване. В такива случаи обратно устойчивият алгоритъм дава резултат, който е напълно приемлив, доколкото грешките в данните могат да доведат до същия резултат.

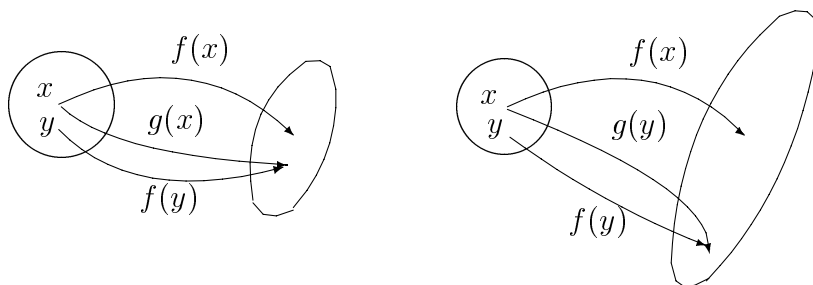
Нека разгледаме дадена изчислителна задача като функция  $f$  на нейния вход  $x$ . Ще казваме, че функцията е *добре обусловена*, ако  $f(x)$  е нечувствителна към малки смущения в  $x$ , т.е. ако  $f(y)$  е близо до  $f(x)$ , когато  $y$  е близо до  $x$ . (Точното определение на термините "нечувствителен", "малко" и "близо" зависи от природата на задачата.) От друга страна, ако  $f$  се изменя силно при малки смущения в  $x$  ще казваме, че тя е *лошо обусловена*. Примери на добре- и лошо-обусловени функции са показани на фиг. 7.1. Кръговете отляво представят областите на неопределеност, в които лежат данните; елипсите отдясно представят областите, в които лежат стойностите на  $f(x)$ . По-голямата елипса отдясно представя лошата обусловеност на  $f$ .



Фигура 7.1: Добре- и лошо-обусловени задачи

Алгоритъмът за решаване на изчислителната задача, представена с функцията  $f$ , може да се разглежда като друга функция  $g$ , която апроксимира  $f$ . Да предположим, че алгоритъмът  $g$  е обратно устойчив и е приложен към данните

$x$ . Тогава  $g(x) = f(y)$  за някакво  $y$  близко до  $x$ . Ако задачата е добре обусловена, то  $f(x)$  трябва да е близо до  $f(y)$  и изчисленото решение  $g(x)$  е точно. От друга страна, ако задачата е лошо обусловена, то изчисленият резултат в общия случай ще е неточен, независимо от устойчивостта на алгоритъма. Но при условие, че грешките в данните са по-големи от обратната грешка, отговорът ще лежи в областта на неопределеност, дължаща се на грешките в данните. Тези две ситуации са илюстрирани на фиг. 7.2.



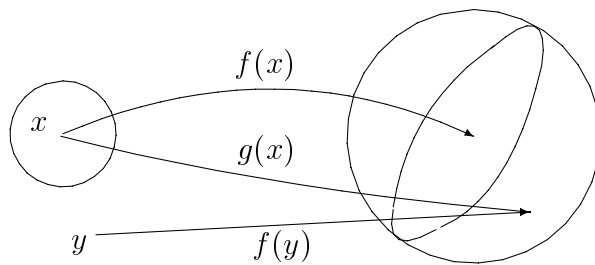
Фигура 7.2: Поведение на устойчив алгоритъм

### Слаба устойчивост

По-горе бе отбелязано, че обратно устойчивият алгоритъм решава добре обусловена задача точно. Не всички алгоритми, обаче, които решават добре обусловени задачи точно, са обратно устойчиви. Така например, даден алгоритъм би могъл да даде решение, което е така точно, както полученото от устойчив алгоритъм, но решението не произлиза от малки смущения в данните. Тази ситуация, която се нарича *слаба устойчивост*, е илюстрирана на фиг. 7.3. Елипсата вдясно представя стойностите на  $f$ , отговарящи на кръга вляво. Големият кръг вдясно представя стойностите, върнати от слабо устойчивия алгоритъм. Тъй като радиусът на кръга е равен на големината на голямата ос на елипсата, алгоритъмът връща стойност, която не е по-малко точна от стойността, върната от устойчив алгоритъм. Но ако стойността не лежи в елипсата, тя не съответства на данни, лежащи в кръга вляво, т.е. на точка близо до входните данни.

### Числа на обусловеност

Колкото и информативен да е обратният анализ на грешките на алгоритъм, той не показва каква точност може да се очаква в изчисленото решение. В означенията, въведени по-горе, обратният анализ на грешките само осигурява, че



Фигура 7.3: Слаба устойчивост

съществува  $y$  близо до  $x$ , такова че  $g(x) = f(y)$ . Но той не дава информация колко далече е  $f(x)$  от  $f(y)$ .

В действителност, задачата за оценяване на точността може да се разглежда в една по-обща постановка, която няма нищо общо с грешките от закръгляване.

*Дадена е функция  $f$  и два аргумента  $x$  и  $y$ , да се оцени разстоянието между  $f(x)$  и  $f(y)$  посредством разстоянието между  $x$  и  $y$ .*

Решаването на такива задачи е обект на математическата дисциплина *теория на смущенията*. Смущенията, които пораждат  $y$  от  $x$ , могат да имат какъвто и да е произход, който не е необходимо да се дължи на грешките от закръгляване. Ето защо ще разгледаме общата задача за смущение на сума от  $n$  числа, а след това ще въведем отново грешките от закръгляване.

Разглеждаме смущенията на сумата

$$s = x_1 + x_2 + \cdots + x_n. \quad (7.12)$$

Смутената задача е

$$\tilde{s} = x_1(1 + \eta_1) + x_2(1 + \eta_2) + \cdots + x_n(1 + \eta_n), \quad (7.13)$$

където се предполага, че смущенията  $\eta_k$  имат обща граница

$$|\eta_k| \leq \varepsilon. \quad (7.14)$$

За да изведем граница за изменението в сумата, изваждаме (7.12) от (7.13) и получаваме

$$\tilde{s} - s = x_1\eta_1 + x_2\eta_2 + \cdots + x_n\eta_n.$$

Вземайки абсолютните стойности и прилагайки (7.14), получаваме

$$|\tilde{s} - s| \leq (|x_1| + |x_2| + \cdots + |x_n|)\varepsilon.$$

По такъв начин относителната грешка в променената сума е

$$\frac{|\tilde{s} - s|}{|s|} \leq \kappa\varepsilon, \quad (7.15)$$

където

$$\kappa = \frac{|x_1| + |x_2| + \cdots + |x_n|}{|x_1 + x_2 + \cdots + x_n|}. \quad (7.16)$$

Дясната страна на (7.15) е относителната грешка в  $\tilde{s}$ . Числото  $\varepsilon$  ограничава относителната грешка в  $x$ -овете. По такъв начин  $\kappa$  е множител, който дава връзката между границата върху смущенията в аргументите на една функция и границата върху смущението, въведено в самата функция. Такова число се нарича *число на обусловеност*.

Така както обратният анализ на грешките дава възможност да се разграничат удовлетворителните от неудовлетворителните алгоритми, числата на обусловеност дават възможност да се разграничат лесните от трудните изчислителни задачи. За разглежданата задача числото на обусловеност никога не е по-малко от единица. То е равно на единица, когато абсолютната стойност на сумата е равна на сумата от абсолютните стойности, което е налице когато всички  $x$ -ове имат еднакви знаци. От друга страна, то е голямо, когато сумата е малка в сравнение с  $x$ -овете. По такъв начин числото на обусловеност не само ограничава грешката, но и хвърля светлина върху причините, които правят сумата трудна за изчисление.

Нека сега свържем резултатите от анализа на смущенията (използува се и понятието *пертурбационен анализ*) на сумата на  $n$  числа с резултатите от обратния анализ на грешките. Обратните грешки  $\eta_k$  удовлетворяват

$$|\eta_1| \leq (n-1)\varepsilon'_M,$$

$$|\eta_k| \leq (n-k+1)\varepsilon'_M, \quad k = 2, 3, \dots, n.$$

По-просто

$$|\eta_k| \leq n\varepsilon'_M.$$

Следователно, грешката в изчислената сума е ограничена от

$$n\kappa\varepsilon'_M.$$

От получения израз се вижда, че натрупването на грешките е представено с множителя  $n\varepsilon'_M$ , който нараства бавно с  $n$ . Така например, ако  $\kappa$  е единица, то можем да загубим осем десетични знака точност само ако  $n$  е по-голямо от 100 милиона. Ето защо даже за големи  $n$  числото на обусловеност може да има много по-силно влияние, отколкото натрупването на грешките от закръгляване. В действителност, даже единствена грешка от закръгляване може да доведе до неточно решаване на лошо обусловена задача.

Необходимо е да се отбележи, че сама по себе си границата върху грешката почти винаги надценява истинската грешка, което се дължи на опроставанията, извършени по време на анализа.

Възможно е да се покаже, че най-малка грешка при сумирането се получава, ако числата се сумират в нарастващ порядък.

Задълбочено изложение на матричната теория на смущенията може да се намери в [23].



## 7.3 Решаване на системи линейни уравнения

Решаването на системата линейни уравнения

$$Ax = b, \quad (7.17)$$

където  $A$  е неособена  $n \times n$  матрица и  $b$  е даден  $n$ -вектор, обикновено се извършва посредством Гаусова елиминация с частичен избор на водещ елемент. При този метод матрицата  $A$  се декомпозира във вида

$$PA = LU, \quad (7.18)$$

където  $L$  е долна триъгълна матрица с единичен диагонал,  $U$  е горна триъгълна матрица и  $P$  е пермутационна матрица (единична матрица с разместени колони), която отразява разместването на редовете при избора на водещ елемент. Решаването на системата (7.17) се свежда тогава до решаването на две системи уравнения с триъгълни матрици

$$Ly = Pb \quad (7.19)$$

и

$$Ux = y. \quad (7.20)$$

Решаването на (7.19) е известно под името *директна елиминация*, а решаването на (7.20) – *обратно заместване*.

Получаването на декомпозицията (7.18) изисква  $n^3/3$  операции с плаваща запетая (flops – от английското floating point operations), а решаването на (7.19) и (7.20) –  $n^2 - n$  такива операции. Следователно решението на системата (7.17) се получава общо с  $n^3/3 + n^2 - n$  flops.

Ако  $A$  е симетрична положително определена матрица, тя се представя с *декомпозицията на Холецки* във вида

$$A = U^T U,$$

където  $U$  е горна триъгълна матрица, след което се решават триъгълните системи

$$U^T y = b$$

и

$$Ux = y.$$

Декомпозицията на Холецки изисква  $n^3/6 - n/6$  flops, така че решението на система уравнения с положително определена матрица се получава с

$$n^3/6 + n^2 - 7n/6$$

flops.

Когато декомпозицията (7.18) е известна, обратната на матрицата  $A$  може да се намери от

$$A^{-1} = U^{-1} L^{-1} P,$$

което изисква около  $2n^3/3$  flops ( $n^3/3$  flops, ако  $A$  е положително определена). Това показва, че решението на системата  $Ax = b$  чрез изчисляване най-напред на  $A^{-1}$  и след това умножаване на  $A^{-1}$  с  $b$  е неефективно, тъй като решението може да се намери директно с  $n^3/3$  flops.

Подробният числен анализ на Гаусовата елиминация показва [4], че матриците  $L$  и  $U$ , получени при използването на частичен избор на водещ елемент, удовлетворяват

$$LU = AP + E, \quad (7.21)$$

където

$$\|E\|_1 \leq n^2 \rho \varepsilon_M \|A\|_1.$$

Уравнението (7.21) показва, че ако множителят  $\rho$  не е голям, изчислените  $L$  и  $U$  образуват точната декомпозиция на леко променена матрица. Вероятността за получаване на големи стойности на  $\rho$  е много малка, поради което на практика Гаусовата елиминация с частичен избор на водещ елемент може да се смята за числено устойчив метод.

Анализът на грешките от закръгляване при решаване на триъгълните системи (7.19) и (7.20) показва, че изчислените решения удовлетворяват

$$(L + \Delta L)y = P^{-1}b$$

и

$$(U + \Delta U)x = y,$$

където

$$\|\Delta L\|_1 \leq n^2 \varepsilon_M, \quad \|U\|_1 \leq n^2 \rho \varepsilon_M \|A\|_1.$$

Следователно изчисленото решение на системата  $Ax = b$  удовлетворява

$$(A + \Delta A)x = b, \quad (7.22)$$

където

$$\|\Delta A\|_1 \leq (2n^3 + n^2) \rho \varepsilon_M.$$

На практика множителят  $2n^3 + n^2$  почти винаги може да се замени с  $n$ , тъй като анализът на грешките от закръгляване отговаря на възможно най-лошия случай.

За да се оцени точността на полученото решение, трябва да се познава обусловеността на задачата, т.е. чувствителността на  $x$  към промени в матрицата  $A$  и вектора  $b$ . Възможно е да се покаже, че ако в системата  $Ax = b$  матрицата  $A$  и векторът  $b$  са смутени с  $\Delta A$  и  $\Delta b$  съответно, то относителното изменение в решението  $x$  удовлетворява

$$\frac{\|\Delta x\|}{\|x\|} \leq \frac{\text{cond}(A)}{1 - \|A^{-1}\Delta A\|} \left( \frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta b\|}{\|b\|} \right), \quad (7.23)$$

където

$$\text{cond}(A) = \|A\| \|A^{-1}\|$$

е числото на обусловеност на матрицата  $A$  по отношение на обръщането.

Неравенството (7.23) означава, че  $\text{cond}$  ограничава отгоре отношението на относителната неопределеност в решението към относителната неопределеност в  $A$  и  $b$ . Тъй като  $\text{cond} \geq 1$ , тази граница не може да е по-малка от 1.

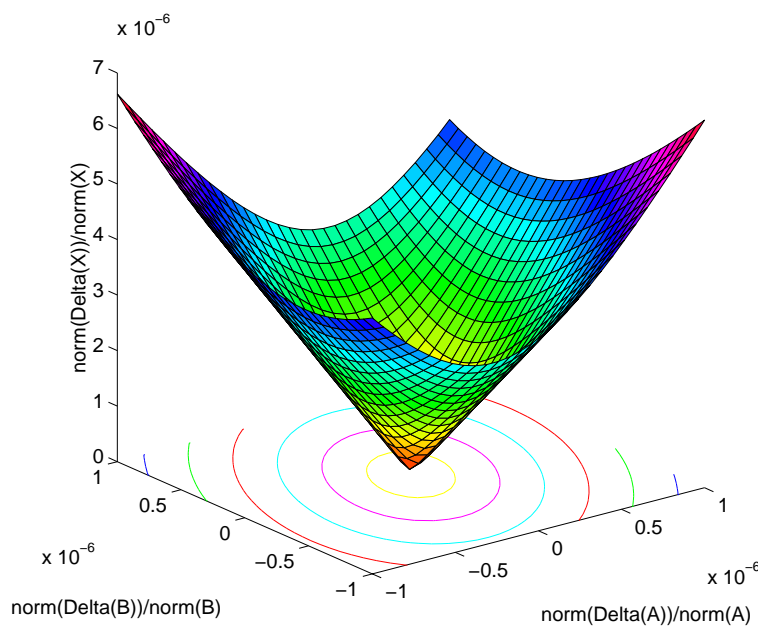
Ако  $\text{cond} = 1$ , то относителната неопределеност в  $x$  е равна на относителната неопределеност в  $A$  и  $b$ . Ако обаче  $\text{cond} = 10^6$ , то  $\frac{\|\Delta x\|}{\|x\|}$  може да е милион пъти по-голямо от  $\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta b\|}{\|b\|}$ . Ако  $\text{cond}$  е относително малко, то матрицата  $A$  се нарича *добре обусловена*, в противен случай  $A$  е *лошо обусловена* по отношение на задачата за решаване на система линейни уравнения.

**Пример 7.3** Разглежда се система линейни уравнения от 5-ти ред с матрици

$$A = \begin{bmatrix} 0 & 0 & -0.5 & 1 & 0 \\ 1 & 1 & -1.5 & 0 & 1 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 1 & 0.5 & -1 & 0 \\ 1 & 0 & -2 & 2 & 1 \end{bmatrix}, b = \begin{bmatrix} 0.5 \\ 1.5 \\ 1 \\ 0.5 \\ 2 \end{bmatrix}.$$

Точното решение на тази система е

$$x^T = [1 \ 1 \ 1 \ 1 \ 1].$$



Фигура 7.4: Смутени решения на система линейни уравнения

На фиг. 7.4 са показани относителните изменения в решението, дължащи се на смущения в данните

$$\Delta A = xE, \quad x \in [-10^{-6}, 10^{-6}] \times \|A\|_2,$$

$$\Delta b = yf, \quad y \in [-10^{-6}, 10^{-6}] \times \|b\|_2,$$

където  $E$ ,  $f$  са генерирани като случайни матрици с единична 2-норма. Системата уравнения е добре обусловена,

$$\text{cond}_2 = 33.01,$$

поради което относителните изменения в данните имат порядъка на относителните изменения в данните.

Нека  $A$  е неособена матрица и  $x^* = A^{-1}b$  е точното решение на  $Ax = b$ . Като се приложи (7.23) към смутеното уравнение (7.22), може да се докаже, че при малки стойности на  $\|A^{-1}\Delta A\|$  изчисленото решение на  $Ax = b$  удовлетворява

$$\frac{\|x - x^*\|_1}{\|x\|} \leq (2n^3 + n^2)\rho\varepsilon_M \text{cond}(A). \quad (7.24)$$

Този резултат показва, че ако  $\text{cond}$  е голямо, т.е. ако системата уравнения е лошо обусловена, относителната грешка в  $x$  може да е голяма. Ето защо, въпреки че решението е получено с помощта на числено устойчив метод, то може да е много неточно, ако матрицата  $A$  е лошо обусловена.

Оценяването на точността при решаване на система линейни уравнения може да се извърши по следното приблизително правило. Ако решението е получено при точнос  $\varepsilon_M \approx 10^{-p}$  и системата има число на обусловеност  $\text{cond}(A) = 10^{-q}$ , то полученото решение ще има в най-лошия случай само  $p - q$  верни десетични значещи цифри. Правилото е приблизително, тъй като  $\text{cond}$  дава горната граница на грешката. Ако  $q \geq p$ , т.е. ако  $\text{cond} \geq 1/\varepsilon_M$ , възможно е в полученото решение да няма нито една верна значеща цифра.

**Пример 7.4** Дадена е системата линейни уравнения  $Ax = b$ , където

$$A = \begin{bmatrix} -6.49 & -4.95 & 3.60 & 9.70 & 38.50 \\ -13.46 & -18.80 & -14.10 & -28.70 & 59.50 \\ -7.78 & -9.90 & -5.10 & -10.60 & 36.40 \\ 7.58 & 9.90 & 5.90 & 12.00 & -35.00 \\ -4.19 & -4.95 & -2.10 & -3.80 & 20.60 \end{bmatrix}, \quad b = \begin{bmatrix} 40.36 \\ -15.36 \\ 3.02 \\ 0.38 \\ 5.56 \end{bmatrix}.$$

Точното решение на системата е

$$x = [1, 1, 1, 1, 1]^T.$$

Матрицата  $A$  е лошо обусловена, като  $\text{cond}_2(A) = 1.1199 \times 10^6$ . При това число на обусловеност може да се очаква, че в аритметика с двойна точност (около 16 верни десетични знака) компонентите на изчисленото решение ще имат около 10 верни десетични знака. Решението на системата, получено с MATLAB при двойна точност ( $\varepsilon_M = 2.22 \times 10^{-16}$ ) е (до четиринадесетия десетичен знак)

$$\bar{x} = \begin{bmatrix} 0.99999999999963 \\ 0.999999999999766 \\ 0.999999999999889 \\ 1.000000000000113 \\ 0.999999999999946 \end{bmatrix}.$$

Относителната грешка в резултата

$$\frac{\|\bar{x} - x\|}{\|x\|} = 1.296 \times 10^{-12}$$

е приблизително от същия порядък, както величината

$$\text{cond}_2(A)\varepsilon_M = 2.487 \times 10^{-10}.$$

## 7.4 Елементарни ортогонални преобразувания

Матричните задачи обикновено се решават чрез привеждане на матриците, присъстващи в тези задачи, в някаква специална форма (например диагонална, триъгълна и т.н.), което дава възможност лесно да се намери решението. Привеждането в специална форма най-често се извършва с последователност от елементарни ортогонални преобразувания, с които може да се осигури числена устойчивост на съответния алгоритъм. Използват се два вида елементарни ортогонални преобразувания: равнинни ротации и елементарни отражения.

*Равнинна ротация (или ротация на Гивънс)* в  $(i, j)$ -тата равнина е матрица във вида

$$R = \begin{bmatrix} 1 & & & & & & & 0 \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & p & & & q & \\ & & & & 1 & & & \\ & & & & & \ddots & & \\ & & & & & & 1 & \\ & & & -q & & & p & \\ & & & & & & & 1 \\ & & & & & & & \ddots \\ 0 & & & & & & & & 1 \end{bmatrix},$$

където нетривиалните елементи  $p$  и  $q$  на ротацията са разположени в  $(i, j)$ -тите редове и колони и

$$p^2 + q^2 = 1.$$

Равнинната ротация се прилага лесно върху вектор. Ако  $y = Rx$ , то  $y$  се различава от  $x$  само в  $i$ -тата и  $j$ -тата компоненти, които се дават от

$$\begin{aligned} y_i &= px_i + qx_j, \\ y_j &= px_j - qx_i. \end{aligned}$$

Равнинната ротация се използва за въвеждане на нула в елемент на вектор или матрица. Ако  $a$  е даден вектор, то може да се конструира равнинна ротация,

която анулира  $j$ -тия елемент на  $a$ . Действително, ако се обозначи  $b = Ra$  и се положи

$$p = \frac{a_i}{\sqrt{a_i^2 + a_j^2}}, \quad q = \frac{a_j}{\sqrt{a_i^2 + a_j^2}}$$

се получава  $p^2 + q^2 = 1$  и

$$\begin{aligned} b_i &= \sqrt{a_i^2 + a_j^2} \\ b_j &= 0. \end{aligned}$$

Последователност от равнинни ротации може да се използва за въвеждане на няколко нули в определени позиции на вектори или матрици.

Изчислената ротация  $\bar{R}$  удовлетворява

$$\bar{R} = R + \Delta R,$$

където

$$\|\Delta R\|_F \leq 5\sqrt{2}\varepsilon_M.$$

Следователно изчислената ротация  $\bar{R}$  е близка до точната ротация, съответстваща на  $a_i, a_j$ .

Конструираната ротация обикновено се използва за извършване на преобразуване на подобие  $RAR^T$  върху  $n \times n$  матрица  $A$ . Лявото умножение с равнинна ротация

$$RA = \begin{bmatrix} 1 & & & & & & 0 \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & p & q & & \\ & & & -q & p & & \\ & & & & & 1 & \\ & & & & & & \ddots \\ 0 & & & & & & & 1 \end{bmatrix} \begin{bmatrix} \times & \cdots & \times & a_{ij} & \times & \cdots & \times \\ \times & \cdots & \times & a_{i+1,j} & \times & \cdots & \times \end{bmatrix}$$

променя само  $i$ -тия и  $(i+1)$ -вия редове на  $A$ . Полученото произведение може да презапише  $A$  в същия масив, за да се икономиса памет. Анализът на грешките от закръгляване върху изчисленото произведение показва, че

$$\|\mathfrak{fl}(\bar{R}A) - RA\|_F < 8\varepsilon_M \|A\|_F. \quad (7.25)$$

Като се вземе предвид, че  $R$  е ортогонална, от (7.25) следва, че

$$\mathfrak{fl}(\bar{R}A) = R(A + \Delta A),$$

където

$$\|\Delta A\|_F < 8\varepsilon_M \|A\|_F.$$

Това показва, че изчисленото произведение остава точно за леко променена матрица  $A$ , т.е. съответният алгоритъм е числено устойчив.

Дясното умножение с равнинна ротация променя само  $j$ -тата и  $(j+1)$ -вата колони на матрицата  $A$ . Преобразуването на подобие на матрица  $A$  с равнинна ротация дава

$$\mathfrak{H}(\bar{R}A\bar{R}^T) = R(A + \Delta A)R^T,$$

където

$$\|\Delta A\|_F < 16\varepsilon_M \|A\|_F.$$

*Елементарно отражение (или отражение на Хаусхолдер)* е матрица във вида

$$U = I - uu^T/h,$$

където  $u$  е вектор и

$$h = \|u\|_2^2/2. \quad (7.26)$$

От това определение следва, че  $U$  е симетрична и ортогонална матрица.

Елементарното отражение се използва за едновременно въвеждане на няколко нули във вектор или матрица. За даден вектор  $a$  е възможно да се намери отражение  $U$ , такова че

$$Ua = -ge_1, \quad (7.27)$$

където  $e_1 = [1, 0, \dots, 0]^T$ . (Знакът  $-$  (минус) е избран за удобство.) Константата  $g$  е определена с точност до знака от факта, че  $U$  е ортогонална,

$$g = \pm \|a\|_2. \quad (7.28)$$

Може да се покаже, че ако векторът  $u$  е избран от

$$u = a + ge_1 \quad (7.29)$$

и ако  $h$  удовлетворява (7.26), то (7.27) е в сила.

Анализът на (7.29) показва, че ако  $g$  и  $a_1$  имат различни знаци, може да се получи унищожаване при изчисляването на  $u_1$  посредством (7.29). Ето защо  $g$  се взема със същия знак като  $a_1$ , т.е.

$$g = \text{sign}(a_1) \|a\|_2.$$

Подробният анализ на грешките показва, че изчисленото отражение  $\bar{U}$  удовлетворява

$$\|\bar{U} - U\|_2 \leq (4n + 22)\varepsilon_M.$$

Това неравенство показва, че изчисленото отражение  $\bar{U}$  е близко до точното отражение  $U$ , съответстващ на дадения вектор  $a$ .

След определянето на  $U$  често се изисква да се изчисли произведението  $UA$ , където  $A$  е дадена  $n \times m$  матрица. Ако  $A$  се запише като

$$A = [A_1, A_2, \dots, A_m],$$

то

$$UA = [UA_1, UA_2, \dots, UA_m].$$

За да се намали обемът на изчисленията, матрицата  $U$  не се формира в явен вид, а се използва фактът, че

$$UA_j = A_j - h^{-1}(u^T A_j)u.$$

Анализът на грешките от закръгляване показва, че

$$\mathfrak{fl}(\bar{U}A) = U(A + \Delta A),$$

$$\|\Delta A\|_F < (6n + 27)\varepsilon_M \|A\|_F.$$

Следователно изчисленото произведение  $\mathfrak{fl}(\bar{U}A)$  може да се разглежда като резултат от точното умножаване на  $U$  с леко променената матрица  $A$ .

## 7.5 QR–декомпозиция

Нека  $A$  е реална  $n \times m$ -матрица. Тогава съществуват ортогонална матрица  $Q$ ,  $Q^T Q = I_n$  и горна триъгълна матрица  $R$  със същите размерности като  $A$ , такива че  $A = QR$ . Ако  $A$  е комплексна,  $Q$  е унитарна матрица,  $Q^H Q = I_n$ . Декомпозицията  $A = QR$  се нарича *QR–декомпозиция* и има редица приложения, като решаване на задачата за най-малки квадрати, определяне на ранга на матрица и т.н.

QR–декомпозицията на  $n \times m$  матрица  $A$  може да се получи по следния начин. Най-напред се конструира елементарно отражение  $U_1$ , такова, че

$$U_1 a_1 = -g_1 e_1,$$

където  $a_1$  е първата колона на  $A$ . По такъв начин

$$A_2 = U_1 A = \begin{bmatrix} -g_1 & \times & & \dots & \times \\ 0 & | & & & \\ \vdots & | & & & \\ 0 & | & & A'_2 & \end{bmatrix}.$$

Второто отражение  $U'_2$  се избира така, че да въведе нули в първата колона  $a'_2$  на  $A'_2$ , т.е.

$$U'_2 a'_2 = -g_2 e'_1.$$

Лявото умножение на  $A_2$  с

$$U_2 = \text{diag}(1, U'_2)$$

дава

$$A_3 = U_2 A_2 = \begin{bmatrix} -g_1 & \times & \times & \dots & \times \\ 0 & -g_2 & \times & \dots & \times \\ 0 & 0 & | & & \\ \vdots & \vdots & | & & \\ 0 & 0 & | & A'_3 & \end{bmatrix}.$$



Това преобразуване не изменя елементите в първата колона на  $A_2$ .

Привеждането продължава, докато се нулират всички поддиагонални елементи на  $A$ .

В общия случай е възможно колоните на  $A$  да са линейно зависими, поради което някои от диагоналните елементи на получената горна триъгълна матрица са нулеви. Например, за  $5 \times 4$  матрица  $A$  резултатът от привеждането може да изглежда по следния начин:

$$R = Q^T A = \begin{bmatrix} \times & \times & \times & \times \\ 0 & \times & 0 & \times \\ 0 & 0 & 0 & \times \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

В този случай първата и третата колони на  $A$  са линейно зависими и матрицата е от ранг 3. В такива случаи се използва QR-декомпозиция с избор на водеща колона, при която се извършва размяна на колоните на  $A$  по време на привеждането, така че линейно независимите колони да се окажат първи. При тази декомпозиция първите  $r = \text{rank}(A)$  диагонални елементи на  $R$  се получават ненулеви.

QR-декомпозицията с избор на водеща колона се получава по следния начин. На  $k$ -тата стъпка се търси колоната на  $(n - k + 1) \times (n - k + 1)$  матрицата  $A'_k$ , разположена в долния десен ъгъл на  $A_k$ , която има максимална норма. Тази колона се разменя с  $k$ -тата колона на  $A_k$ , след което се извършва привеждането. За да се намалят изчисленията, нормите на колоните на  $A$  се определят предварително и техните стойности се подновяват на всяка стъпка. Тази стратегия води до матрица  $R$ , която удовлетворява условието

$$r_{kk}^2 \geq r_{kj}^2 + \dots + r_{jj}^2, \quad j = k + 1, k + 2, \dots, m.$$

Това означава, че ако  $r_{kk}$  е нула, редовете от  $k$  до  $n$  са също нулеви, което може да се използва за определяне на ранга на  $A$ .

Анализът на грешките при QR-декомпозицията показва, че изчислената горна триъгълна матрица  $R$  удовлетворява

$$\bar{R} = Q^T (A + \Delta A),$$

$$\|\Delta A\|_F \leq (6n - 3\ell + 30)\ell\varepsilon\|A\|_F,$$

където  $\ell = \min\{n - 1, m\}$ . Това показва, че алгоритъмът, при който се използват елементарни отражения, е числено устойчив.

## 7.6 Изчисляване на собствени стойности

### Чувствителност на собствените стойности

Числено устойчивите методи за решаване на задачата за собствени стойности дават множество от приблизителни собствени стойности на матрицата  $A$ , които са точните собствени стойности на леко променената матрица  $A + E$ . Ето защо точността на изчислените собствени стойности зависи от тяхната чувствителност към смущения в матричните елементи. Ако малки смущения в елементите на  $A$  водят до големи изменения в някои от нейните собствени стойности, съответната задача за собствени стойности се нарича лошо обусловена.

Нека разгледаме ефекта върху собствените стойности на смущения в  $A$  от вида  $A + \mu E$ , където  $E$  е фиксирана произволна матрица с  $\|E\|_2 = 1$  и  $\mu$  е променлив параметър. Да предположим най-напред, че  $A$  не е дефектна матрица, т.е. че тя има пълна система от линейно независими собствени вектори  $x_j$ ,  $j = 1, 2, \dots, n$ .

Най-простата мярка за чувствителността на собствените стойности е числото на обусловеност на матрицата от собствените вектори

$$\text{cond}_2(X) = \|X\|_2 \|X^{-1}\|_2.$$

Съгласно *теоремата на Бауер–Файк* [4] измененията в елементите на  $A$  могат да водят до  $\text{cond}_2(X)$  пъти по-големи изменения в собствените стойности. Поради зависимостта си от спектралната норма на  $X$  числото  $\text{cond}_2(X)$  нарича спектрално число на обусловеност. Тъй като  $\text{cond}_2(X)$  характеризира цялостната чувствителност на собствените стойности на  $A$ , то може да се разглежда като число на обусловеност на  $A$  по отношение на задачата за собствени стойности.

Недостатък на теоремата на Бауер–Файк е, че тя може да даде твърде пессимистични оценки, ако някои от собствените стойности на матрицата са много по-нечувствителни от останалите.

Ако матрицата  $A$  е реална и симетрична (или комплексна и Ермитова), то матрицата от собствените вектори  $X$  е ортогонална (респективно унитарна) и  $\text{cond}_2(X) = 1$ . Ето защо задачата за собствени стойности е винаги добре обусловена. В съответствие с *теоремата на Виландт–Хофман*, ако  $A$  е симетрична, то

$$\sqrt{\sum_{j=1}^n (\delta\lambda_j)^2} \leq n\alpha,$$

където  $\delta\lambda_j$  е изменението в  $\lambda_j$ , дължащо се на смущение  $\alpha E$  в матрицата  $A$ . Очевидно е, че  $|\delta\lambda_j| \leq n\alpha$ , за  $j = 1, 2, \dots, n$ .

Изследването на чувствителността на отделните собствени стойности включва разглеждането на десните и левите собствени вектори на  $A$ . Нека тези собствени вектори са нормализирани, така че  $\|x_j\|_2 = \|y_j\|_2 = 1$ . Може да се покаже [4], че при малки стойности на параметъра  $\mu$  е изпълнено

$$|\lambda_j(\mu) - \lambda_j| \leq \text{cond}(\lambda_j) |\mu|, \quad (7.30)$$

където

$$\text{cond}(\lambda_j) = \frac{1}{|y_j^H x_j|}$$

се нарича *число на обусловеност на собствената стойност*  $\lambda_j$ .

Уравнението (7.30) показва, че ако смущенията в  $A$  имат порядък  $\mu$ , собствената стойност  $\lambda_j$  може да бъде променена с

$$|\mu| \text{cond}(\lambda_j).$$

Ето защо ако  $\text{cond}(\lambda_j)$  е голямо, задачата за собствената стойност  $\lambda_j$  е лошо обусловена.

Спектралното число на обусловеност  $\text{cond}_2(X)$  и числата на обусловеност  $\text{cond}(\lambda_j)$  са свързани с отношението

$$\text{cond}(\lambda_j) \leq \text{cond}_2(X), \quad j = 1, 2, \dots, n.$$

За симетрична матрица  $y_j^H x_j = 1$  и  $\text{cond}(\lambda_j) = 1$ , така че от (7.30) също следва, че задачата за собствени стойности е добре обусловена. Когато матрицата е несиметрична, векторите  $x_j$  и  $y_j$  могат да са почти ортогонални и чувствителността на собствените стойности може да е изключително висока.

**Пример 7.5** Ако

$$A = \begin{bmatrix} a & 1 \\ 0 & a + \mu \end{bmatrix},$$

то  $\lambda_1 = a$ ,  $\lambda_2 = a + \mu$  и

$$x_1^H = [1, 0], \quad y_1^H = [-\mu, 1]/[1 + \mu^2]^{\frac{1}{2}},$$

$$x_2^H = [1, \mu]/[1 + \mu^2]^{\frac{1}{2}}, \quad y_2^H = [0, 1].$$

По такъв начин

$$y_1^H x_1 = -\mu/[1 + \mu^2]^{\frac{1}{2}}, \quad y_2^H x_2 = \mu/[1 + \mu^2]^{\frac{1}{2}}.$$

Тези величини клонят към нула, когато  $\mu \rightarrow 0$ , така че  $\text{cond}(\lambda_1)$  и  $\text{cond}(\lambda_2)$  могат да станат произволно големи. При  $\mu = 0$  матрицата  $A$  е дефектна, т.е. тя има нелинеен елементарен делител и числата на обусловеност може да се разглеждат като безкрайно големи. Разбира се, в този случай измененията в собствените стойности са крайни, но линейната оценка на чувствителността (7.30) повече не е в сила.

Задачата за кратни собствени стойности, съответстващи на нелинеен елементарен делител, обикновено е лошо обусловена. В общия случай, ако  $\lambda$  е кратна собствена стойност с кратност  $k$ , която отговаря на нелинеен елементарен делител, то смущения в  $A$  с порядък  $\mu$  могат да доведат до изменения в  $\lambda$  с порядък  $\sqrt[k]{\mu}$ .

**Пример 7.6** Нека

$$J_{-1}^{(5)} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

Ако елементът в позицията (5,1) се замени с  $\alpha = 10^{-5}$ , то характеристичното уравнение става

$$(\lambda + 1)^5 = \alpha$$

и кратната собствена стойност  $\lambda = -1$  се замества с пет различни собствени стойности, дадени по-долу с точност до осмата значеща цифра

$$\begin{aligned} \lambda_1 &= -0.900\,000\,00 + 0.000\,000\,00i, \\ \lambda_{2,3} &= -0.969\,098\,30 \pm 0.095\,105\,652i, \\ \lambda_{4,5} &= -1.080\,901\,7 \pm 0.058\,778\,525i \end{aligned}$$

Вижда се, че измененията в собствените стойности имат порядък  $10^{-1}$ , както се предвижда от анализа.

На практика, кратностите на собствените стойности и степените на елементарните делители трябва да се определят от изчислените собствени стойности. Ето защо определянето на кратностите на собствените стойности и точното получаване на Жордановата форма на матрица може да представлява трудна задача.

Анализът на чувствителността в присъствието на кратни собствени стойности може да се извърши с теоремата на Гершгорин (вж. Приложението). Съгласно тази теорема всяка собствена стойност на  $A$  лежи в един от кръговете в комплексната равнина, центърът на който е в  $a_{jj}$ , а радиусът му е

$$\sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|,$$

т.е.

$$|\lambda - a_{jj}| \leq \sum_{j \neq i} |a_{ij}|, \quad i = 1, 2, \dots, n.$$

**Пример 7.7** Ако

$$A = \begin{bmatrix} -1 \times 10^0 & 2 \times 10^{-5} & -1 \times 10^{-5} \\ -1 \times 10^{-5} & -1 \times 10^0 & -1 \times 10^{-5} \\ 3 \times 10^{-5} & 2 \times 10^{-5} & -2 \times 10^0 \end{bmatrix}$$

то кръговете на Гершгорин  $C_1$ ,  $C_2$ ,  $C_3$  на  $A$  лежат в -1, -1, -2 и техните радиуси съответно са  $3 \times 10^{-5}$ ,  $2 \times 10^{-5}$  и  $5 \times 10^{-5}$ . Следователно всички собствени стойности на  $A$  лежат в обединението на кръговете  $C_1$  и  $C_3$ .

Необходимо е да се отбележи, че теоремата на Гершгорин не уточнява в кои кръгове лежат собствените стойности. Възможно е собствените стойности да лежат в някои от кръговете, като останалите кръгове са празни, т.е. невинаги е вярно, че всеки кръг на Гершгорин съдържа собствена стойност.

Ако  $k$  кръгове на Гершгорин на матрицата  $A$  са отделени от останалите кръгове, то точно  $k$  собствени стойности лежат в обединението на тези кръгове. В частност, ако всеки кръг е изолиран от останалите, то той съдържа точно една собствена стойност на  $A$ . За матрицата в пример 7.7 кръговете  $C_1$  и  $C_2$  са изолирани от  $C_3$ . Ето защо две собствени стойности на  $A$  са в обединението на  $C_1$  и  $C_2$ , а кръгът  $C_3$  съдържа точно една собствена стойност. Точните собствени стойности на  $A$  до петнадесетия десетичен знак са

$$\begin{aligned}\lambda_{1,2} &= -1.000\,000\,000\,249\,998 \pm 0.000\,014\,142\,277\,046i, \\ \lambda_3 &= -1.999\,999\,999\,500\,004.\end{aligned}$$

За получаване на по-добри приближения за собствените стойности радиусите на изолираните кръгове на Гершгорин могат да се намаляват, като се използват диагонални преобразувания на подобие върху матрицата  $A$ . За целта съответният елемент на диагоналното преобразуване се избира така, че радиусът на кръга да бъде колкото се може по-малък, като се поддържат в същото време останалите кръгове достатъчно малки, така че да останат изолирани.

За анализ на чувствителността на кратни собствени стойности може да се използва модификация на теоремата на Хенричи [23], която се формулира по следния начин.

Нека  $S^{-1}AS = J$  е жордановата канонична форма на  $A$  и нека  $m$  е размерността на най-големият жорданов блок в  $J$ . Тогава за всяка собствена стойност  $\tilde{\lambda} \in \lambda(A + E)$ , съществува собствена стойност  $\lambda$  на  $A$ , такава че

$$\frac{|\tilde{\lambda} - \lambda|^m}{1 + |\tilde{\lambda} - \lambda| + \dots + |\tilde{\lambda} - \lambda|^{m-1}} \leq \|S^{-1}ES\|_2. \quad (7.31)$$

Ако с  $\omega(m, \delta)$  се означаи неотрицателният корен на уравнението

$$\eta^m = \delta(1 + \eta + \dots + \eta^{m-1}), \quad (7.32)$$

то може да се установи взаимно еднозначно съответствие между собствените стойности на матриците  $A$  и  $A + E$ :  $\tilde{\lambda}_i \leftrightarrow \lambda_{\tau_i}$ ,  $i = 1, 2, \dots, n$ , при което

$$|\tilde{\lambda}_i - \lambda_{\tau_i}| \leq (2n - 1)\omega(m, \|S^{-1}ES\|_2). \quad (7.33)$$

По такъв начин собствените стойности на  $A + E$  се намират в кръгове с центрове собствените стойности на  $A$  и радиуси определени от (7.33).

Практическото използване на тази теорема изисква познаването на жордановата форма на матрицата  $A$ . При определянето на корените на (7.32) може да се използва оценката

$$\|S^{-1}ES\|_2 \leq \text{cond}_2(S)\|E\|_2.$$

За илюстриране на чувствителността на собствените стойности могат да се използват псевдоспектърът на матрицата и нейният спектрален портрет. За дадено  $\varepsilon > 0$ ,  $\varepsilon$ -псевдоспектърът на матрица  $A \in \mathbb{C}^{n \times n}$  се дефинира като множеството

$$\Lambda_\varepsilon(A) = \{z : z \text{ е собствена стойност на } A + E \text{ за някакво } E \text{ с } \|E\|_2 \leq \varepsilon\}.$$

Иначе казано, това е множеството на всички комплексни числа, които са собствени стойности на  $A + E$  за някакво смущение  $E$  с 2-норма, която е най-много равна на  $\varepsilon$ .

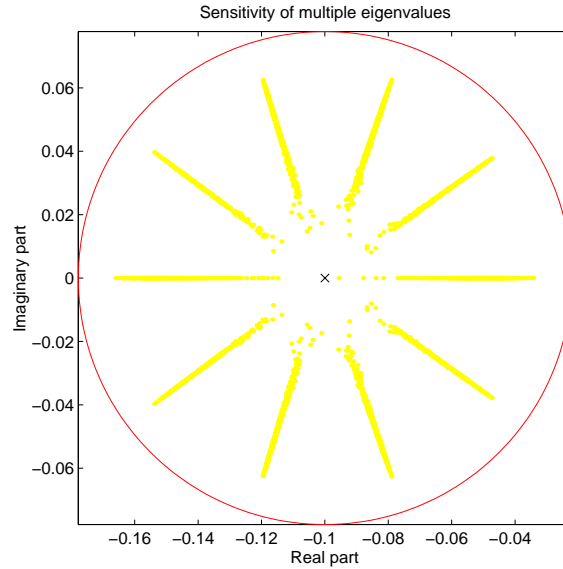
*Спектралният портрет* на матрицата  $A$  се дефинира като функцията

$$f(z) = \log_{10} \|(zI - A)^{-1}\|_2 \|A\|_2 = \log_{10} \frac{\sigma_{\max}(A)}{\sigma_{\min}(zI - A)}.$$

над комплексната равнина.

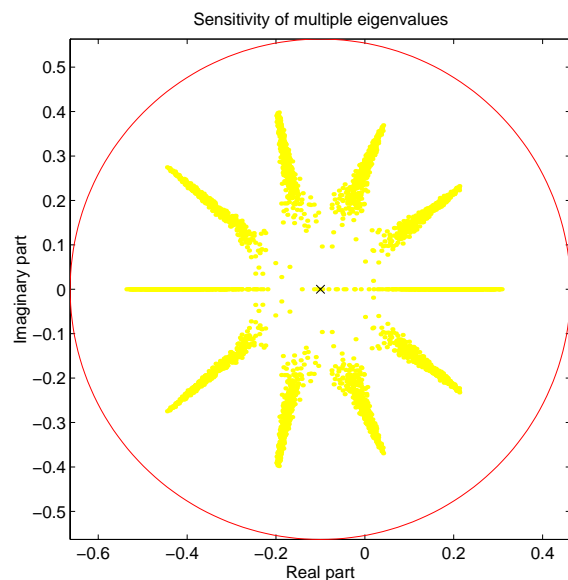
**Пример 7.8** Дадена е матрицата

$$A = \begin{bmatrix} -35.1 & 27.5 & 8.0 & -37.0 & 63.0 \\ 12.0 & -12.1 & 0.5 & 12.0 & -24.0 \\ -20.0 & 12.0 & 7.9 & -21.5 & 33.0 \\ -16.0 & 8.0 & 8.0 & -18.1 & 25.5 \\ -32.0 & 24.0 & 8.0 & -34.0 & 56.9 \end{bmatrix}.$$



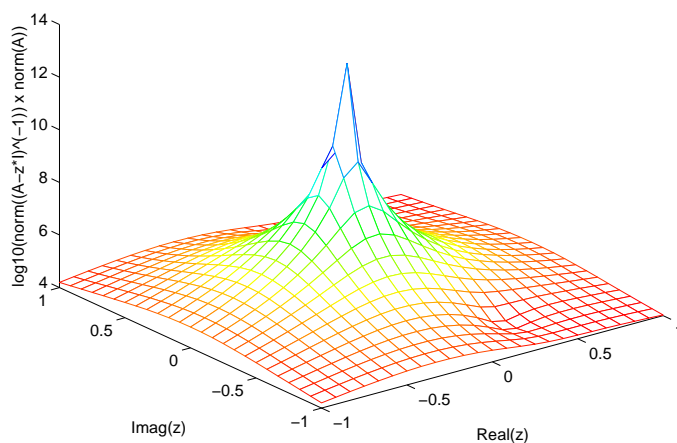
Фигура 7.5: Псевдоспектър при  $\varepsilon = 10^{-8}$

Тази матрица е подобна на Жорданов блок  $J_{-0.1}^{(5)}$ , т.е. има петкратна собствена стойност в  $-0.1$ . При  $\|E\|_2 = 10^{-8}$  и  $\|E\|_2 = 10^{-4}$ , за неотрицателните корени на



Фигура 7.6: Псевдоспектър при  $\varepsilon = 10^{-4}$

(7.32) получаваме съответно  $7.78 \times 10^{-2}$  и  $5.63 \times 10^{-1}$ . На фиг. 7.5 и 7.6 са показани кръгове с тези радиуси и псевдоспектрите на матриците  $A + E$ , съответно за  $\varepsilon = 10^{-8}$  и  $\varepsilon = 10^{-4}$ . Псевдоспектрите са получени с функцията `ps` от Test Matrix Toolbox [17]. Правят впечатление големите изменения в собствените стойности на матрицата при сравнително малки изменения в нейните елементи, което се дължи на присъствието на всички кратни собствени стойности в един Жорданов блок.



Фигура 7.7: Спектрален портрет на матрицата от пример 7.8

На фиг. 7.7 е показан спектралният портрет на същата матрица, получен с

функцията `portrait.m` от Precise Toolbox [9]. Максимумът на функцията  $f(z)$  се получава при  $z = -0.1$ , което отговаря на кратната собствена стойност.

### QR-метод за определяне на собствените стойности

Съвременните числени методи за определяне на собствените стойности и собствените вектори се състоят в изчисляване на преобразуване на подобие, което привежда матрицата  $A$  от общ вид в матрица  $\tilde{A} = P^{-1}AP$  в специален вид, за който задачата за собствени стойности може да се реши просто. На практика, обаче, не е целесъобразно да се използва преобразуване на подобие в общ вид, тъй като неговата матрица може да е лошо обусловена, което ще доведе до грешки при изчисляването на  $P^{-1}AP$ . Поради тази причина преобразуванията се ограничават до някъкъв клас от матрици, които са винаги добре обусловени. Един такъв клас от преобразувания е класът на унитарните матрици, които се характеризират с число на обусловеност, равно на единица. Разбира се, при използването на такива преобразувания не може да се очаква, че ще се постигне значително опростяване на задачата (например, матрицата  $\tilde{A}$  да е в диагонален вид), тъй като не се използват възможностите на общите преобразувания на подобие.

Нека матрицата  $A$  има собствени стойности  $\lambda_1, \lambda_2, \dots, \lambda_n$ . Може да се докаже, че съществува унитарно преобразуване с матрица  $U$ , такава че  $U^H AU$  е горна триъгълна матрица с диагонални елементи  $\lambda_1, \lambda_2, \dots, \lambda_n$ . Тази триъгълна матрица се нарича *канонична форма на Шур* на матрицата  $A$ . Колоните на  $U$  се наричат *вектори на Шур*. Ако  $A$  е ермитова матрица ( $A^H = A$ ), триъгълната форма се превръща в диагонална, т.е. посредством унитарно преобразуване на подобие всяка ермитова матрица може да се приведе в диагонална форма.

Ако първоначалната матрица е реална, използването на унитарно преобразуване ще въведе комплексни елементи в  $U^H AU$ . Ето защо при изчисления с реални матрици е желателно да се използват реални унитарни, т.е. ортогонални преобразувания. Ако  $A$  има комплексни собствени стойности, очевидно е, че тя не може да се приведе в строго триъгълна форма с комплексни елементи върху диагонала, използвайки ортогонални преобразувания. Тъй като собствените стойности на реална матрица  $A$  присъстват в спрегнати двойки, естествено е да се очаква, че  $A$  може да се приведе в квазитриъгълна форма, в която  $2 \times 2$  диагоналните блокове имат спрегнати собствени стойности. Действително, ако  $A$  е реална матрица, винаги съществува ортогонална матрица  $U$ , такава че  $U^T AU$  е квазитриъгълна. Нещо повече,  $U$  може да се избере така, че всеки  $2 \times 2$  диагонален блок на  $U^T AU$  да има само комплексни собствени стойности (които следователно трябва да са спрегнати). Получената квазитриъгълна форма се нарича *реална форма на Шур* на матрицата  $A$ .

Формата на Шур на дадена матрица не може да се получи с помощта на крайна последователност от преобразувания. Ето защо за тази цел се използват итеративни методи, при които поддиагоналните елементи се анулират, като се използва безкрайна последователност от преобразувания на подобие. Най-широко прилаганият метод за итеративно привеждане във форма на Шур е QR-алгоритъмът,



при който се използват ортогонални преобразувания на подобие на всяка стъпка. За да се намали обемът на изчисленията, дадената матрица се преобразува предварително в някаква проста форма, която може да се получи с директен алгоритъм (т.е. с краен брой преобразувания на подобие) и съдържа голям брой нули. За тази цел матрицата се обикновено се преобразува в горна форма на Хесенберг, която се получава с ортогонални преобразувания на подобие и има нулеви елементи под първия диагонал.

Нека  $A_1 = A$  е комплексна  $n \times n$  матрица. QR-алгоритъмът произвежда последователност от подобни матрици  $A_1, A_2, \dots$  по следния начин. За дадена матрица  $A_s$  от елементите и се определя скалар  $k_s$ , наречен *изместване на началото* (със сходимостта на итеративния процес  $k_s$  приближава собствена стойност на  $A$ ). След това матрицата  $A_s - k_s I$  се факторизира във вида

$$A_s - k_s I = Q_s R_s, \quad (7.34)$$

където  $Q_s$  е унитарна, а  $R_s$  е горна триъгълна матрица. След това се изчислява матрица  $A_{s+1}$ , такава че

$$A_{s+1} = R_s Q_s + k_s I. \quad (7.35)$$

От (7.34) имаме, че

$$R_s = Q_s^H (A_s - k_s I)$$

и следователно от (7.35) получаваме

$$A_{s+1} = Q_s^H (A_s - k_s I) Q_s + k_s I = Q_s A_s Q_s,$$

така че действително  $A + s + 1$  е унитарно подобна на  $A_s$ .

Може да се покаже, че

$$A_{s+1} = (Q_1 Q_2 \cdots Q_s)^H A Q_1 Q_2 \cdots Q_s$$

и

$$(Q_1 Q_2 \cdots Q_s)(R_s \cdots R_1) = (A_1 - k_1 I) \cdots (A_s - k_s I).$$

Доказва се, че при подходящ избор на изместването  $k_s$  може да се постигне извъндиагоналните елементи в последния ред на  $A_s$  да приближават нулата много бързо. Освен това другите поддиагонални елементи също могат да приближават нулата, макар и по-бавно.

Когато се прилага към плътна матрица, QR-алгоритъмът изисква голям обем изчисления, тъй като на всяка итерация са необходими около  $n^3$  flops. Ако обаче матрицата  $A_1$  е във форма на Хесенберг, то такива са и следващите матрици  $A_s$  ( $s = 2, 3, \dots$ ) и изчисленията могат така да се организират, че всяка итерация да изисква от порядъка на  $n^2$  flops.

Ускорената сходимост на QR-алгоритъма може да се постигне, без да е необходимо да се изважда изместването от диагонала на  $A_s$  и след това то да се възстановява явно. Това води до техниката на неявното изместване [4], чието главно приложение е изчисляването на формата на Шур за реална матрица, която има комплексни собствени стойности.

Изчисляването на собствени стойности на плътна матрица с QR-алгоритъма изискват около  $5n^3$  flops. За получаването на формата на Шур  $T = Q^T A Q$  са необходими около  $7n^3$  flops за намиране на  $T$  и  $6n^3$  flops за натрупване на  $Q$ .

QR-алгоритъмът е числено устойчив. Може да се докаже, че изчислената форма на Шур е винаги ортогонално подобна на матрицата  $A + E$ , където

$$\|E\|_F \leq c_1 n \varepsilon \|A\|_F,$$

а  $c_1$  е малка константа.

Точността, с която се определят собствените стойности, зависи от тяхната чувствителност към изменения в матрицата  $A$ . Ето защо, въпреки числената устойчивост на QR-алгоритъма, собствените стойности могат да се получат с големи грешки, ако техните спектрални числа на обусловеност са големи.

Когато матрицата е дефектна, вследствие на грешки от закръгляване кратните собствени стойности се получават като близки, но различни. Ето защо за всяка от тях се изчислява отделен собствен вектор (теоретически, на един Жорданов блок съответства само един собствен вектор). Получените по този начин собствени вектори са почти линейно зависими, поради което матрицата от собствените вектори ще е лошо обусловена. Това може да доведе до числена неустойчивост на алгоритмите, в които се използва обръщане на тази матрица.

Собствените стойности на матрица, получени с QR-алгоритъма, обикновено не са подредени по някакъв специален начин върху диагонала на формата на Шур. В някои случаи обаче се изисква собствените стойности да присъстват в определен ред, например най-напред да са собствените стойности с отрицателни реални части. Желаното подреждане на собствените стойности може да се постигне, като се използват по определен начин ортогонални преобразувания на подобие, които разместват две съседни собствени стойности [21].

## 7.7 Изчисляване на обобщени собствени стойности

Нека  $A$  и  $B$  са  $n \times n$  матрици. Множеството на всички матрици във вида  $A - \lambda B$ , където  $\lambda$  е произволен комплексен скалар, се нарича *матричен сноп*. Обобщената задача за собствени стойности се състои в определяне на тези стойности на  $\lambda$ , за които уравнението

$$Ax = \lambda Bx \tag{7.36}$$

има нетривиално решение. Тези стойности на  $\lambda$  се наричат собствени стойности на снопа  $A - \lambda B$  или обобщени собствени стойности. Решението  $x$  на (7.36) се нарича обобщен собствен вектор.

Обобщената задача за собствени стойности има две необичайни свойства. Ако  $A$  и  $B$  са особени матрици и съществува ненулев вектор  $z$ , такъв че  $Az = Bz = 0$ , то всяко  $\lambda$  е собствена стойност.

По-нататък се предполага, че снопът  $A - \lambda B$  е регулярен, т.е.  $\det(A - \lambda B)$  не е тъждествено равно на нула. На второ място, ако  $B$  е особена матрица, то

степенята на характеристичния полином  $\det(A - \lambda B)$  е по-малка от  $n$ , т.е. снопът  $A - \lambda B$  има по-малко от  $n$  собствени стойности. В този случай липсващите собствени стойности на  $A - \lambda B$  могат да се разглеждат като безкрайни собствени стойности.

За да се реши обобщената задача за собствени стойности, е необходимо матриците  $A$  и  $B$  да се приведат в някакви прости форми, от които лесно да се намери решението. Може да се докаже, че за дадени  $n \times n$  матрици  $A$  и  $B$  съществуват унитарни матрици  $Q$  и  $Z$ , такива че  $QAZ$  и  $Q B Z$  са горни триъгълни матрици. Това привеждане се нарича обобщена декомпозиция на Шур. Тъй като

$$QAZ\tilde{x} = \lambda Q B Z\tilde{x},$$

където  $\tilde{x} = Z^H x$ , посредством тази декомпозиция собствените стойности на  $A - \lambda B$  могат да се намерят като отношения на диагоналните елементи на  $A$  и  $B$ :

$$\lambda_j = a_{jj}/b_{jj}.$$

Привеждането на  $A$  и  $B$  в триъгълна форма се осъществява с т.н. QZ-алгоритъм [15]. За целта матрицата  $A$  се привежда предварително в горна форма на Хесенберг, а  $B$  се привежда в горна триъгълна форма. Това привеждане се осъществява с помощта на елементарни отражения и изисква около  $3n^3/2$  flops. Ако матрицата  $Z$  се натрупва (това е необходимо при намирането на собствените вектори), са необходими още  $3n^3/2$  flops. Матрицата  $Q$  не е необходима при изчисляването на собствените вектори. По-нататък се прилага аналог на QR-алгоритъма, при който с помощта на еквивалентни унитарни преобразувания  $A$  и  $B$  се привеждат от Хесенберг-триъгълна в триъгълна форма.

Цялостното прилагане на QZ-алгоритъма изисква около  $13n^3$  flops. Този алгоритъм е числено устойчив, като изчислената обобщена форма на Шур е точна за матрици  $A + E$ ,  $B + F$ , където  $\|E\| \leq \varepsilon_M \|A\|$ ,  $\|F\| \leq \varepsilon_M \|B\|$ .

В някои приложения е необходимо да се препоредят триъгълните форми на  $A$  и  $B$ , получени с QZ-алгоритъма, така че собствените стойности на  $A - \lambda B$  да се подредят по определен начин. Аналогично на препоредждането на формата на Шур препоредждането на снопа  $A - \lambda B$  може да се извърши чрез последователно прилагане на ортогонални преобразувания, които разместват две съседни собствени стойности.

## 7.8 Решаване на линейни матрични уравнения

В анализа и синтеза на системите за управление широко приложение намират линейните матрични уравнения на Ляпунов и Силвестър.

Матричното уравнение на Силвестър има вида

$$AX + XB = C,$$

където  $A$  е  $n \times n$  матрица,  $B$  е  $m \times m$  и  $C$ ,  $X$  са  $n \times m$  матрици. Това уравнение има единствено решение за  $X$ , ако и само ако матриците  $A$  и  $B$  нямат общи собствени стойности [1].

Матричното уравнение на Ляпунов  $A^T X + X A = C$ , където  $A$  е  $n \times n$  матрица, а  $C$  е  $n \times n$  симетрична матрица, може да се разглежда като частен случай на уравнението на Силвестър. То има единствено решение, ако и само собствените стойности на  $A$  удовлетворяват условието

$$\lambda_i + \lambda_j \neq 0.$$

Дискретното матрично уравнение на Ляпунов

$$A^T X A - X = C$$

има единствено решение за  $X$ , ако и само ако матрицата  $A$  няма реципрочни собствени стойности, т.е.

$$\lambda_i \lambda_j \neq 1.$$

### Обусловеност на уравненията на Ляпунов и Силвестър

Когато уравнението  $A X + X B = C$  се решава с компютър с относителна точност  $\varepsilon_M$ , в матриците  $A$ ,  $B$  и  $C$  ще има грешки от закръгляване с порядък  $\varepsilon_M \|A\|_F$ ,  $\varepsilon_M \|B\|_F$ ,  $\varepsilon_M \|C\|_F$ . Ето защо в най-добрия случай можем да очакваме, че изчисленото решение  $\bar{X}$  ще удовлетворява уравнението

$$(A + E)\bar{X} + \bar{X}(B + F) = C + G,$$

където

$$\|E\|_F \leq \varepsilon_M \|A\|_F, \|F\|_F \leq \varepsilon_M \|B\|_F, \|G\|_F \leq \varepsilon_M \|C\|_F.$$

Точността на решението ще зависи от чувствителността на  $X$  към смущения в  $A$ ,  $B$ ,  $C$ . В някои случаи малки изменения в тези матрици ще причинят големи изменения в решението, независимо от използвания метод за решаване и от свойствата на компютърната аритметика.

**Пример 7.9** Решението на уравнението на Силвестър за

$$A = \begin{bmatrix} 21.6 & -23.46 & 66.3 & -42.46 \\ 8.4 & -7.84 & 25.2 & -17.94 \\ 3.4 & -2.34 & 9.7 & -7.94 \\ 12.2 & -11.82 & 36.6 & -25.17 \end{bmatrix},$$

$$B = \begin{bmatrix} 15.2 & 30.42 \\ -7.8 & -15.61 \end{bmatrix}, C = \begin{bmatrix} 29.38 & 36.79 \\ 15.22 & 22.63 \\ 10.22 & 17.63 \\ 19.21 & 26.62 \end{bmatrix}$$

е

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Ако матрицата  $A$  е смутена с

$$E = 10^{-7} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix},$$

то решението става (с точност до петия десетичен знак)

$$\bar{X} = \begin{bmatrix} 1.0151 & 1.0294 \\ 0.98553 & 0.97180 \\ 0.98549 & 0.97171 \\ 0.99301 & 0.98637 \end{bmatrix}.$$

По такъв начин относително изменение в нормата на матрицата  $A$ ,

$$\frac{\|E\|_F}{\|A\|_F} = 3.86 \times 10^{-9}$$

води до относително изменение в решението

$$\frac{\|\bar{X} - X\|_F}{\|X\|_F} = 2.06 \times 10^{-2},$$

т.е. относителните смущения в данните се увеличават  $10^7$  пъти в решението.

Възможно е да се докаже следният резултат. Нека матриците  $A$  и  $B$  нямат общи собствени стойности,  $C \neq 0$  и

$$(\|E\|_F + \|F\|_F)/\text{sep}(A, -B) \stackrel{\text{def}}{=} \delta < 1,$$

където

$$\text{sep}(A, -B) = \min_{X \neq 0} \|AX + XB\|_F / \|X\|_F$$

е *разделянето* между матриците  $A$  и  $-B$ . Тогава относителното изменение в решението на уравнението на Силвестър, дължащо се на смущенията  $E$ ,  $F$ ,  $G$  в матриците  $A$ ,  $B$ ,  $C$ , удовлетворява неравенството

$$\frac{\|\bar{X} - X\|_F}{\|X\|_F} \leq \frac{1}{1 - \delta} \frac{\|E\|_F + \|F\|_F + \|G\|_F / \|X\|_F}{\text{sep}(A, -B)}.$$

При относителни смущения в  $A$ ,  $B$  и  $C$  с порядък  $\varepsilon$  се получава, че

$$\frac{\|\bar{X} - X\|_F}{\|X\|_F} \leq \frac{2\varepsilon}{1 - \delta} \frac{\|A\|_F + \|B\|_F}{\text{sep}(A, -B)}.$$

Ето защо величината

$$\frac{\|A\|_F + \|B\|_F}{\text{sep}(A, -B)}$$

може да се разглежда като число на обусловеност на уравнението на Силвестър. То може да е много голямо, ако  $\text{sep}(A, -B)$  е малко. За уравнението от пример 7.9

$$\text{sep}(A, -B) = 8.49 \times 10^{-7},$$

така че измененията в  $X$  могат да са около  $10^7$  пъти по-големи от смущенията в  $A$ , както бе показано в примера.

Разделянето  $\text{sep}(A, -B)$  може да се изчисли като минималното сингулярно число на матрицата

$$I_m \otimes A + B^T \otimes I_n.$$

В случая на уравнението на Ляпунов, ако  $A$  няма противоположни собствени стойности и

$$2\varepsilon \|A\|_F / \text{sep}(A^T, -A) \stackrel{\text{def}}{=} \delta < 1,$$

се получава

$$\frac{\|\bar{X} - X\|_F}{\|X\|_F} \leq \frac{4\varepsilon}{1 - \delta} \frac{\|A\|_F}{\text{sep}(A^T, -A)}.$$

Аналогичен резултат може да се докаже за дискретното уравнение на Ляпунов. Нека матрицата  $A$  няма реципрочни собствени стойности,  $C \neq 0$  и

$$(2\varepsilon + \varepsilon^2) \|A\|_F^2 / \text{sep}_d(A^T, A) \stackrel{\text{def}}{=} \delta < 1,$$

където

$$\text{sep}_d(A^T, A) = \min_{X \neq 0} \|A^T X A - X\|_F / \|X\|_F.$$

Тогава

$$\frac{\|\bar{X} - X\|_F}{\|X\|_F} \leq \frac{\varepsilon}{1 - \delta} \frac{(3 + \varepsilon) \|A\|_F^2 + 1}{\text{sep}_d(A^T, A)}.$$

Величината  $\text{sep}_d(A^T, A)$  може да се изчисли като минималното сингулярно число на матрицата

$$A^T \otimes A^T - I_{n^2}.$$

### Методи за решаване на уравненията на Ляпунов и Силвестър

Най-добри числени свойства имат методите за решаване на уравненията на Ляпунов и Силвестър, при които се използват ортогонални (или унитарни в комплексния случай) преобразувания върху матриците  $A$ ,  $B$  и  $C$ . Такъв е методът на Бартелс–Стюърт, при който матриците  $A$  и  $B$  се привеждат във форма на Шур.

Уравнението на Силвестър се решава с алгоритъма на Бартелс–Стюърт по следния начин. Матрицата  $A$  се привежда в реална долна форма на Шур  $\tilde{A}$  посредством ортогонално преобразувание с матрица  $U$ :

$$\tilde{A} = U^T A U = \begin{bmatrix} \tilde{A}_{11} & & & 0 \\ \tilde{A}_{21} & \tilde{A}_{22} & & \\ \vdots & \vdots & \ddots & \\ \tilde{A}_{p1} & \tilde{A}_{p2} & \cdots & \tilde{A}_{pp} \end{bmatrix},$$

където всяка матрица  $\tilde{A}_{ii}$  е с размерност  $1 \times 1$  или  $2 \times 2$ . По подобен начин матрицата  $B$  се привежда в реална горна форма на Шур

$$\tilde{B} = V^T B V = \begin{bmatrix} \tilde{B}_{11} & \tilde{B}_{12} & \cdots & \tilde{B}_{1q} \\ & \tilde{B}_{22} & \cdots & \tilde{B}_{2q} \\ & & \ddots & \vdots \\ 0 & & & \tilde{B}_{qq} \end{bmatrix},$$

където отново всяко  $B_{ii}$  е най-много от втори ред. Ако

$$\tilde{C} = U^T C V = \begin{bmatrix} \tilde{C}_{11} & \cdots & \tilde{C}_{1q} \\ \vdots & & \vdots \\ \tilde{C}_{p1} & \cdots & \tilde{C}_{pq} \end{bmatrix}$$

и

$$\tilde{X} = U^T X V = \begin{bmatrix} \tilde{X}_{11} & \cdots & \tilde{X}_{1q} \\ \vdots & & \vdots \\ \tilde{X}_{p1} & \cdots & \tilde{X}_{pq} \end{bmatrix},$$

то уравнението на Силвестър е еквивалентно на

$$\tilde{A}\tilde{X} + \tilde{X}\tilde{B} = \tilde{C}.$$

Ако разделянето на матриците  $\tilde{A}$ ,  $\tilde{B}$ ,  $\tilde{C}$  и  $\tilde{X}$  е извършено по съвместим начин, то

$$\tilde{A}_{kk}\tilde{X}_{kl} + \tilde{X}_{kl}\tilde{B}_{ll} = \tilde{C}_{kl} - \sum_{j=1}^{k-1} \tilde{A}_{kj}\tilde{X}_{jl} - \sum_{i=1}^{l-1} \tilde{X}_{ki}\tilde{B}_{il} \quad \begin{cases} k = 1, 2, \dots, p \\ l = 1, 2, \dots, q \end{cases} \quad (7.37)$$

Получените зависимости се решават последователно за  $\tilde{X}_{11}$ ,  $\tilde{X}_{21}$ ,  $\dots$ ,  $\tilde{X}_{p1}$ ,  $\tilde{X}_{12}$ ,  $\tilde{X}_{22}$ ,  $\dots$ ,  $\tilde{X}_{pq}$ . Решението на първоначалното уравнение тогава се дава от  $X = U\tilde{X}V^T$ .

Привеждането на  $A$  и  $B$  във форма на Шур се извършва чрез QR-алгоритъма.  $\tilde{A}$  в долна форма на Шур се получава, като  $A^T$  се привежда в горна форма на Шур и се вземе транспонираното на резултата. Решението за  $X_{kl}$  в (7.37) изисква решаването на матрично уравнение на Силвестър, в което матриците  $\tilde{A}_{kk}$  и  $\tilde{B}_{ll}$  са най-много от втори ред. Следователно то може да се получи чрез решаването на линейна система, която е най-много от четвърти ред. Ако, например,  $\tilde{A}_{kk}$  и  $\tilde{B}_{ll}$  са от втори ред, то

$$\begin{bmatrix} \tilde{a}_{11} + \tilde{b}_{11} & \tilde{a}_{12} & \tilde{b}_{21} & 0 \\ \tilde{a}_{21} & \tilde{a}_{22} + \tilde{a}_{11} & 0 & \tilde{b}_{21} \\ \tilde{b}_{12} & 0 & \tilde{a}_{11} + \tilde{b}_{22} & \tilde{a}_{12} \\ 0 & \tilde{b}_{12} & \tilde{a}_{21} & \tilde{a}_{22} + \tilde{b}_{22} \end{bmatrix} \begin{bmatrix} \tilde{x}_{11} \\ \tilde{x}_{21} \\ \tilde{x}_{12} \\ \tilde{x}_{22} \end{bmatrix} = \begin{bmatrix} \tilde{d}_{11} \\ \tilde{d}_{21} \\ \tilde{d}_{12} \\ \tilde{d}_{22} \end{bmatrix}$$

където  $\tilde{a}_{ij}$ ,  $\tilde{b}_{ij}$  и  $\tilde{x}_{ij}$  означават елементите на  $\tilde{A}_{kk}$ ,  $\tilde{B}_{ll}$  и  $\tilde{X}_{kl}$ , съответно, а  $\tilde{d}_{ij}$  означават елементите в дясната страна на (7.37). Тази система може да се реши посредством Гаусова елиминация с частичен избор на водещ елемент. След

като се изчисли  $\tilde{X}_{kl}$ , то може да запамети на мястото на  $\tilde{C}_{kl}$ , което повече не е необходимо. По такъв начин имаме следният алгоритъм.

*Алгоритъм SYLV. Решаване на уравнението на Силвестър  $AX + XB = C$*

- 1 Привежда се  $A$  в долна форма на Шур  $\tilde{A} = U^T A U$
- 2 Привежда се  $B$  в горна форма на Шур  $\tilde{B} = V^T B V$
- 3 Изчислява се  $\tilde{C} = U^T C V$
- 4 За  $l = 1, 2, \dots, q$ 

$$\tilde{C}_{kl} \leftarrow \tilde{C}_{kl} - \sum_{i=1}^{l-1} \tilde{X}_{ki} \tilde{B}_{il} \quad k = 1, 2, \dots, p$$
 За  $k = 1, 2, \dots, p$ 

$$\tilde{C}_{kl} \leftarrow \tilde{C}_{kl} - \sum_{j=1}^{k-1} \tilde{A}_{kj} \tilde{X}_{jl}$$
 Решава се  $\tilde{A}_{kk} \tilde{X}_{kl} + \tilde{X}_{kl} \tilde{B}_{ll} = \tilde{C}_{kl}$  за  $\tilde{X}_{kl}$
- 5 Изчислява се  $X = U \tilde{C} V^T$

Приемайки, че декомпозициите на Шур за  $A$  и  $B$  изискват съответно  $13n^3$  и  $13m^3$  flops, решаването на уравнението на Силвестър с алгоритъм SYLV изисква  $13(n^3 + m^3) + 5(nm^2 + mn^2)/2$  flops. Численият анализ на алгоритъма показва, че изчисленото решение удовлетворява

$$\frac{\|\tilde{X} - X\|_F}{\|X\|_F} \leq k\varepsilon \frac{\|A\|_F + \|B\|_F}{\text{sep}(A, -B)}$$

за някаква малка константа  $k$ . Това показва, че алгоритъмът на Бартелс–Стюърт е числено устойчив.

Алгоритъм SYLV може да се модифицира лесно за решаване уравнението на Ляпунов. В този случай е необходимо да се приведе във форма на Шур само матрицата  $A$ . При изчисляване на решението може да се използва симетрията на матриците  $C$  и  $X$ . Това дава възможност уравнението на Ляпунов да се реши само с  $13n^3 + 7n^3/2$  flops. И в този случай точността на решението се определя от обусловеността на решаваното уравнение. Дискретното уравнение на Ляпунов може да се реши по подобен начин.

**Пример 7.10** За илюстрация на точността, с която се решава уравнението на Ляпунов, разглеждаме семейство от уравнения, чиито матрици са избрани във вида

$$A = T A_0 T^{-1}, \quad C = T^{-T} C_0 T^{-1},$$

където  $A_0$ ,  $C_0$  са диагонални матрици, а  $T$  е неособена матрица. Решението на уравнението на Ляпунов се дава от

$$X = T^{-T} X_0 T^{-1},$$

където  $X_0$  е диагонална матрица, чиито елементи се определят лесно от диагоналните елементи на  $A_0$ ,  $C_0$ . За да се избегнат грешки от закръгляване при конструирането и оръщането на  $T$ , тази матрица е избрана като

$$T = H_2 S H_1$$



където

$$\begin{aligned} H_1 &= I_n - 2ee^T/n, e = [1, 1, \dots, 1]^T \\ H_2 &= I_n - 2ff^T/n, f = [1, -1, 1, \dots, (-1)^{n-1}]^T, \\ S &= \text{diag}(1, s, s^2, \dots, s^{n-1}), s > 1. \end{aligned}$$

Задавайки различни стойности на  $s$ , е възможно да се променя числото на обусловеност на матрицата  $T$  по отношение на обръщането,

$$\text{cond}_2(T) = s^{n-1}.$$

Вземайки предвид формата на  $T$  получаваме, че

$$\begin{aligned} A &= H_2 S H_1 A_0 H_1 S^{-1} H_2, \\ C &= H_2 S^{-1} H_1 C_0 H_1 S^{-1} H_2. \end{aligned}$$

Тези матрици се изчисляват лесно с относителна точност  $\varepsilon$ .

В случая се разглежда решението на уравнение на Ляпунов от 18-ти ред, за което

$$A_0 = \text{diag}(A_1, A_1, A_1, A_1, A_1, A_1),$$

$$C_0 = \text{diag}(C_1, C_1, C_1, C_1, C_1, C_1)$$

и

$$\begin{aligned} A_1 &= \text{diag}(-1 \times 10^{-k}, -2, -3 \times 10^k), \\ C_1 &= \text{diag}(-3 \times 10^{-k}, -5, -7 \times 10^k). \end{aligned}$$

Точното решение се дава от

$$X_0 = \text{diag}(X_1, X_1, X_1, X_1, X_1, X_1),$$

където

$$X_1 = \text{diag}(3/2, 5/4, 7/6).$$

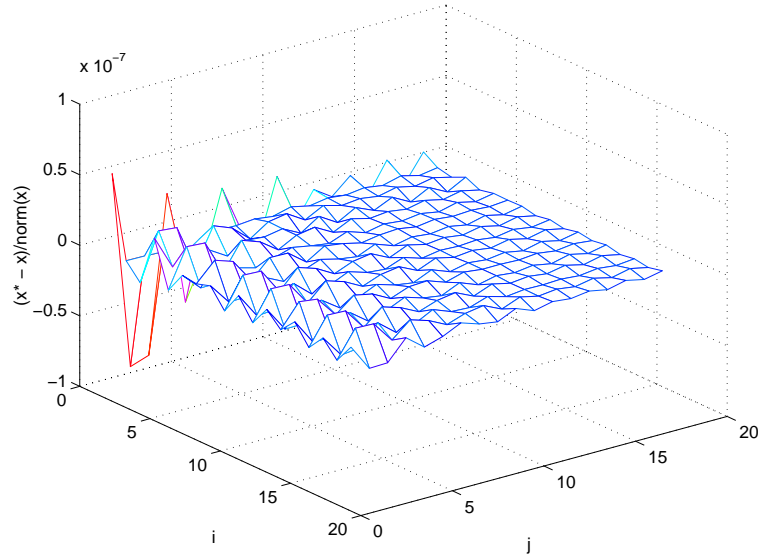
На фиг. 7.8 са показани покомпонентните относителни грешки в изчисленото решение за  $k = 4$ ,  $s = 1.3$  в аритметика с относителна точност  $\varepsilon = 2.22 \times 10^{-16}$ . При тези стойности на параметрите  $k$  и  $s$  уравнението на Ляпунов има число на обусловеност равно на  $4.49 \times 10^{10}$  и относителната грешка в изчисленото решение

$$\frac{\|\bar{X} - X\|_F}{\|X\|_F} = 1.52 \times 10^{-7}$$

е близка до предвидената от числения анализ грешка

$$\text{cond}\varepsilon = 9.96 \times 10^{-6}.$$

Изчисленията са извършени с функцията `lyap` в MATLAB.



Фигура 7.8: Покомпонентни грешки в решението на уравнение на Ляпунов от 18-ти ред

## 7.9 Привеждане в ортогонална канонична форма

Привеждането на двойката  $(A, B)$  в ортогонална канонична форма може да се извърши използвайки декомпозицията по сингулярни числа или, по-ефективно, използвайки QR декомпозицията с водеща колона.

Нека матрицата  $B$  е декомпозирана като

$$B = Q_1 \begin{bmatrix} R_1 \\ 0 \end{bmatrix} P_1,$$

където  $Q_1$  е ортогонална,  $R_1$  е  $m_1 \times m$  горна трапецоидална и  $P_1$  е пермутационна матрица ( $P_1^{-1} = P_1^T$ ), отразяваща преместването на колоните по време на декомпозицията. Първият спрегнат индекс на Кронекер  $m_1$  се определя като ранга на  $R_1$  (и следователно на  $B$ ). Чрез ортогонално преобразуване на подобие с матрицата  $Q_1$  се получава

$$A_2 = Q_1^T A Q_1 = \begin{bmatrix} A_{11}^{(2)} & A_{12}^{(2)} \\ A_{21}^{(2)} & A_{22}^{(2)} \end{bmatrix}, \quad B^0 = Q_1^T B = \begin{bmatrix} B_{10} \\ 0 \end{bmatrix},$$

където  $A_{11}^{(2)} \in R^{m_1 \times m_1}$  и  $B_{10} = R_1 P_1$ .

На втората стъпка на привеждането матрицата  $A_{21}^{(2)}$  се декомпозира в

$$A_{21}^{(2)} = Q_2' \begin{bmatrix} R_2 \\ 0 \end{bmatrix} P_2,$$

където  $R_2$  е  $m_2 \times m_1$  горна трапецоидална матрица и  $m_2$  се определя от

$$m_2 = \text{rank}(R_2) = \text{rank} A_{21}^{(2)} \leq m_1.$$

На следващата стъпка се извършва ортогонално преобразуване с матрица

$$Q_2 = \text{diag}(I_{m_1}, Q_2'),$$

след което процедурата продължава по подобен начин до окончателното преобразуване на системата. В резултат се получава  $(A^0, B^0) = (U^T A U, U^T B)$ , където

$$A^0 = \begin{bmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1,p-1} & A_{1p} \\ A_{21} & A_{22} & A_{23} & \cdots & A_{2,p-1} & A_{2p} \\ 0 & A_{32} & A_{33} & \cdots & A_{3,p-1} & A_{3p} \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & A_{p,p-1} & A_{pp} \end{bmatrix},$$

$$B^0 = \begin{bmatrix} B_{10} \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

матриците

$$B_{10} \in \mathbb{R}^{m_1 \times m}; \quad A_{i,i-1} \in \mathbb{R}^{m_i \times m_{i-1}}; \quad i = 2, \dots, p$$

имат пълен редови ранг и  $U$  е произведението на натрупаните ортогонални преобразувания.

Привеждането на системата в ортогонална канонична форма се извършва с алгоритъм DREMCF, който презаписва  $A$  и  $B$  с  $A^0$  и  $B^0$ , съответно, и натрупва ортогоналните преобразувания на подобие в матрицата  $U$ . Блоковете  $A_{i,i-1}$  се получават като горни трапецоидални матрици с пермутирани колони. Алгоритъмът произвежда също така цяло число  $d$ , равно на размерността на управляемото подпространство на системата, цяло число  $p$ , равно на индекса на управляемост на системата и спрегнатите индекси на Кронекер  $m_1, m_2, \dots, m_p$ . За запамятаване на матриците, които трябва да се декомпонират, се използва  $n \times m$  спомагателен масив  $W$ .

*Алгоритъм DREMCF. Привеждане на многовходова система в ортогонална канонична форма*

- 1 Полага се  $k = 0$ ,  $d = 0$ ,  $nb = n$ ,  $mb = m$ ,  $p = 0$
- 2 Полага се  $U = I_n$
- 3  $W \leftarrow B$ ,  $B \leftarrow 0$
- 4  $k \leftarrow k + 1$
- 5 Изчислява се QR декомпозицията с избор на водеща колона на  $nb \times mb$  матрицата  $W$ ,  

$$Q'_k \begin{bmatrix} R_k \\ 0 \end{bmatrix} = WP_k^T$$
- 6 Определя се числения ранг  $irnk$  на  $R_k$  използвайки допуск  $tol$
- 7 Ако  $irnk = 0$ , край на изчисленията
- 8 Полага се  $ni = d$
- 9  $d \leftarrow d + irnk$ ,  $p \leftarrow p + 1$
- 10  $m_p = irnk$
- 11  $A \leftarrow Q_k^T A Q_k$ ,  $Q_k = \text{diag}(I_{n-nb}, Q'_k)$
- 12  $U \leftarrow U Q_k$
- 13 Ако  $k = 1$   

$$B \leftarrow \begin{bmatrix} R_1 P_1 \\ 0 \end{bmatrix}$$
- 14 Ако  $irnk = nb$ , край на изчисленията
- 15  $mb \leftarrow irnk$ ,  $nb \leftarrow nb - irnk$
- 16 За  $i = 1, 2, \dots, nb$   

$$\begin{aligned} w_{ij} &\leftarrow a_{d+i, ni+j} & j = 1, 2, \dots, mb \\ a_{d+i, ni+j} &\leftarrow 0 & j = 1, 2, \dots, mb \end{aligned}$$

Този алгоритъм изисква около  $\frac{8}{3}n^3 + mn^2$  операции с плаваща запетая. Може да се докаже, че алгоритъм DREMCF е числено устойчив, като изчислената двойка  $(A^0, B_2^0)$  е точна за леко смутените данни  $(A + \Delta A, B_2 + \Delta B_2)$ , където

$$\|\Delta A\|_F \leq (6n^2 + 60n)\varepsilon_M \|A\|_F,$$

$$\|\Delta B_2\|_F \leq (6nm - 3m^2 + 30m)\varepsilon_M \|B_2\|_F.$$

## 7.10 Синтез при зададена собствена структура

В тази точка се разглежда числен алгоритъм за синтез при зададена собствена структура, който може да се използва и за синтез при зададени полюси.

Нека е дадена многовходовата система

$$\dot{x}(t) = Ax(t) + Bu(t).$$

Търсим закон на управление

$$u(t) = -Kx(t),$$

за който затворената система

$$\dot{x}(t) = (A - BK)x(t)$$

има желаното множество полюси  $\lambda$ .

Предварителната стъпка на алгоритъма е привеждането на двойката  $(A, B)$  в ортогонална канонична форма  $(A^0, B^0) = (U^T A U, U^T B)$ .

Втората стъпка на алгоритъма е да се определи матрицата на обратната връзка на преобразуваната система. Тъй като преобразуваната матрица на затворената система  $A^0 - B^0 K^0$ ,  $K^0 = KU$ , е в блочна форма на Хесенберг, то може да се намери собствен вектор  $v^0$  на тази матрица, познавайки само  $A^0$  и желаната собствена стойност  $\lambda^0 \in \lambda$ . Действително, това следва от уравнението

$$(A^0 - B^0 K^0)v^0 = \lambda^0 v^0 \quad (7.38)$$

където

$$v^0 = [v_1 \ \dots \ v_p]^T; \ v_i \in \mathbb{C}^{m_i}$$

и

$$A_{i,i-1}v_{i-1} = v_i\lambda^0 - \sum_{k=i}^p A_{ik}v_k; \ i = p, \dots, 2, \quad (7.39)$$

така че посредством задаване на  $v_p$  е възможно да се изчислят рекурсивно  $v_{p-1}, \dots, v_1$ . Елементите на  $v_p$  са свободни (с изключение, че поне един от тях трябва да е ненулев, така че  $v_p \neq 0$ ) и различното избиране на тези елементи ще води до различни решения  $K^0$ .

Освен свободата в  $v_p$  съществува допълнителна свобода при решаването на уравнението (7.39), ако  $m_i < m_{i-1}$ .

Да предположим сега, че е избрана последователност от равнинни ротации в съответните равнини, така че да се нулират последователно всички елементи на  $v^0$ , с изключение на първия, движейки се от долу на горе. В резултат получаваме

$$\begin{aligned} Q^H(A^0 - B^0 K^0)Qv^1 &= v^1\lambda^0, \\ v^1 = Q^H v^0 &= [v_{10} \ 0 \ \dots \ 0]^T, \\ &v_{10} \neq 0 \end{aligned} \quad (7.40)$$

където  $Q$  е ортогонална матрица, произведение на приложените равнинните ротации и  $(n-1) \times (n-1)$  блокът в долния десен ъгъл на  $Q^H A^0 Q$  е отново в блочна форма на Хесенберг. Този блок има същата структура както  $A^0$  с изключение на това, че индексът  $m_p$  е намалял с 1.

За да се илюстрира преобразуването на собствения вектор, нека разгледаме система с  $n = 6$ ,  $m_1 = m_2 = m_3 = 2$ . За тази система уравнението (7.38) изглежда по следния начин:

$$\begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \end{bmatrix} \begin{bmatrix} \times \\ \times \\ \times \\ \times \\ \times \\ \times \end{bmatrix} = \begin{bmatrix} \times \\ \times \\ \times \\ \times \\ \times \\ \times \end{bmatrix} \lambda^0$$

където неопределените елементи са означени с  $\times$ .

Изчислявайки третия и четвъртия елементи на собствения вектор и прилагайки ротации в равнините (5,6) и (4,5) за да се анулират 6-тия и 5-тия елементи на собствения вектор, съответно, получаваме

$$\begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ + & + & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \end{bmatrix} \begin{bmatrix} \times \\ \times \\ \times \\ \times \\ \otimes \\ 0 \end{bmatrix} = \begin{bmatrix} \times \\ \times \\ \times \\ \times \\ \otimes \\ 0 \end{bmatrix} \lambda^0$$

където ротацията в равнината (4,5) анулира елемента, означен с  $\otimes$  и въвежда ненулевите елементи, означени с  $+$ . Отбележете, че по необходимост четвъртият елемент на преобразувания собствен вектор ще бъде ненулев.

След прилагането на ротациите  $R_{3,4}$  и  $R_{2,3}$  се получава

$$\begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & 0 & \otimes & \times & \times & \times \end{bmatrix} \begin{bmatrix} \times \\ \times \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \times \\ \times \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \lambda^0$$

където елементът, означен с  $\otimes$ , се анулира след ротацията в равнината (3,4).

Последната ротация в равнината (1,2) дава

$$\begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \otimes & \times & \times & \times & \times & \times \\ \otimes & \times & \times & \times & \times & \times \\ \otimes & \times & \times & \times & \times & \times \\ \otimes & \times & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \end{bmatrix} \begin{bmatrix} \times \\ \otimes \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \times \\ \otimes \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \lambda^0$$

анулирайки едновременно всички елементи, означени с  $\otimes$ .  $5 \times 5$  матрицата, разположена в долния десен ъгъл на преобразуваната матрица на затворената система е в блочна форма на Хесенберг с индекси  $m_1 = m_2 = 2$ ,  $m_3 = 1$ .

От (7.40) се получава линейно уравнение за първата колона на  $K^0 Q$  и полюсът  $\lambda^0$  се изчерпва от задачата.

Възможно е на следващата стъпка да се процедира по същия начин, работейки с подсистема от ред  $n - 1$ . По такъв начин всеки полюс се задава независимо, което позволява задаването на кратни полюси.

По този начин чрез задаване на свободните елементи на собствените вектори на подсистемата, е възможно на всяка стъпка да се получи определено решение за съответната колона на матрицата на обратната връзка.

Определянето на матрицата на обратната връзка  $K$ , която съответства на зададени елементи на собствените вектори и матрицата на обратната връзка, се

извършва със следния алгоритъм, който използва матриците  $A^0, B^0$  на ортогоналната канонична форма, както и матрицата на ортогоналното преобразуване  $U$ .

*Алгоритъм DEIGEN. Синтез при зададена собствена структура на многомерна система чрез ортогонално изчерпване.*

- 1 Задава се  $\ell = 0$ ,  $Q = U$
- 2  $\ell \leftarrow \ell + 1$
- 3 Ако  $p = 1$  се отива на 9
- 4 Полага с  $v_p$
- 5 За  $i = p, p - 1, \dots, 1$ 
  - 5.1 Ако  $i > 1$  се изчислява  $v_{i-1}$   
задавайки свободните елементи
  - 5.2 Определя се произведението  $Q_i$  на равнинните  
ротации, преобразуващи  $v_i$
  - 5.3  $A^0 \leftarrow Q_i^H A^0 Q_i$
  - 5.4 Ако  $i = 1$ ,  $B^0 \leftarrow Q_i^H B^0$
  - 5.5  $Q \leftarrow Q Q_i$
- 6 Определят се  $\ell$ -тите колони на  $K^0$  и  
горната триъгълна форма на  $Q^H(A - BK)Q$   
задавайки свободните елементи
- 7  $m_p \leftarrow m_p - 1$
- 8 Ако  $m_p = 0$ ,  $p \leftarrow p - 1$   
Отива се на 2
- 9 Определят се последните (от  $\ell$ -тата до  $n$ -тата)  
колони на  $K^0$  и горната триъгълна  
форма на матрицата  $Q^H(A^0 - B^0 K^0)Q$   
задавайки свободните елементи
- 10  $K = K^0 Q^H$

В случая на реални системи горният алгоритъм може да се изведе като реална алгоритмична процедура за реални и комплексно спрегнати полюси.

Числените свойства на алгоритъма *DEIGEN* са много благоприятни. Може да се покаже, че той е обратно устойчив, т.е. изчислената матрица на обратната връзка е точна за матрици  $A + \Delta A$ ,  $B + \Delta B$ , където  $\|\Delta A\|$ ,  $\|\Delta B\|$  са малки относно  $\|A\|$ ,  $\|B\|$ , съответно.

## 7.11 Решаване на матричното уравнение на Рикати

Нека разгледаме непрекъснатото матрично уравнение на Рикати

$$A^T X + X A + C - X D X = 0, \quad (7.41)$$

където  $A$ ,  $C$ ,  $D$  са квадратни  $n \times n$  матрици и матриците  $C$ ,  $D$  са симетрични. Предполагаме, че  $C$  и  $D$  са положително полуопределени матрици, като двойката  $(A, D)$  е стабилизируема, а двойката  $(C, A)$  – детектируема. При тези условия съществува единствено положително полуопределено решение  $X$  на (7.41) и матрицата  $A - DX$  е устойчива.

### Обусловеност на уравнението на Рикати

Когато уравнение (7.41) се решава с компютър, изчисленият резултат ще удовлетворява смутеното уравнение

$$(A + E)^T \bar{X} + \bar{X}(A + E) + C + F - \bar{X}(D + G)\bar{X} = 0, \quad (7.42)$$

където  $E$ ,  $F$ ,  $G$  са матрици на смущенията. Ако се използва числено устойчив метод за решаване на (7.41), може да се очаква, че в най-добрия случай тези матрици удовлетворяват условията

$$\|E\| \leq \varepsilon_M \|A\|, \quad \|F\| \leq \varepsilon_M \|C\|, \quad \|G\| \leq \varepsilon_M \|D\|.$$

В някои случаи уравнението на Рикати е много чувствително към смущения в данните, т.е. малки смущения  $E$ ,  $F$ ,  $G$  причиняват големи изменения в решението. В такива случаи се казва, че уравнението на Рикати е лошо обусловено.

Чувствителността на уравнението на Рикати (7.41) може да се оцени по следния начин. Нека  $\Delta X = \bar{X} - X$ . Изваждайки (7.41) от (7.42) и пренебрегвайки членовете от ред по-висок от първи по отношение на  $\Delta X$ , получаваме

$$(A - DX)^T \Delta X + \Delta X(A - DX) + E^T X + XE + F - XGX = 0. \quad (7.43)$$

Уравнението (7.43) е уравнение на Ляпунов по отношение на смущението  $\Delta X$ .

Нека вместо матриците  $\Delta X$ ,  $E$ ,  $F$ ,  $G$  въведем техните векторни представления  $\text{Vec}(\Delta X)$ ,  $\text{Vec}(E)$ ,  $\text{Vec}(F)$ ,  $\text{Vec}(G)$ , които се получават чрез "наставяне" на колоните на съответната матрица. Тогава уравнението (7.43) може да се запише във векторния вид

$$(A_C^T \otimes I_n + I_n \otimes A_C^T) \text{Vec}(\Delta X) = -\text{Vec}(F) - (I_n \otimes X + (X \otimes I_n)W) \text{Vec}(E) + (X \otimes X) \text{Vec}(G), \quad (7.44)$$

където са използвани представянията

$$\begin{aligned} \text{Vec}(E^T) &= W \text{Vec}(E), \\ \text{Vec}(MZN) &= (N^T \otimes M) \text{Vec}(Z) \end{aligned}$$

и  $W$  е т.н. векторно-пермутационна матрица.

Решението на уравнение (7.44) се записва във вида

$$\text{Vec}(\Delta X) = -\Omega^{-1} \text{Vec}(F) - \Theta \text{Vec}(E) + \Pi \text{Vec}(G), \quad (7.45)$$



където

$$\begin{aligned}\Omega^{-1} &= (A_C^T \otimes I_n + I_n \otimes A_C^T)^{-1}, \\ \Theta &= (A_C^T \otimes I_n + I_n \otimes A_C^T)^{-1}(I_n \otimes X + (X \otimes I_n)W), \\ \Pi &= (A_C^T \otimes I_n + I_n \otimes A_C^T)^{-1}(X \otimes X).\end{aligned}\tag{7.46}$$

Вземайки 2-нормите на двете страни в (7.45), получаваме

$$\|\text{Vec}(\Delta X)\|_2 \leq \|\Omega^{-1}\|_2 \|\text{Vec}(F)\|_2 + \|\Theta\|_2 \|\text{Vec}(E)\|_2 + \|\Pi\|_2 \|\text{Vec}(G)\|_2,$$

което е еквивалентно на

$$\|\Delta X\|_F \leq \|\Omega^{-1}\|_2 \|F\|_F + \|\Theta\|_2 \|E\|_F + \|\Pi\|_2 \|G\|_F,\tag{7.47}$$

тъй като

$$\|\text{Vec}(M)\|_2 = \|M\|_F.$$

Уравнение (7.47) показва, че чувствителността на решението по отношение на смущения в матриците  $A$ ,  $C$ ,  $D$  се определя съответно от величините  $\|\Theta\|_2$ ,  $\|\Omega^{-1}\|_2$ ,  $\|\Pi\|_2$ . Може да се отбележи, че величината

$$\|\Omega^{-1}\|_2 = \|(A_C^T \otimes I_n + I_n \otimes A_C^T)^{-1}\|_2 = \frac{1}{\text{sep}(A_C^T, -A_C)}$$

характеризира чувствителността на уравнението на Ляпунов

$$A_C^T X + X A_C = -C.$$

От уравнение (7.47) получаваме, че ако смущенията  $E$ ,  $F$ ,  $G$  удовлетворяват условията

$$\|E\|_F \leq \delta \|A\|_F, \quad \|F\|_F \leq \delta \|C\|_F, \quad \|G\|_F \leq \delta \|D\|_F,$$

то

$$\frac{\|\Delta X\|_F}{\|X\|_F} \leq K_B \delta,$$

където числото

$$K_B = \frac{\|\Omega^{-1}\|_2 \|C\|_F + \|\Theta\|_2 \|A\|_F + \|\Pi\|_2 \|D\|_F}{\|X\|_F}$$

може да се разглежда като число на обусловеност на уравнението на Рикати.

Уравненията (7.46) позволяват да се получат точните стойности на  $\|\Omega^{-1}\|_2$ ,  $\|\Theta\|_2$  и  $\|\Pi\|_2$  за сметка на конструирането на  $n^2 \times n^2$  матрици, което не е ефективно при големи  $n$ .

**Пример 7.11** Нека разгледаме семейство от уравнения на Рикати, чиито матрици са определени от

$$A = T A_0 T^{-1}, \quad C = T^{-T} C_0 T^{-1}, \quad D = T D_0 T^T,$$

където  $A_0$ ,  $C_0$ ,  $D_0$  са диагонални матрици, а  $T$  е неособена матрица, която се избира по същия начин, както в пример 7.10. Решението на уравнението на Рикати се дава от

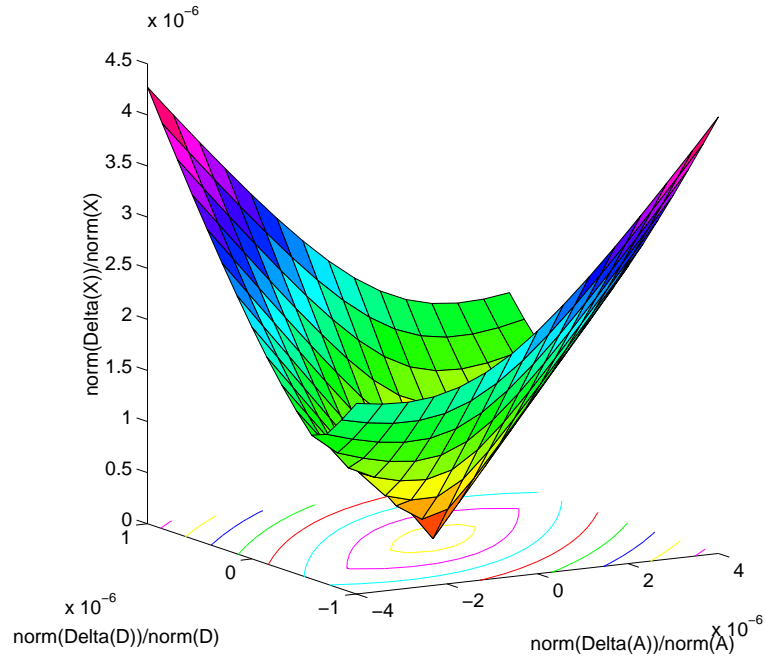
$$X = T^{-T} X_0 T^{-1}.$$

В случая се разглежда уравнение на Рикати от шести ред, за което

$$\begin{aligned} A_0 &= \text{diag}(A_1, A_1), \\ C_0 &= \text{diag}(C_1, C_1), \\ D_0 &= \text{diag}(D_1, D_1), \\ X_0 &= \text{diag}(X_1, X_1) \end{aligned}$$

и

$$\begin{aligned} A_1 &= \text{diag}(1 \times 10^k, 2 \times 10^k, 3 \times 10^k), \\ C_1 &= \text{diag}(3 \times 10^k, 5 \times 10^k, 7 \times 10^k), \\ D_1 &= \text{diag}(10^k, 10^k, 10^k), \\ X_1 &= \text{diag}(3, 5, 7). \end{aligned}$$



Фигура 7.9: Смутени решения на уравнение на Рикати

Това уравнение на Рикати е добре обусловено. На фиг. 7.9 са показани относителните изменения в 2-нормите на изчислените решения при  $k = 1.1$ ,  $s = 1.2$ , получени за смутените матрици

$$A + E, \quad E = x\hat{E}, \quad x \in [-4 \times 10^{-6}, 4 \times 10^{-6}] \times \|A\|_2,$$

$$D + G, \quad G = y\hat{G}, \quad y \in [-10^{-6}, 10^{-6}] \times \|D\|_2,$$

където  $\hat{E}$ ,  $\hat{G}$  са генерирани като случайни матрици с единична 2-норма. От фигурата се вижда, че относителните изменения в решението имат порядъка на относителните изменения в данните, което се дължи на ниското число на обусловеност

$$K_B = 6.15.$$

### Методи за решаване на уравнението на Рикати

Уравнението на Рикати (7.41) е тясно свързано с Хамилтоновата матрица

$$H = \begin{bmatrix} A & -D \\ -C & -A^T \end{bmatrix}.$$

Тъй като тази матрица може да се представи във вида

$$H = \begin{bmatrix} I_n & 0 \\ X & I_n \end{bmatrix} \begin{bmatrix} A - DX & D \\ 0 & -(A - DX)^T \end{bmatrix} \begin{bmatrix} I_n & 0 \\ -X & I_n \end{bmatrix},$$

то следва, че  $H$  има точно  $n$  собствени стойности с отрицателни реални части (собствените стойности на  $A - DX$ ) и  $n$  собствени стойности с положителни реални части (собствените стойности на  $-(A - DX)^T$ ).

Най-широко приложение за решаване на уравнението на Рикати е намерил методът на Шур, при който Хамилтоновата матрица се привежда в реална горна форма на Шур посредством ортогонални преобразувания на подобие.

Съгласно декомпозицията на Шур съществува ортогонална матрица  $U$ , която преобразува Хамилтоновата матрица  $H$  във форма на Шур

$$T = U^T H U = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}, \quad (7.48)$$

където  $T_{11}$  и  $T_{22}$  са  $n \times n$  горни квазитриъгълни матрици.

Привеждането (7.48) не е единствено и винаги е възможно матрицата  $U$  да се избере так, че собствените стойности на  $T_{11}$  да имат отрицателни реални части, а тези на  $T_{22}$  – положителни.

Нека матрицата  $U$  е разделена като

$$U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix},$$

където всеки блок има размерност  $n \times n$ . Тогава може да се покаже, че  $U_{11}$  е неособена матрица и положително полуопределеното решение на (7.41) се намира от

$$X = U_{21} U_{11}^{-1}.$$

Освен това собствените стойности на  $T_{11}$  са собствените стойности на матрицата  $A - DX$ , т.е. полюсите на оптималната затворена система при синтеза на  $H_2$  оптимален регулатор на състоянието.

Ортогоналното привеждане на Хамилтоновата матрица във форма на Шур се извършва с QR-метода. Тъй като матрицата  $H$  е несиметрична, то не може да се гарантира специално подреждане на собствените стойности върху диагонала на  $T$ . Тези диагонални елементи могат обаче да се препоредят по произволен начин, като се разместват последователно съседни собствени стойности посредством ортогонални (или унитарни при комплексни собствени стойности) преобразувания на подобие.

По такъв начин се получава следния алгоритъм.

*Алгоритъм RICCATI. Решаване на уравнението на Рикати  $A^T X + X A + C - X D X = 0$*

1 Полага се

$$H = \begin{bmatrix} A & -D \\ -C & -A^T \end{bmatrix}$$

2 Привежда се  $H$  във форма на Шур  $T = U^T H U$  посредством ортогонални преобразувания на подобие

3 Препоредва се формата на Шур, така че блокът  $T_{11}$  в левия горен ъгъл на  $T$  да има собствени стойности с отрицателни реални части.

4 Решава се  $U_{11}^T X = U_{21}^T$ , където  $[U_{11}^T, U_{21}^T]^T$  са векторите на Шур, съответстващи на  $T_{11}$ .

Алгоритъм RICCATI изисква около  $75n^3$  flops.

**Пример 7.12** Разглежда се уравнение на Рикати от 15-ти ред, чийто матрици са определени както в пример 7.11 за

$$\begin{aligned} A_0 &= \text{diag}(A_1, A_1, A_1, A_1, A_1), \\ C_0 &= \text{diag}(C_1, C_1, C_1, C_1, C_1), \\ D_0 &= \text{diag}(D_1, D_1, D_1, D_1, D_1). \end{aligned}$$

където

$$\begin{aligned} A_1 &= \text{diag}(-1 \times 10^{-k}, -2, -3 \times 10^k), \\ C_1 &= \text{diag}(3 \times 10^{-k}, 5, 7 \times 10^k), \\ D_1 &= \text{diag}(10^{-k}, 1, 10^k). \end{aligned}$$

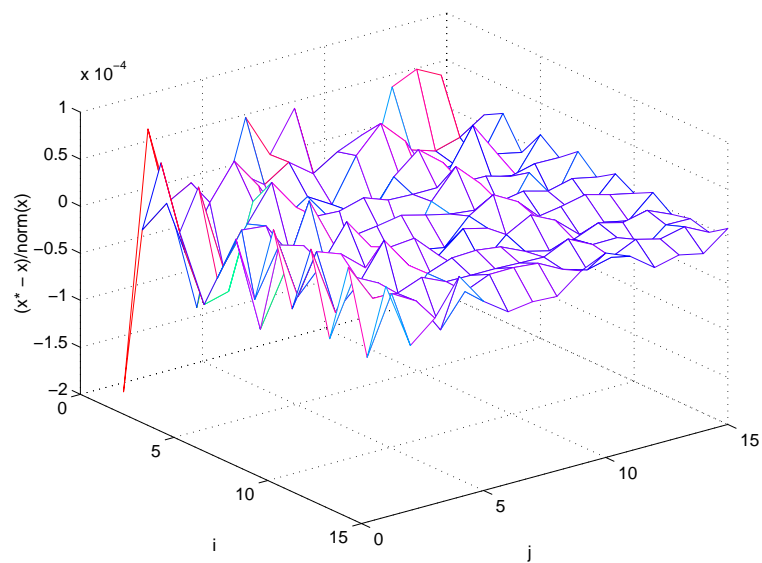
Точното решение на уравнението се дава от

$$X_0 = \text{diag}(X_1, X_1, X_1, X_1, X_1).$$

където

$$X_1 = \text{diag}(1, 1, 1).$$

Това уравнение е лошо обусловено, като при  $k = 6$ ,  $s = 1.2$  числото на обусловеност е  $K_B = 3.33 \times 10^{13}$ . Следователно, при машинна точност  $\varepsilon_M = 2.22 \times 10^{-16}$  може да се очаква, че в изчисленото решение ще има около три верни десетични знака. На фиг. 7.10 са показани покомпонентните грешки в решението,



Фигура 7.10: Покомпонентни грешки в решението на уравнение на Рикати от 15-ти ред

получено с алгоритъма RICCATI. Относителната грешка в решението

$$\frac{\|\bar{X} - X\|_F}{\|X\|_F} = 3.21 \times 10^{-4}$$

е близка до предвидената от анализа на чувствителността.



# Приложение

## МАТРИЧЕН АНАЛИЗ

В това приложение се дават някои основни сведения от матричния анализ, които се използват в книгата. Основните определения, отнасящи се до множества и реални и комплексни вектори и матрици, се привеждат в т. П.1 и П.2. В т. П.3 са дадени някои основни сведения за линейните векторни подпространства. В т. П.4 са разгледани редица свойства на собствените стойности и собствените вектори на матрици. Декомпозицията по сингулярни числа, която играе важна роля в матричния анализ и теорията на многомерните системи, се разглежда в т. П.5. В т. П.6 се описват най-важните векторни и матрични норми. И накрая, в т. П.7 са изложени някои основни факти за линейните дробни преобразувания.

Задълбочено изложение на разглежданите в това приложение въпроси може да се намери в [1, 2, 6, 15].

### П.1 Множества

Казваме, че множество  $A$  е всяка добре дефинирана съвкупност от обекти. Тези обекти се наричат *членове* или *елементи* на  $A$ .

Множество, не съдържащо нито един елемент, се нарича *празно множество* и се означава с  $\emptyset$ .

Множеството на всички реални (комплексни) числа се означава с  $\mathbb{R}(\mathbb{C})$ .

Нека  $A$  е произволно множество. Записът  $a \in A$  означава, че  $a$  е елемент на  $A$ . Ако  $a$  не е елемент на  $A$ , записваме  $a \notin A$ . Множество, чиито елементи са  $a_1, a_2, \dots, a_n$  се означава с  $\{a_1, a_2, \dots, a_n\}$ . Множеството от всички елементи  $a$ , такива че предложението  $P(a)$  е истинно, се означава с

$$A = \{a : P(a)\}.$$

За произволни множества  $A$  и  $B$ ,  $A$  се нарича *подмножество* на  $B$  (записано като  $A \subset B$ ), ако и само ако за всички  $x$ ,  $x \in A$  означава  $x \in B$ . (Това определение очевидно включва случая  $A = B$ .) Самото  $B$  и празното множество  $\emptyset$  са винаги подмножества на  $B$ . Множеството  $A$  е *собствено подмножество* на множеството  $B$ , ако  $A \subset B$  и  $A \neq B$ .

Нека  $A$  и  $B$  са произволни множества; тяхната *сума* или *обединение*  $C = A \cup B$  се нарича множеството, състоящо се от всички елементи, принадлежащи поне на едното от множествата  $A$  и  $B$ .

*Сечение*  $C = A \cap B$  на множествата  $A$  и  $B$  се нарича множеството, състоящо се от всички елементи, принадлежащи както на  $A$ , така и на  $B$ .

*Разлика*  $A \setminus B$  на множествата  $A$  и  $B$  се нарича съвкупността на тези елементи на  $A$ , които не се съдържат в  $B$ .

Ако  $A$  е подмножество на  $B$ , то *допълнението* на  $A$ , означавано с  $A'$ , е множеството, образувано от всички елементи на  $B$ , които не принадлежат на  $A$ , т.е.  $A' = \{x \in B : x \notin A\} = B \setminus A$ .

Нека  $A$  е множество от реални числа, ограничено отгоре. Да предположим, че съществува число  $x$ , притежаващо следните свойства:

- (а)  $x$  е горна граница на  $A$ ,
- (б) ако  $y$  е горна граница на  $A$ , то  $x \leq y$ .

Тогава  $x$  се нарича най-малка горна граница или *супремум* на множеството  $A$  и се означава със  $\sup(A)$ .

Най-голямата долна граница или *инфимум* на кое да е множество  $A$ , което е ограничено отдолу, се дефинира по подобен начин и се означава с  $\inf(A)$ .

В общият случай супремумът или инфимумът на множество могат да принадлежат или могат да не принадлежат на множеството.

Ако  $A$  е непразно множество от реални числа, което е ограничено отгоре (отдолу), то супремумът (инфимумът) на  $A$  съществува.

## П.2 Вектори и матрици

Комплексен *вектор-колона*  $a$  с  $m$  компоненти се записва като

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix},$$

където  $a_i$  е комплексен скалар. Векторът-ред  $a^T$  означава транспонираният на вектора  $a$ .

Да разгледаме комплексна  $l \times m$  матрица  $A$  с елементи  $a_{ij} = \operatorname{Re} a_{ij} + j \operatorname{Im} a_{ij}$ .  $l$  е броят на редовете (броят на "изходите", когато  $A$  се разглежда като *оператор*), а  $m$  е броят на колоните ("вховете"). Математически, записваме  $A \in \mathbb{C}^{l \times m}$  ако  $A$  е комплексна матрица или  $A \in \mathbb{R}^{l \times m}$  ако  $A$  е реална матрица. Отбележете, че вектор-колоната  $a$  с  $m$  елементи може да се разглежда като  $m \times 1$  матрица.

Транспонираната на матрицата  $A$  се означава с  $A^T$  (с елементи  $a_{ji}$ ), спрегнатата на  $A$  – с  $\bar{A}$  (с елементи  $\operatorname{Re} a_{ij} - j \operatorname{Im} a_{ij}$ ), спрегнато-транспонираната (Ермитово спрегнатата) на  $A$  – с  $A^H \stackrel{\text{def}}{=} \bar{A}^T$  (с елементи  $\operatorname{Re} a_{ji} - j \operatorname{Im} a_{ji}$ ), следата (сумата от диагоналните елементи) на  $A$  – с  $\operatorname{tr} A$  и детерминантата на  $A$  – с  $\det A$ . Квадратната матрица  $A$  е симетрична, ако  $A^T = A$  и Ермитова, ако



$A^H = A$ . Една Ермитова матрица е положително определена, ако  $x^H A x > 0$  за произволен ненулев вектор  $x$ .

Ако  $A \in \mathbb{R}^{p \times q}$  и  $B \in \mathbb{R}^{r \times s}$ , то *Кронекеровото произведение* на  $A$  и  $B$  се дефинира като

$$A \otimes B \stackrel{\text{def}}{=} \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1q}B \\ a_{12}B & a_{22}B & \cdots & a_{2q}B \\ \vdots & \vdots & & \vdots \\ a_{p1}B & a_{p2}B & \cdots & a_{pq}B \end{bmatrix} \in \mathbb{R}^{pr \times qs}.$$

## П.3 Линейни подпространства

Нека  $\mathbb{R}$  означава полето на реалните числа, а  $\mathbb{C}$  – полето на комплексните числа. Нека  $\mathbb{F}$  да е или  $\mathbb{R}$ , или  $\mathbb{C}$  и нека  $\mathbb{F}^n$  да е векторното пространство над  $\mathbb{F}$ , т.е.  $\mathbb{F}^n$  е или  $\mathbb{R}^n$  или  $\mathbb{C}^n$ . Нека сега  $x_1, x_2, \dots, x_k \in \mathbb{F}^n$ . Тогава елемент от вида  $\alpha_1 x_1 + \cdots + \alpha_k x_k$  с  $\alpha_i \in \mathbb{F}$  е *линейна комбинация* над  $\mathbb{F}$  на  $x_1, \dots, x_k$ . Множеството на всички линейни комбинации на  $x_1, x_2, \dots, x_k \in \mathbb{F}^n$  е подпространство, което се нарича *обвивка* на  $x_1, \dots, x_k$  и се означава с

$$\text{span}\{x_1, \dots, x_k\} \stackrel{\text{def}}{=} \{x = \alpha_1 x_1 + \cdots + \alpha_k x_k : \alpha_i \in \mathbb{F}\}.$$

Множеството от вектори  $x_1, x_2, \dots, x_k \in \mathbb{F}^n$  се наричат *линейно зависими* над  $\mathbb{F}$ , ако съществуват  $\alpha_1, \dots, \alpha_k \in \mathbb{F}$ , които не са едновременно нули, такива че  $\alpha_1 x_1 + \cdots + \alpha_k x_k = 0$ ; в противен случай те се наричат *линейно независими*.

Нека  $S$  е подпространство на  $\mathbb{F}^n$ , тогава множеството от вектори  $\{x_1, x_2, \dots, x_k\} \in S$  се нарича *базис* на  $S$  ако  $x_1, x_2, \dots, x_k$  са линейно независими и  $S = \text{span}\{x_1, \dots, x_k\}$ . Такъв базис за подпространството  $S$ , обаче, не е единствен, но всички базиси на  $S$  имат един и същ брой елементи. Този брой се нарича *размерност* на  $S$  и се означава с  $\dim S$ .

*Сума*  $Y + Z$  на две линейни подпространства  $Y$  и  $Z$  се нарича подмножеството на всички вектори във вида  $x = y + z$ , където  $y \in Y$ ,  $z \in Z$ .

*Сечение*  $Y \cap Z$  на линейните подпространства  $Y$  и  $Z$  се нарича множеството на всички вектори, които принадлежат едновременно и на  $Y$ , и на  $Z$ .

В общият случай имаме, че

$$\dim(Y + Z) = \dim Y + \dim Z - \dim(Y \cap Z).$$

Множеството от вектори  $x_1, x_2, \dots, x_k$  в  $\mathbb{F}^n$  са взаимно *ортogonalни*, ако  $x_i^H x_j = 0$  за всички  $i \neq j$  и *ортонормални*, ако  $x_i^H x_j = \delta_{ij}$ , където  $\delta_{ij}$  е Кронекеровата делта функция с  $\delta_{ij} = 1$  за  $i = j$  и  $\delta_{ij} = 0$  за  $i \neq j$ .

*Ортogonalното допълнение* на подпространство  $S \subset \mathbb{F}^n$  се дефинира чрез

$$S^\perp \stackrel{\text{def}}{=} \{y \in \mathbb{F}^n : y^H x = 0 \text{ за всички } x \in S\}.$$

Множеството от вектори  $\{u_1, u_2, \dots, u_k\}$  се нарича *ортонормален* базис на подпространството  $S \subset F^n$ , ако те образуват базис на  $S$  и са ортонормални. Такъв базис винаги може да се разшири до пълен ортонормален базис  $\{u_1, u_2, \dots, u_n\}$  на  $F^n$ . Отбележете, че в този случай

$$S^\perp = \text{span} \{u_{k+1}, \dots, u_n\},$$

и  $\{u_{k+1}, \dots, u_n\}$  се нарича *ортogonalно допълнение* на  $\{u_1, u_2, \dots, u_k\}$ .

Нека  $A \in F^{m \times n}$  е линейно преобразуване от  $F^n$  към  $F^m$ , т.е.

$$A : F^n \longrightarrow F^m.$$

(Обърнете внимание, че вектор  $x \in F^n$  може също да се разглежда като линейно преобразуване от  $F$  към  $F^n$ , така че всичко казано за общия матричен случай е в сила също за векторния случай.) Тогава *ядрото* или *нулевото пространство* на линейното преобразуване  $A$  се дефинира с

$$\text{Ker} A = N(A) \stackrel{\text{def}}{=} \{x \in F^n : Ax = 0\},$$

а *образът* на  $A$  е

$$\text{Rg} A = R(A) \stackrel{\text{def}}{=} \{y \in F^m : y = Ax, x \in F^n\},$$

Ясно е, че  $\text{Ker} A$  е подпространство на  $F^n$ , а  $\text{Rg} A$  е подпространство на  $F^m$ . Нещо повече, може да се покаже, че  $\dim(\text{Ker} A) + \dim(\text{Rg} A) = n$  и  $\dim(\text{Rg} A) = \dim(\text{Ker} A)^\perp$ . Обърнете внимание, че  $(\text{Ker} A)^\perp$  е подпространство на  $F^n$ .

## П.4 Собствени стойности и собствени вектори

**Определение П.1 Собствени стойности и собствени вектори** Нека  $A$  е квадратна  $n \times n$  матрица. Собствениите стойности  $\lambda_i$ ,  $i = 1, \dots, n$ , са  $n$ -те корена на характеристичното уравнение от  $n$ -ти ред

$$\det(A - \lambda I) = 0. \quad (\text{П.1})$$

(Десният) *собствен вектор*  $t_i$ , *съответстващ* на *собствената стойност*  $\lambda_i$ , е *нетривиалното решение* ( $t_i \neq 0$ ) на

$$(A - \lambda_i I)t_i = 0 \Leftrightarrow At_i = \lambda_i t_i. \quad (\text{П.2})$$

Съответните *леви собствени вектори*  $q_i$  *удовлетворяват*

$$q_i^H (A - \lambda_i I) = 0 \Leftrightarrow q_i^H A = \lambda_i q_i^H. \quad (\text{П.3})$$

Когато се казва само *собствен вектор* се има предвид *десен собствен вектор*.

**Забележка.** Левите собствени вектори на  $A$  са (десните) собствени вектори на  $A^H$ .

Собствените стойности понякога се наричат характеристични числа. Множеството от собствени стойности на  $A$  се нарича *спектър* на  $A$  и се означава с  $\lambda(A)$ . Най-голямата по абсолютна стойност от собствените стойности на  $A$  е *спектралният радиус* на  $A$ ,  $\rho(A) \stackrel{\text{def}}{=} \max_i |\lambda_i(A)|$ .

Ако  $t$  е собствен вектор, то  $\alpha t$  е също собствен вектор за произволна константа  $\alpha \neq 0$ . Ето защо собствените вектори обикновено се нормализират, така че да имат единична дължина, т.е.  $t_i^H t_i = 1$ . Важно свойство на собствените вектори е, че *собствените вектори съответстващи на различни собствени стойности са винаги линейно независими*. При кратни собствени стойности това може да не е изпълнено, т.е. не всички  $n \times n$  матрици имат  $n$  линейно независими собствени вектори (вж. по-нататък).

Собствените вектори могат да се разглеждат като колони на матрицата  $T$ , а собствените стойности като диагонални елементи на матрицата  $\Lambda$ :

$$T = [t_1, t_2, \dots, t_n]; \quad \Lambda = \text{diag} \{ \lambda_1, \dots, \lambda_n \}. \quad (\text{П.4})$$

Тогава (П.4) може да се запише в следния вид

$$AT = T\Lambda. \quad (\text{П.5})$$

Матрицата  $T$  може да се използва за "диагонализация" на матрицата  $A$ , при условие, че собствените вектори са линейно независими и  $T^{-1}$  съществува. Това е винаги изпълнено, ако собствените стойности са различни и може да е изпълнено в други случаи, напр. ако  $A = I$ . От (П.5) тогава получаваме, че посредством матрицата от собствените вектори матрицата  $A$  се диагонализира по следния начин

$$\Lambda = T^{-1}AT. \quad (\text{П.6})$$

### Свойства на собствените стойности

Нека  $\lambda_i$  означава собствена стойност на  $A$ . Тогава

1. Сумата от собствените стойности на  $A$  е равна на следата на  $A$  (сумата от диагоналните елементи):  $\text{trace}(A) = \sum_i \lambda_i$ .
2. Произведението от собствените стойности на  $A$  е равно на детерминантата на  $A$ :  $\det(A) = \prod_i \lambda_i$ .
3. Собствените стойности на горна или долна триъгълна матрица са равни на диагоналните елементи на матрицата.
4. За реална матрица собствените стойности са или реални или участвуват в спрегнати двойки.
5.  $A$  и  $A^T$  имат едни и същи собствени стойности (но изобщо различни собствени вектори).

6. Обратната матрица  $A^{-1}$  съществува, ако и само ако собствените стойности на  $A$  са различни от нула. Собствените стойности на  $A^{-1}$  тогава са  $1/\lambda_1, \dots, 1/\lambda_n$ .
7. Матрицата  $A + cI$  има собствени стойности  $\lambda_i + c$ .
8. Матрицата  $cA^k$ , където  $k$  е цяло число, има собствени стойности  $c\lambda_i^k$ .
9. Да разгледаме  $l \times m$  матрицата  $A$  и  $m \times l$  матрицата  $B$ . Тогава  $l \times l$  матрицата  $AB$  и  $m \times m$  матрицата  $BA$  имат едни и същи ненулеви собствени стойности. Ако  $l > m$ , матрицата  $AB$  има същите  $m$  собствени стойности както  $BA$  плюс  $l - m$  собствени стойности, които са идентично равни на нула.
10. Собствените стойности са инвариантни към преобразувания на подобие, т.е.  $A$  и  $S^{-1}AS$  имат едни и същи собствени стойности.
11. Една и съща матрица на собствените вектори диагонализира матрицата  $A$  и матрицата  $(I+A)^{-1}$ . (Доказателство:  $T^{-1}(I+A)^{-1}T = (T^{-1}(I+A)T)^{-1} = (I + \Lambda)^{-1}$ .)
12. *Теорема на Гершгорин.* Собствените стойности на  $n \times n$  матрицата  $A$  лежат в обединението на  $n$  кръгове в комплексната равнина (тези кръгове се наричат *кръгове на Гершгорин*), всеки с център  $a_{ii}$  и радиус  $r_i = \sum_{j \neq i} |a_{ij}|$  (сумата на извъндиагоналните елементи в реда  $i$ ). Те също лежат в обединението на  $n$  кръгове, всеки с център  $a_{ii}$  и радиус  $r'_i = \sum_{j \neq i} |a_{ji}|$  (сумата на извъндиагоналните елементи в колоната  $i$ ).
13. Дадена матрица е положително определена, ако и само ако нейните собствени стойности са реални и положителни.

Нека  $J_\lambda^{(k)}$  е матрица от  $k$ -ти ред във вида

$$J_\lambda^{(k)} = \begin{bmatrix} \lambda & 1 & & & 0 \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ 0 & & & \lambda & 1 \end{bmatrix}. \quad (\text{П.7})$$

Тогава  $J_\lambda^{(k)}$  има  $\lambda$  за собствена стойност с кратност  $k$ . Собствените вектори на  $J_\lambda^{(k)}$  са кратни на  $[1, 0, \dots, 0]^T$ . Матрица във вида (П.7) се нарича *Жорданов блок*.

Квадратната матрица  $A$  се нарича *дефектна*, ако тя има собствена стойност с кратност  $k$ , на която съответстват по-малко от  $k$  линейно независими собствени вектори. По такъв начин дефектната матрица има непълна система от собствени вектори. Жордановият блок (П.7) е пример за дефектна матрица. Съществуват теоретични и изчислителни трудности, свързани с дефектните матрици [4].

Ако матрицата  $A$  е дефектна, тя не може да се приведе в диагонална форма посредством преобразуания на подобие. В този случай съществуват единствени числа  $\lambda_1, \lambda_2, \dots, \lambda_r \in \lambda(A)$  и единствени цели числа  $m_1, m_2, \dots, m_r$ ;  $m_1 + m_2 + \dots + m_r = n$ , такива че  $A$  е подобна на матрицата

$$\text{diag}(J_{\lambda_1}^{(m_1)}, J_{\lambda_2}^{(m_2)}, \dots, J_{\lambda_r}^{(m_r)}). \quad (\text{П.8})$$

Формата (П.8) се нарича *Жорданова канонична форма* на  $A$ . Тя е единствена с точност до подреждането на диагоналните блокове. Типичната Жорданова форма може да изглежда така

$$\begin{bmatrix} -2 & 1 & & & & 0 \\ & 0 & -2 & & & \\ & & & -2 & 1 & 0 \\ & & & 0 & -2 & 1 \\ & & & 0 & 0 & -2 \\ & & & & & 4 & 1 \\ 0 & & & & & 0 & 4 \end{bmatrix}. \quad (\text{П.9})$$

Трябва да се отбележи, че една и съща собствена стойност може да участва в няколко блокове на Жордановата форма.

Общият брой на линейно независимите собствени вектори на  $A$  е равен на броя на блоковете в нейната Жорданова форма.

Ако  $X$  е матрицата, която привежда  $A$  в нейната Жорданова форма (П.8), то  $x_1, x_2, \dots, x_m$  удовлетворяват

$$\begin{aligned} Ax_1 &= \lambda_1 x_1, \\ Ax_{i+1} &= \lambda_1 x_{i+1} + x_i \quad i = 1, 2, \dots, m_1 - 1, \end{aligned}$$

като подобни зависимости са в сила и за векторите, съответстващи на другите блокове. Векторите  $x_i$  се наричат *корневи вектори* на матрицата  $A$ .

Полиномите

$$p_i(\lambda) = \det(\lambda I - J_{\lambda_i}^{(m_i)}) = (\lambda - \lambda_i)^{m_i}$$

се наричат *елементарни делители* на  $A$ . Те делят характеристичния полином на  $A$ , който може да се запише във вида

$$p(\lambda) = p_1(\lambda)p_2(\lambda) \dots p_r(\lambda).$$

Жордановата форма е диагонална само когато  $m_1 = m_2 = \dots = m_r = 1$  в който случай всеки делител  $p_i(\lambda)$  е линеен. Ето защо, ако  $A$  е дефектна, тя има нелинейни елементарни делители.

Матрици във вида (П.9), в които някоя собствена стойност присъства в повече от един блок, се наричат *непълни матрици*. Матрица в общ вид е непълна, ако е непълна нейната Жорданова канонична форма. В този случай на дадена  $\lambda_i$  съответстват няколко линейно независими собствени вектори.

### Собствени стойности на матрицата на състоянието

Да разгледаме система, описана с линейните диференциални уравнения

$$\dot{x} = Ax + Bu. \quad (\text{П.10})$$

Освен ако матрицата  $A$  не е диагонална, уравнението (П.10) представлява множество от свързани диференциални уравнения. Да предположим за простота, че собствените вектори на  $A$  са линейно независими и да въведем новия вектор на състоянието  $z = T^{-1}x$ , т.е.  $x = Tz$ . Тогава получаваме

$$T\dot{z} = ATz + Bu \Leftrightarrow \dot{z} = \Lambda z + T^{-1}Bu, \quad (\text{П.11})$$

което е множество от несвързани диференциални уравнения в термините на новите състояния  $z = Tx$ . Свободното решение (т.е. решението при  $u = 0$ ) за всяко състояние  $z_i$  е  $z_i = z_{0i}e^{\lambda_i t}$ , където  $z_{0i}$  е стойността на състоянието за  $t = 0$ . Ако  $\lambda_i$  е реално, то тази мода е устойчива ( $z_i \rightarrow 0$  при  $t \rightarrow \infty$ ), ако и само ако  $\lambda < 0$ . Ако  $\lambda_i = \text{Re}\lambda_i + j\text{Im}\lambda_i$  е комплексно, то получаваме  $e^{\lambda_i t} = e^{\text{Re}\lambda_i t}(\cos(\text{Im}\lambda_i t) + j\sin(\text{Im}\lambda_i t))$  и модата е устойчива ( $z_i \rightarrow 0$  при  $t \rightarrow \infty$ ), ако и само ако  $\text{Re}\lambda_i < 0$ . Фактът, че в този случай новите състояния  $z_i$  са комплексни, не е от особено значение, тъй като действителните физически състояния  $x = Tz$  са реални. Следователно, една линейна система е устойчива, ако и само ако всички собствени стойности на матрицата на състоянието  $A$  имат реални части, които са по-малки от нула, т.е. лежат в лявата комплексна полуравнина.

### Собствени стойности на предавателни матрици

Собствените стойности на предавателната матрица на отворената система,  $\lambda_i(L(j\omega))$ , изчислени като функции на честотата, понякога се наричат характеристичен ходограф и могат да се разглеждат като обобщение на  $L(j\omega)$  за скалярна система.

## П.5 Декомпозиция по сингулярни числа

**Определение П.2 Унитарна матрица.** *Комплексната матрица  $U$  е унитарна, ако*

$$U^H = U^{-1}. \quad (\text{П.12})$$

Всичките собствени стойности на унитарна матрица имат абсолютни стойности равни на 1.

**Определение П.3 Декомпозиция по сингулярни числа.** *Всяка комплексна  $l \times m$  матрица  $A$  може да се факторизира в декомпозицията по сингулярни числа*

$$A = U\Sigma V^H, \quad (\text{П.13})$$

където  $l \times l$  матрицата  $U$  и  $m \times m$  матрицата  $V$  са унитарни, а  $l \times m$  матрицата  $\Sigma$  съдържа диагонална матрица  $\Sigma_1$  от реални, неотрицателни сингулярни числа  $\sigma_i$ , подредени в ненарастващ ред в

$$\Sigma = \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix}; \quad l \geq m \quad (\text{П.14})$$

или

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \end{bmatrix}; \quad l \leq m, \quad (\text{П.15})$$

където

$$\Sigma_1 = \text{diag} \{ \sigma_1, \sigma_2, \dots, \sigma_k \}; \quad k = \min(l, m) \quad (\text{П.16})$$

и

$$\bar{\sigma} \stackrel{\text{def}}{=} \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k \stackrel{\text{def}}{=} \underline{\sigma}. \quad (\text{П.17})$$

Унитарните матрици  $U$  и  $V$  формират ортонормални базиси за колонното (входно) пространство и редовото (изходно) пространство на  $A$ . Колоните на  $V$ , означавани с  $v_i$ , се наричат десни или входни сингулярни вектори, а колоните на  $U$ , означавани с  $u_i$ , се наричат леви или изходни сингулярни вектори. Дефинираме  $\bar{u} \stackrel{\text{def}}{=} u_1, \bar{v} \stackrel{\text{def}}{=} v_1; \underline{u} \stackrel{\text{def}}{=} u_k$  и  $\underline{v} \stackrel{\text{def}}{=} v_k$ .

Отбележете, че декомпозицията в (П.13) не е единствена, тъй като  $A = U' \Sigma V'^H$ , където  $U' = US$ ,  $V' = VS^{-1}$ ,  $S = \text{diag} \{ e^{j\Theta_i} \}$  и  $\Theta_i$  е кое да е реално число, е също декомпозиция по сингулярните числа на  $A$ . Обаче, сингулярните числа  $\sigma_i$  са единствени.

Сингулярните числа са положителните квадратни корени на  $k = \min(l, m)$  най-големи собствени стойности едновременно на  $AA^H$  и  $A^H A$ . Имаме, че

$$\sigma_i(A) = \sqrt{\lambda_i(A^H A)} = \sqrt{\lambda_i(AA^H)}. \quad (\text{П.18})$$

Също така, колоните на  $U$  и  $V$  са собствени вектори с единична дължина на  $AA^H$  и  $A^H A$ , съответно. За да изведем (П.18), записваме

$$AA^H = (U \Sigma V^H)(U \Sigma V^H)^H = (U \Sigma V^H)(V \Sigma^H U^H) = U \Sigma \Sigma^H U^H \quad (\text{П.19})$$

или еквивалентно, тъй като  $U$  е унитарна и удовлетворява  $U^H = U^{-1}$ , получаваме

$$(AA^H)U = U \Sigma \Sigma^H. \quad (\text{П.20})$$

Тогава виждаме, че  $U$  е матрицата от собствените вектори на  $AA^H$  и  $\{\sigma_i^2\}$  са нейните собствени стойности. По подобен начин се установява, че  $V$  е матрицата от собствените вектори на  $A^H A$ .

### Ранг на матрица

**Определение П.4 Рангът** на матрица е равен на броя на ненулевите собствени стойности на матрицата. Нека  $\text{rank}(A) = r$ , тогава матрицата  $A$  се нарича матрица с непълен ранг, ако  $r < k = \min(l, m)$  и имаме сингулярни числа  $\sigma_i = 0$  за  $i = r + 1, \dots, k$ . Квадратна матрица с непълен ранг е особена

матрица (матриците, които не са квадратни, се разглеждат като особени матрици).

Рангът на матрица остава непроменен след умножаване отляво или отдясно с неособена матрица. Освен това, за  $l \times m$  матрица  $A$  и  $m \times p$  матрица  $B$ , рангът на тяхното произведение  $AB$  е ограничен от (неравенство на Силвестър):

$$\text{rank}(A) + \text{rank}(B) - m \leq \text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B)). \quad (\text{П.21})$$

### Сингулярни числа на $2 \times 2$ матрица

В общият случай сингулярните числа трябва да се изчисляват по числен път. За  $2 \times 2$  матрици, обаче, могат да се изведат аналитични изрази. Да въведем

$$b \stackrel{\text{def}}{=} \text{trace}(A^H A) = \sum_{i,j} |a_{ij}|^2, \quad c \stackrel{\text{def}}{=} \det(A^H A).$$

Тъй като сумата на собствените стойности на матрица е равна на следата ѝ, а произведението е равно на детерминантата ѝ, то

$$\lambda_1 + \lambda_2 = b, \quad \lambda_1 \lambda_2 = c.$$

Решавайки за  $\lambda_1$  и  $\lambda_2$  и използвайки  $\sigma_i(A) = \sqrt{\lambda_i(A^H A)}$ , получаваме

$$\bar{\sigma}(A) = \sqrt{\frac{b + \sqrt{b^2 - 4c}}{2}}; \quad \underline{\sigma}(A) = \sqrt{\frac{b - \sqrt{b^2 - 4c}}{2}}. \quad (\text{П.22})$$

Например, за  $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$  имаме, че  $b = \sum |a_{ij}|^2 = 1 + 4 + 9 + 16 = 30$ ,  $c = (\det A)^2 = (-2)^2 = 4$ , което дава  $\bar{\sigma}(A) = 5.465$  и  $\underline{\sigma}(A) = 0.366$ . Отбележете, че за особени  $2 \times 2$  матрици (с  $\det A = 0$  и  $\underline{\sigma}(A) = 0$ ) получаваме  $\bar{\sigma}(A) = \sqrt{\sum |a_{ij}|^2} \stackrel{\text{def}}{=} \|A\|_F$  (Фробениусовата норма, което е специален случай на (П.91).

### Декомпозиция по сингулярните числа на обратна матрица

Тъй като  $A = U \Sigma V^H$ , получаваме, при условие, че  $m \times m$  матрицата  $A$  е неособена, че

$$A^{-1} = V \Sigma U^H. \quad (\text{П.23})$$

Това е декомпозицията по сингулярни числа на  $A^{-1}$ , но със сингулярни числа в обратен ред. Нека  $j = m - i + 1$ . Тогава от (П.23) следва, че

$$\sigma_i(A^{-1}) = 1/\sigma_j(A), \quad u_i(A^{-1}) = v_j(A), \quad v_j(A^{-1}) = u_j(A) \quad (\text{П.24})$$

и в частност

$$\bar{\sigma}(A^{-1}) = 1/\underline{\sigma}(A). \quad (\text{П.25})$$



### Неравенства за сингулярните числа

Сингулярните числа ограничават абсолютните стойности на собствените стойности (вж. също (П.83)):

$$\underline{\sigma}(A) \leq |\lambda_i(A)| \leq \bar{\sigma}(A). \quad (\text{П.26})$$

Следното неравенство е очевидно от определението за декомпозиция по сингулярни числа:

$$\bar{\sigma}(A^H) = \bar{\sigma}(A) \text{ и } \bar{\sigma}(A^T) = \bar{\sigma}(A). \quad (\text{П.27})$$

Следващото важно свойство е доказано по-нататък (уравнение П.67))

$$\bar{\sigma}(AB) \leq \bar{\sigma}(A)\bar{\sigma}(B). \quad (\text{П.28})$$

За неособена матрица  $A$  (или  $B$ ) имаме също долна граница за  $\bar{\sigma}(B)$

$$\underline{\sigma}(A)\bar{\sigma}(B) \leq \bar{\sigma}(AB) \text{ или } \bar{\sigma}(A)\underline{\sigma}(B) \leq \bar{\sigma}(AB). \quad (\text{П.29})$$

Имаме също така долна граница за минималното сингулярно число

$$\underline{\sigma}(A)\underline{\sigma}(B) \leq \underline{\sigma}(AB). \quad (\text{П.30})$$

При блочна матрица са полезни следните неравенства

$$\max \{ \bar{\sigma}(A), \bar{\sigma}(B) \} \leq \bar{\sigma} \begin{bmatrix} A \\ B \end{bmatrix} \leq \sqrt{2} \max \{ \bar{\sigma}(A), \bar{\sigma}(B) \}, \quad (\text{П.31})$$

$$\bar{\sigma} \begin{bmatrix} A \\ B \end{bmatrix} \leq \bar{\sigma}(A) + \bar{\sigma}(B). \quad (\text{П.32})$$

Следното равенство за блочни матрици се използва често:

$$\bar{\sigma} \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} = \max \{ \bar{\sigma}(A) + \bar{\sigma}(B) \}. \quad (\text{П.33})$$

Един друг много важен резултат е *теоремата на Фан Цзи*

$$\sigma_i(A) - \bar{\sigma}(B) \leq \sigma_i(A+B) \leq \sigma_i(A) + \bar{\sigma}(B). \quad (\text{П.34})$$

Два специални случая на (П.34) са:

$$|\bar{\sigma}(A) - \bar{\sigma}(B)| \leq \bar{\sigma}_i(A+B) \leq \bar{\sigma}(A) + \bar{\sigma}(B), \quad (\text{П.35})$$

$$\underline{\sigma}(A) - \bar{\sigma}(B) \leq \underline{\sigma}(A+B) \leq \underline{\sigma}(A) + \bar{\sigma}(B). \quad (\text{П.36})$$

Неравенството (П.36) дава

$$\underline{\sigma}(A) - 1 \leq \underline{\sigma}(I+A) \leq \underline{\sigma}(A) + 1. \quad (\text{П.37})$$

Комбинирайки (П.25) и (П.37) получаваме зависимост, която е полезна при изчисляване усилването на затворената система:

$$\underline{\sigma}(A) - 1 \leq \frac{1}{\underline{\sigma}(I+A)^{-1}} \leq \underline{\sigma}(A) + 1. \quad (\text{П.38})$$

### Декомпозицията по сингулярни числа като сума на матрици с ранг 1

Нека  $r$  означава рангът на  $l \times m$  матрицата  $A$ . Ние можем да разглеждаме декомпозицията по сингулярни числа като декомпозиция на  $A$  в  $r$   $l \times m$  матрици, всяка с ранг 1. Имаме, че

$$A = U \Sigma V^H = \sum_{i=1}^r \sigma_i u_i v_i^H. \quad (\text{П.39})$$

Останалите членове от  $r+1$  до  $k = \min(l, m)$  имат сингулярни числа, равни на нула и нямат принос към сумата. Първата и най-важна подматрица се дава от  $A_1 = \sigma_1 u_1 v_1^H$ . Ако сега разгледаме матрицата на остатъка

$$A^1 = A - A_1 = A - \sigma_1 u_1 v_1^H, \quad (\text{П.40})$$

то

$$\sigma_1(A^1) = \sigma_2(A), \quad (\text{П.41})$$

т.е. най-голямото сингулярно число на  $A^1$  е равно на второто сингулярно число на първоначалната матрица. По подобен начин най-голямото сингулярно число на  $A - A_1 - A_2$ , където  $A_2 = \sigma_2 u_2 v_2^H$ , е  $\sigma_3$  и т.н.

### Особеност на матрицата $A + E$

От лявото неравенство в (П.36) намираме, че

$$\underline{\sigma}(E) < \underline{\sigma}(A) \Rightarrow \underline{\sigma}(A + E) > 0 \quad (\text{П.42})$$

и  $A + E$  е неособена. От друга страна, винаги съществува някакво  $E$  със  $\underline{\sigma}(E) = \underline{\sigma}(A)$ , което прави  $A + E$  особена, напр. избирайки  $E = -\underline{u} \underline{\sigma} \underline{v}^H$ ; вж. (П.39). По такъв начин най-малкото сингулярно число на  $A$  характеризира доколко матрицата  $A$  е близо до особена матрица или до матрица с непълен ранг. Проверката (П.42) се използва често в числения анализ и е важно неравенство при проверката за робастност.

### Декомпозиция по сингулярни числа с понижена размерност

Тъй като има само  $r = \text{rank}(A) \leq \min(l, m)$  ненулеви сингулярни числа и тъй като само ненулевите сингулярни числа имат принос към резултата, декомпозицията на  $A$  понякога се записва като декомпозиция по сингулярните числа с понижена размерност по следния начин

$$A^{l \times m} = U_r^{l \times r} \Sigma_r^{r \times r} (v_r^{m \times r})^H, \quad (\text{П.43})$$

където матриците  $U_r$  и  $V_r$  съдържат само първите  $r$  колони на матриците  $U$  и  $V$ , въведени по-горе. Тук бе използвано означението  $A^{l \times m}$ , за да се индикира, че  $A$  е  $l \times m$  матрица. Декомпозицията по сингулярни числа се използва при изчисляването на псевдо-обратната, вж. (П.46).

### Псевдо-обратна (обобщена обратна)

Да разгледаме системата линейни уравнения

$$y = Ax \quad (\text{П.44})$$

за даден  $l \times 1$  вектор  $y$  и дадена  $l \times m$  матрица  $A$ . Решение по *метода на най-малките квадрати* на (П.44) е  $m \times 1$  вектор  $x$ , такъв че

$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_m^2}$  се минимизира над множеството на всички вектори, за които се минимизира  $\|y - Ax\|_2$ . Решението се получава в термините на псевдо-обратната (обобщената обратна на Муур-Пенроуз) на  $A$ :

$$x = A^\dagger y. \quad (\text{П.45})$$

Псевдо-обратната може да се получи от декомпозицията по сингулярни числа на  $A = U\Sigma V^H$  чрез

$$A^\dagger = V_r \Sigma_r^{-1} U_r^H = \sum_{i=1}^r \frac{1}{\sigma_i(A)} v_i u_i^H, \quad (\text{П.46})$$

където  $r$  е броят на ненулевите сингулярни числа на  $A$ . Имаме, че

$$\underline{\sigma}(A) = 1/\bar{\sigma}(A^\dagger). \quad (\text{П.47})$$

Обърнете внимание, че  $A^\dagger$  съществува за всяка матрица  $A$ , даже за особена квадратна матрица и правоъгълна матрица. Псевдо-обратната също удовлетворява

$$AA^\dagger A = A \text{ и } A^\dagger AA^\dagger = A^\dagger.$$

Отбележете следните случаи (където  $r$  е рангът на  $A$ ):

1.  $r = l = m$ , т.е.  $A$  е неособена. В този случай  $A^\dagger = A^{-1}$  е обратната матрица.
2.  $r = m \leq l$  има пълен колонен ранг. Това е "стандартната задача за най-малки квадрати", когато искаме да минимизираме  $\|y - Ax\|_2$  и решението е

$$A^\dagger = (A^H A)^{-1} A^H. \quad (\text{П.48})$$

В този случай  $A^\dagger A = I$ , така че  $A^\dagger$  е *лявата обратна* на  $A$ .

3.  $r = l \leq m$ , т.е.  $A$  има пълен редови ранг. В този случай (П.44) има безброй много решения и търсим това, което минимизира  $\|x\|_2$ . Имаме, че

$$A^\dagger = A^H (AA^H)^{-1}. \quad (\text{П.49})$$

В този случай  $AA^\dagger A = I$ , така че  $A^\dagger$  е *дясната обратна* на  $A$ .

4.  $r < k = \min(l, m)$  (общият случай). В този случай двете матрици  $A^H A$  и  $AA^H$  са с непълен ранг и за да се получи псевдо-обратната трябва да се използва декомпозицията по сингулярни числа. В този случай  $A$  няма нито лява, нито дясна псевдо-обратна.

### Регресия по главните компоненти

Трябва да се отбележи, че псевдо-обратната в (П.46) може да е много чувствителна към шумове и може да се получи с много големи по модул елементи, ако най-малкото сингулярно число  $\sigma_r$  е малко. В регресията по главните компоненти този проблем се избягва, използвайки само първите  $q \leq r$  сингулярни числа, които могат да се разграничат от шума. Тогава за псевдо-обратната се получава

$$A^\dagger = \sum_{i=1}^q \frac{1}{\sigma_i} v_i u_i^H. \quad (\text{П.50})$$

Основният проблем в този случай е определянето на  $q$ . Тази задача е еквивалентна на определянето на числения ранг на матрица.

### Число на обусловеност

**Числото на обусловеност** на матрица може да се дефинира като отношението

$$\gamma(A) = \sigma_1(A)/\sigma_k(A) = \bar{\sigma}(A)/\underline{\sigma}(A), \quad (\text{П.51})$$

където  $k = \min(l, m)$ . Матрица с голямо число на обусловеност се нарича *лошо обусловена*. Това определение дава безкрайно голямо число на обусловеност за матрици с непълен ранг. За неособена матрица от (П.25) се получава

$$\gamma(A) = \bar{\sigma}(A) \cdot \bar{\sigma}(A^{-1}). \quad (\text{П.52})$$

Използват с и други определения за число на обусловеност на неособена матрица, например

$$\gamma_p(A) = \|A\|_p \cdot \|A^{-1}\|_p, \quad (\text{П.53})$$

където  $\|A\|_p$  означава матрична  $p$ -норма. Ако се използва индуцираната 2-норма (максималното сингулярно число), се получава (П.52). От (П.52) и (П.28) за неособени матрици се получава, че

$$\gamma(AB) \leq \gamma(A)\gamma(B). \quad (\text{П.54})$$

**Минимизираното число на обусловеност** се получава, минимизирайки числото на обусловеност над всички възможни мащабирания. Имаме, че

$$\gamma^*(G) \stackrel{\text{def}}{=} \min_{D_I, D_O} \gamma(D_O G D_I), \quad (\text{П.55})$$

където  $D_I$  и  $D_O$  са диагонални мащабиращи матрици. Ако се използват едностранни мащабирания, се получават входното и изходното минимизирани числа на обусловеност:

$$\gamma_I^*(G) \stackrel{\text{def}}{=} \min_{D_I} \gamma(G D_I), \quad \gamma_O^*(G) \stackrel{\text{def}}{=} \min_{D_O} \gamma(D_O G). \quad (\text{П.56})$$

## П.6 Векторни и матрични норми

Полезно е да се разполага с единствено число, което дава цялостна характеристика на големината на вектор, матрица, сигнал или система. За тази цел се използват функции, които се наричат норми. Най-често използваната норма е Евклидовата векторна норма,  $\|e\|_2 = \sqrt{|e_1|^2 + |e_2|^2 + \dots + |e_m|^2}$ . Геометрически тя представлява разстоянието между две точки  $y$  и  $x$ , където  $e_i = y_i - x_i$  е разликата между техните  $i$ -ти координати.

**Определение П.5** *Норма на е (което може да е вектор, матрица, сигнал или система) е реално число, означаващо с  $\|e\|$ , което удовлетворява следните свойства:*

1. *Неотрицателност:*  $\|e\| \geq 0$ .
2. *Положителност:*  $\|e\| = 0 \Leftrightarrow e = 0$  (за полунормите имаме, че  $\|e\| = 0 \Leftarrow e = 0$ ).
3. *Хомогенност:*  $\|\alpha \cdot e\| = |\alpha| \cdot \|e\|$  за всички комплексни скалари  $\alpha$ .
4. *Неравенство на триъгълника:*

$$\|e_1 + e_2\| \leq \|e_1\| + \|e_2\|. \quad (\text{П.57})$$

По-точно,  $e$  е елемент на векторно пространство  $V$  над полето  $\mathbb{C}$  на комплексните числа и горните свойства трябва да се удовлетворяват за всяко  $e, e_1, e_2 \in V$  и за всяко  $\alpha \in \mathbb{C}$ .

### Векторни норми

Разглеждаме вектор  $a$  с  $m$  елемента, т.е. векторното пространство е  $V = \mathbb{C}^m$ . За да се илюстрират различните норми, ще изчислим всяка от тях за вектора

$$b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ -5 \end{bmatrix}. \quad (\text{П.58})$$

Ще разгледаме три норми, които са специален случай на векторната  $p$ -норма

$$\|a\|_p = (\sum |a_i|^p)^{1/p}, \quad (\text{П.59})$$

където трябва да имаме  $p \geq 1$  за да се удовлетвори неравенството на триъгълника (свойство 4 на норма). Тук  $a$  е вектор-колона с елементи  $a_i$  и  $|a_i|$  е абсолютната стойност на комплексния скалар  $a_i$ .

**Векторна 1-норма** (или норма-сума). Имаме, че

$$\|a\|_1 \stackrel{\text{def}}{=} \sum_i |a_i| \quad (\|b\|_1 = 1 + 3 + 5 = 9). \quad (\text{П.60})$$

**Векторна 2-норма (Евклидова норма).** Това е най-често използваната векторна норма и съответствува на най-малкото разстояние между две точки

$$\|a\|_2 \stackrel{\text{def}}{=} \sqrt{\sum_i |a_i|^2} \quad (\|b\|_2 = \sqrt{1 + 9 + 25} = 5.916). \quad (\text{П.61})$$

Евклидовата норма удовлетворява свойството

$$a^H a = \|a\|_2^2, \quad (\text{П.62})$$

където  $a^H$  означава спрегнато-транспонираният на вектора  $a$ .

**Векторна норма  $\infty$ .** Това е най-големият по абсолютна стойност елемент на вектора. Използваме означението  $\|a\|_{\max}$ , така че

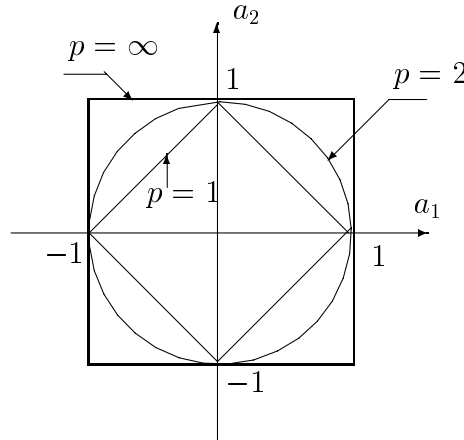
$$\|a\|_{\max} \equiv \|a\|_{\infty} \stackrel{\text{def}}{=} \max_i |a_i| \quad (\|b\|_{\max} = |-5| = 5). \quad (\text{П.63})$$

Тъй като различните норми се различават само с постоянен множител, те често се наричат *еквивалентни*. Така например, за вектор с  $m$  елемента е изпълнено

$$\|a\|_{\max} \leq \|a\|_2 \leq \sqrt{m} \|a\|_{\max}, \quad (\text{П.64})$$

$$\|a\|_2 \leq \|a\|_1 \leq \sqrt{m} \|a\|_2. \quad (\text{П.65})$$

На фиг. П.1 е илюстрирана разликата между векторните норми чрез изчертаване на контурите за  $\|a\|_p = 1$  за случая на  $m = 2$ .



Фигура П.1. Контури за векторната  $p$ -норма,  $\|a\|_p = 1$  за  $p = 1, 2, \infty$

### Матрични норми

Ще разглеждаме постоянна  $l \times m$  матрица  $A$ , която може да представлява, например, стойността за определено  $\omega$  на честотната характеристика  $G(j\omega)$  на система  $G(s)$  с  $m$  входове и  $l$  изходи. За числени илюстрации ще се използва следната  $2 \times 2$  матрица

$$A_0 = \begin{bmatrix} 1 & 2 \\ -3 & 4 \end{bmatrix}. \quad (\text{П.66})$$

**Определение П.6** Нормата  $\|A\|$  е **матрична норма** ако, в допълнение на четирите свойства на нормата от Определение П.4, се удовлетворява също мултипликативното свойство (наричано също условие за съгласуване):

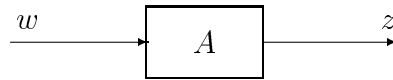
$$\|AB\| \leq \|A\| \cdot \|B\|. \quad (\text{П.67})$$

Условието (П.67) е много важно при разглеждане свойствата на съединения от системи.

Една от най-често използваните матрични норми е **Фробениусовата матрична норма** (или Евклидова норма). Тя се дефинира като квадратният корен от сумата на квадратите на абсолютните стойности на елементите на матрицата

$$\|A\|_F = \sqrt{\sum_{i,j} |a_{i,j}|^2} = \sqrt{\text{trace}(A^H A)}. \quad (\|A_0\|_F = \sqrt{30} = 5.477). \quad (\text{П.68})$$

Фробениусовата норма играе важна роля в теория на управлението, тъй като тя се използва при сумиране на каналите, например при оптималното LQG управление.



Фигура П.2. Представяне на (П.69)

Най-важните матрични норми са т.н. индуцирани матрични норми, които имат тясна връзка с усилването на сигналите в системи. Да разгледаме следното уравнение, което е илюстрирано на фиг. П.2:

$$z = Aw. \quad (\text{П.69})$$

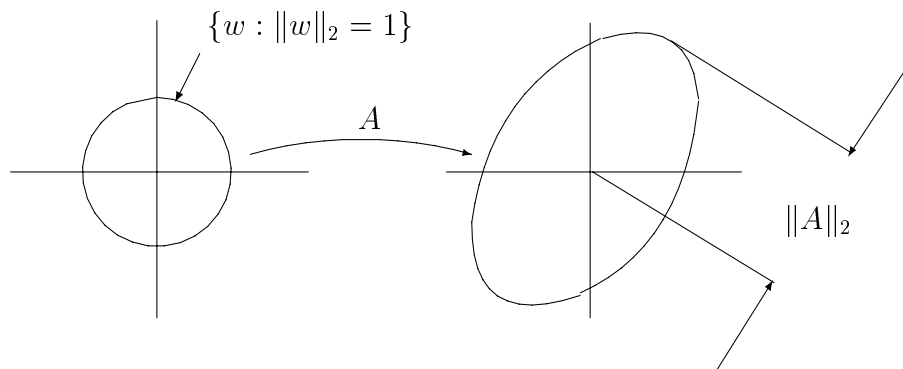
Ние можем да смятаме  $w$  като входен вектор, а  $z$  като изходен вектор и да разгледаме "усилването" на матрицата  $A$ , дефинирано от отношението  $\|z\|/\|w\|$ . Максималното усилване за всички възможни входни направления, т.е. всички възможни посоки на вектора  $w$ , представлява особен интерес. То се дава от *индуцираната норма*, която се дефинира като

$$\|A\|_p \stackrel{\text{def}}{=} \max_{w \neq 0} \frac{\|Aw\|_p}{\|w\|_p}, \quad (\text{П.70})$$

където  $\|w\|_p = (\sum_i |w_i|^p)^{1/p}$  означава векторна  $p$ -норма. Иначе казано, търсим направление на вектора  $w$ , такова че отношението  $\|z\|_p/\|w\|_p$  е максимизирано. По такъв начин индуцираната норма дава най-голямото възможно "усилване" на матрица. Използува се и следното еквивалентно определение

$$\|A\|_p = \max_{\|w\|_p \leq 1} \|Aw\|_p = \max_{\|w\|_p = 1} \|Aw\|_p. \quad (\text{П.71})$$

Геометричната интерпретация на индуцираната норма е, че тя е максималната норма на всички вектори  $z$ , получени след преобразуване с матрицата  $A$  на всички вектори  $w$ , имащи норма единица. Това е илюстрирано на фиг. П.3 за вектори в  $\mathbb{R}^2$  с 2-норма.



Фигура П.3. Геометрична интерпретация на индуцирана норма

За индуцираните 1-, 2- и  $\infty$ -норми са в сила следните тъждества:

$$\|A\|_1 = \max_j \left( \sum_i |a_{ij}| \right) \quad \text{"максимална колонна сума"} \quad (\text{П.72})$$

$$\|A\|_\infty = \max_i \left( \sum_j |a_{ij}| \right) \quad \text{"максимална редова сума"} \quad (\text{П.73})$$

$$\|A\|_2 = \bar{\sigma}(A) = \sqrt{\rho(A^H A)} \quad \text{"сингулярно число или спектрална норма"} \quad (\text{П.74})$$

където спектралният радиус  $\rho(A) = \max_i |\lambda_i(A)|$  е най-голямата собствена стойност на матрицата  $A$ . Отбележете, че индуцираната 2-норма на матрица е равна на най-голямото сингулярно число и често се нарича спектрална норма. За примерната матрица в (П.66) получаваме

$$\|A\|_1 = 6; \quad \|A\|_\infty = 7; \quad \|A\|_2 = \bar{\sigma}(A_0) = 5.117.8 \quad (\text{П.75})$$

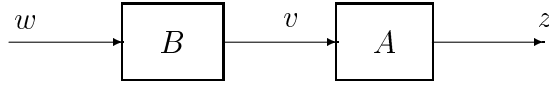
**Теорема П.1** *Всички индуцирани норми  $\|A\|_p$  са матрични норми и следователно удовлетворяват мултипликативното свойство*

$$\|AB\|_p \leq \|A\|_p \cdot \|B\|_p. \quad (\text{П.76})$$

*Доказателство:* Да разгледаме следното множество от уравнения, което е илюстрирано графически на фиг. П.4.

$$z = Av, \quad v = Bw \Rightarrow z = ABw. \quad (\text{П.77})$$





Фигура П.4. Представяне на (П.77)

От определението за индуцирана норма получаваме, чрез въвеждане най-напред на  $v = Bw$ , след това умножаване на числителя и знаменателя с  $\|v\|_p \neq 0$  и най-накрая максимизирайки всеки член включващ  $w$  и  $v$  независимо, че

$$\|AB\|_p \stackrel{\text{def}}{=} \max_{w \neq 0} \frac{\|A \overbrace{Bw}^v\|_p}{\|w\|_p} = \max_{w \neq 0} \frac{\|Av\|_p}{\|v\|_p} \cdot \frac{\|Bw\|_p}{\|w\|_p} \leq \max_{v \neq 0} \frac{\|Av\|_p}{\|v\|_p} \cdot \max_{w \neq 0} \frac{\|Bw\|_p}{\|w\|_p}$$

и (П.77) следва от определението за индуцирана норма.  $\square$

### Следствия от мултипликативното свойство

За матричните норми мултипликативното свойство  $\|AB\| \leq \|A\| \cdot \|B\|$  е в сила за матрици  $A$  и  $B$  с произволна размерност, при условие че произведението  $AB$  съществува. В частност, то е в сила, ако изберем  $A$  и  $B$  като вектори. От тук получаваме:

1. Да изберем  $B$  да е вектор, т.е.  $B = w$ . Тогава за всяка матрична норма от (П.67) имаме, че

$$\|Aw\| \leq \|A\| \cdot \|w\|. \quad (\text{П.78})$$

В този случай се казва, че "матричната норма  $\|A\|$  е съгласувана със съответната й векторна норма  $\|w\|$ ". От (П.70) е очевидно, че всяка индуцирана матрична  $p$ -норма е съгласувана със съответната й векторна  $p$ -норма. По подобен начин, Фробениусовата норма е съгласувана с векторната 2-норма (тъй като, когато  $w$  е вектор,  $\|w\|_F = \|w\|_2$ ).

2. От (П.78) също за всяка матрична норма получаваме, че

$$\|A\| \geq \max_{w \neq 0} \frac{\|Aw\|}{\|w\|}. \quad (\text{П.79})$$

Обърнете внимание, че индуцираните норми са дефинирани така, че в (П.79) имаме равенство. Свойството  $\|A\|_F \geq \bar{\sigma}(A)$  тогава следва от факта, че  $\|w\|_F = \|w\|_2$ .

3. Нека изберем  $A = z^H$  и  $B = w$  като вектори. Тогава използвайки Фробениусовата норма или индуцираната 2-норма (сингулярното число) в (П.67) извеждаме неравенството на Коши-Шварц

$$|z^H w| \leq \|z\|_2 \cdot \|w\|_2, \quad (\text{П.80})$$

където  $z$  и  $w$  са вектор-колони с еднаква размерност и  $z^H w$  е Евклидовото скалярно произведение на векторите  $z$  и  $w$ .

4. Скаларното произведение може също да се използва, за да се дефинира ъгълът  $\phi$  между два вектора  $z$  и  $w$

$$\phi = \cos^{-1} \left( \frac{|z^H w|}{\|z\|_2 \cdot \|w\|_2} \right). \quad (\text{П.81})$$

Отбележете, че при това определение  $\phi$  е между  $0^\circ$  и  $90^\circ$ .

### Спектрален радиус

Спектралният радиус  $\rho(A)$  е абсолютната стойност на най-голямата собствена стойност на матрицата  $A$ ,

$$\rho(A) = \max_i |\lambda_i(A)|. \quad (\text{П.82})$$

Той не е норма, тъй като не удовлетворява свойствата 2 и 4 в Определение П.5. Например, ако

$$A_1 = \begin{bmatrix} 1 & 0 \\ 10 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 10 \\ 0 & 1 \end{bmatrix},$$

то имаме, че  $\rho(A_1) = 1$  и  $\rho(A_2) = 1$ . Обаче,  $\rho(A_1 + A_2) = 12$  и  $\rho(A_1 A_2) = 101.99$ , което не удовлетворява нито неравенството на триъгълника (свойство 4 на норма), нито мултипликативното свойство в (П.67).

Въпреки че спектралният радиус не е норма, той дава долна граница за всяка матрична норма, което може да е много полезно.

**Теорема П.2** *За всяка матрична норма (и в частност за всяка индуцирана норма)*

$$\rho(A) \leq \|A\|. \quad (\text{П.83})$$

*Доказателство:* Тъй като  $\lambda_i(A)$  е собствена стойност на  $A$ , то имаме, че  $At_i = \lambda_i t_i$ , където  $t_i$  означава собственият вектор. Получаваме

$$|\lambda_i| \cdot \|t_i\| = \|\lambda_i t_i\| = \|At_i\| \leq \|A\| \cdot \|t_i\| \quad (\text{П.84})$$

(последното неравенство следва от (П.78)). По такъв начин за всяка матрична норма  $|\lambda_i(A)| \leq \|A\|$  и тъй като това е в сила за всички собствени стойности, следва резултатът.  $\square$

За нашата примерна матрица в (П.66) получаваме  $\rho(A_0) = \sqrt{10} \approx 3.162$ , което е по-малко отколкото всички индуцирани норми ( $\|A_0\|_1 = 6$ ,  $\|A_0\|_\infty = 7$ ,  $\bar{\sigma}(A_0) = 5.117$ ) и е също по-малко, отколкото Фробениусовата норма ( $\|A\|_F = 5.477$ ).

Една проста физическа интерпретация на (П.83) е, че собствената стойност характеризира усилването на матрицата само по някои направления (дадени от собствените вектори) и следователно трябва да е по-малко, отколкото за матрична норма, която позволява всяко направление и дава максималното матрично усилване, вж. (П.79).

### Някои зависимости между матричните норми

Различните норми на матрицата  $A$  са тясно свързани, както може да се види от следните неравенства. Нека  $A$  е  $l \times m$  матрица, тогава

$$\bar{\sigma}(A) \leq \|A\|_F \leq \sqrt{\min(l, m)} \bar{\sigma}(A), \quad (\text{П.85})$$

$$\bar{\sigma}(A) \leq \sqrt{\|A\|_1 \|A\|_\infty}, \quad (\text{П.86})$$

$$\frac{1}{\sqrt{m}} \|A\|_\infty \leq \bar{\sigma}(A) \leq \sqrt{l} \|A\|_\infty, \quad (\text{П.87})$$

$$\frac{1}{\sqrt{l}} \|A\|_1 \leq \bar{\sigma}(A) \leq \sqrt{m} \|A\|_1. \quad (\text{П.88})$$

Всички тези норми удовлетворяват (П.67). Неравенствата са нестроги, т.е. съществуват матрици с каквито и да са размерности, за които равенствата са в сила.

Важно свойство на Фробениусовата норма и максималното сингулярно число (индуцираната 2-норма) е, че те са инвариантни по отношение на унитарни преобразувания, т.е. за унитарни матрици  $U_i$ , удовлетворяващи  $U_i U_i^H = I$  имаме, че

$$\|U_1 A U_2\|_F = \|A\|_F, \quad (\text{П.89})$$

$$\bar{\sigma}(U_1 A U_2) = \bar{\sigma}(A). \quad (\text{П.90})$$

От декомпозицията по сингулярни числа на матрица  $A = U \Sigma V^H$  и (П.89) се получава важната зависимост между Фробениусовата норма и сингулярните числа  $\sigma_i(A)$

$$\|A\|_F = \sqrt{\sum_i \sigma_i^2(A)}. \quad (\text{П.91})$$

Теоремата на Перон–Фробениус, която е приложима към квадратна матрица  $A$ , гласи че

$$\min_D \|D A D^{-1}\|_1 = \min_D \|D A D^{-1}\|_\infty = \rho(|A|), \quad (\text{П.92})$$

където  $D$  е диагонална ”мащабираща” матрица,  $|A|$  означава матрица, чийто елементи са заменени с техните абсолютни стойности и  $\rho(|A|) = \max_i |\lambda_i(|A|)|$  е Пероновият корен (собствената стойност на Перон–Фробениус). Пероновият корен е по-голям или равен на спектралният радиус,  $\rho(A) \leq \rho(|A|)$ .

### Матрични и векторни норми в MATLAB

Следните MATLAB функции се използват за матрици:

$$\begin{array}{ll} \bar{\sigma}(A) = \|A\|_2 & \text{norm(A,2) или max(svd(A))} \\ \|A\|_1 & \text{norm(A,1)} \\ \|A\|_\infty & \text{norm(A,'inf')} \\ \|A\|_F & \text{norm(A,'fro')} \\ \rho(A) & \text{max(abs(eig(A)))} \\ \rho(|A|) & \text{max(eig(abs(A)))} \\ \gamma(A) = \bar{\sigma}(A)/\underline{\sigma}(A) & \text{cond(A)} \end{array}$$

За вектори:

$$\begin{array}{ll} \|a\|_1 & \text{norm(a,1)} \\ \|a\|_2 & \text{norm(a,2)} \end{array}$$

## П.7 Линејни дробни преобразувания

Нека разгледаме матрица  $P$  с размерност  $(n_1 + n_2) \times (m_1 + m_2)$  и да я разделим по следния начин:

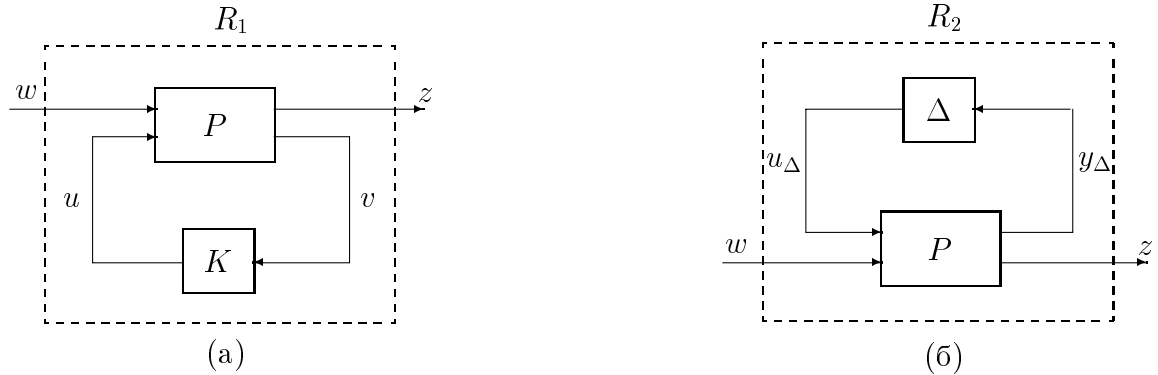
$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}. \quad (\text{П.93})$$

Нека матриците  $\Delta$  и  $K$  имат размерности  $m_1 \times n_1$  и  $m_2 \times n_2$ , съответно (съвместими с горните и долните блокове на  $P$ , съответно). Приемаме следните означения за долно и горно линејни дробни преобразувания

$$F_\ell(P, K) \stackrel{\text{def}}{=} P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}, \quad (\text{П.94})$$

$$F_u(P, \Delta) \stackrel{\text{def}}{=} P_{22} + P_{21}\Delta(I - P_{11}\Delta)^{-1}P_{12}, \quad (\text{П.95})$$

където индексът  $\ell$  означава долно, а индексът  $u$  – горно преобразувание. По-нататък с  $R$  се означава матричната функция, която се получава като резултат от линейно дробно преобразувание.



Фигура П.5. (а)  $R_1$  като долно ЛДП в термините на  $K$ . (б)  $R_2$  като горно ЛДП в термините на  $\Delta$

Долното дробно преобразувание  $F_\ell(P, K)$  е предавателната матрица  $R_1$ , получена като резултат от обхващането с (положителна) обратна връзка  $K$  на долната част на  $P$ , както е илюстрирано на фиг. П.5 (а). За да се види това, обърнете внимание, че блок-схемата на фиг. П.5 (а) може да се запише като

$$z = P_{11}w + P_{12}u, \quad v = P_{21}w + P_{22}u, \quad u = Kv. \quad (\text{П.96})$$

След елиминирание на  $v$  и  $u$  от тези уравнения, получаваме

$$z = R_1 w = F_\ell(P, K)w = [P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}]w. \quad (\text{П.97})$$

Казано накратко,  $R_1$  се записва като долно ЛДП на  $P$  в термините на параметъра  $K$ . По подобен начин, на фиг. П.5 (б) е илюстрирано горното ЛДП  $R_2 = F_u(P, \Delta)$ , получено чрез обхващане с (положителна) обратна връзка  $\Delta$  на горната част на  $P$ .

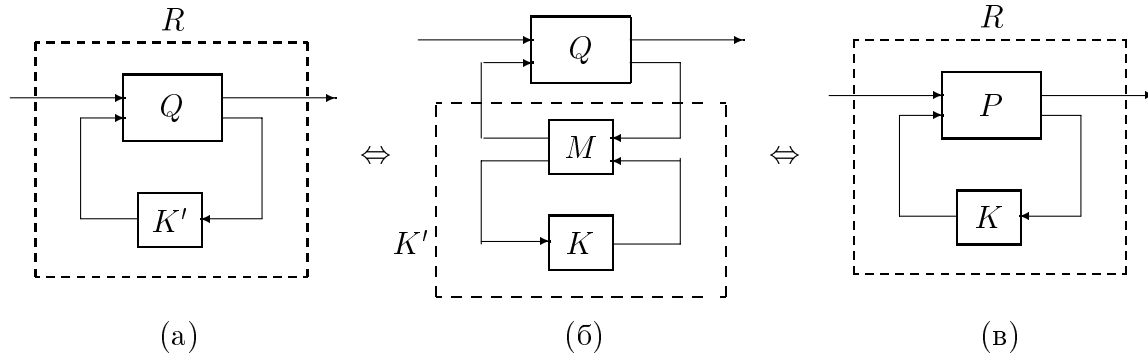
### Съединения на линейни дробни преобразувания

Важно свойство на ЛДП е, че кое да е съединение на ЛДП е отново ЛДП. Нека разгледаме фиг. П.6, където  $R$  е записано в термините на долно ЛДП на  $K'$ , което от своя страна е отново долно ЛДП на  $K$  и желаем да изразим  $R$  директно като ЛДП на  $K$ . Имаме, че

$$R = F_\ell(Q, K') \quad \text{където} \quad K' = F_\ell(M, K) \quad (\text{П.98})$$

и искаме да получим  $P$  (в термините на  $Q$  и  $M$ ), така че

$$R = F_\ell(P, K). \quad (\text{П.99})$$



Фигура П.6. Съединението на ЛДП дава ЛДП

Намираме, че

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} Q_{11} + Q_{12}M_{11}(I - Q_{22}M_{11})^{-1}Q_{21} & Q_{12}(I - M_{11}Q_{22})^{-1}M_{12} \\ M_{21}(I - Q_{22}M_{11})^{-1}Q_{21} & M_{22} + M_{21}Q_{22}(I - M_{11}Q_{22})^{-1}M_{12} \end{bmatrix}. \quad (\text{П.100})$$

Подобни изрази са в сила, когато се използват горни ЛДП. За

$$R = F_u(M, \Delta') \quad \Delta' = F_u(Q, \Delta) \quad (\text{П.101})$$

получаваме  $R = F_u(P, \Delta)$ , където  $P$  се дава в термините на  $Q$  и  $M$  чрез (П.100).

### Връзка между $F_\ell$ и $F_u$

Очевидно е, че  $F_\ell$  и  $F_u$  са тясно свързани. Ако познаваме  $R = F_\ell(M, K)$ , то можем директно да получим  $R$  в термините на горно преобразувание на  $K$  чрез преподреждане на  $M$ . Имаме, че

$$F_u(\tilde{M}, K) = F_\ell(M, K), \quad (\text{П.102})$$

където

$$\tilde{M} = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} M \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}. \quad (\text{П.103})$$

**Обратно ЛДП**

При предположението, че всички съответни обратни матрици съществуват, имаме че

$$(F_\ell(M, K))^{-1} = F_\ell(\tilde{M}, K), \quad (\text{П.104})$$

където  $\tilde{M}$  се дава от

$$\tilde{M} = \begin{bmatrix} M_{11}^{-1} & -M_{11}^{-1}M_{12} \\ M_{21}M_{11}^{-1} & M_{22} - M_{21}M_{11}^{-1}M_{12} \end{bmatrix}. \quad (\text{П.105})$$

# Библиография

- [1] Гантмахер, Ф.Р. (1967). *Теория матриц*, 3-то изд. Наука, М.
- [2] Глазман, И.М., Ю.И. Любич (1969). *Конечномерный линейный анализ*. Наука, М.
- [3] Квакернаак, Х., Р. Сиван (1977). *Линейные оптимальные системы управления*. Мир, М.
- [4] Уилкинсон, Дж.Х. (1970). *Алгебраическая проблема собственных значений*. Наука, М.
- [5] Уонэм, У.М. (1980). *Линейные многомерные системы управления*. Наука, М.
- [6] Хорн, Р., Ч. Джонсон, Ч. (1989). *Матричный анализ*. Мир, М.
- [7] Balas, G.J., J.C. Doyle, K. Glover, A. Packard and R. Smith (1993).  *$\mu$ -Analysis and Synthesis Toolbox User's Guide*. MathWorks, Natick, Mass.
- [8] Boyd, S., L.E. Ghaoui, E. Feron and V. Balakrishnan (1994). *Linear Matrix Inequalities in System and Control Theory*. SIAM, Philadelphia.
- [9] Chaitin–Chatelin, F., V. Frayssé (1996). *Lectures on Finite Precision Computations*. SIAM, Philadelphia.
- [10] Chiang, R.Y. and M.G. Safonov (1992). *Robust Control Toolbox User's Guide*. MathWorks, Natick, Mass.
- [11] Doyle, J.C., B. Francis and A. Tannenbaum (1992). *Feedback Control Theory*. Macmillan Publishing Company, N.Y.
- [12] Doyle, J.C., K. Glover, P.P. Khargonekar and B.A. Francis (1989). State-space solutions to standard  $\mathcal{H}_2$  and  $\mathcal{H}_\infty$  control problems. *IEEE Transactions on Automatic Control*, **AC-34**, 831–847.
- [13] Francis, B.A. (1987). *A Course in  $\mathcal{H}_\infty$  Control*. Springer–Verlag, Berlin.
- [14] Goffe, W.L., G.D. Ferrier and J. Rogers (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, **60**, 65–69.
- [15] Golub, G.H. and C.V. Van Loan (1989). *Matrix Computations*, 2nd ed. The John Hopkins University Press, Baltimore, Maryland.
- [16] Green, M. and D.J.N. Limebeer (1995). *Linear Robust Control*. Prentice–Hall, Englewood Cliffs, N.J.

- [17] Higham, N.J. (1993). *The Test Matrix Toolbox for Matlab*. Num. Anal. Rep. No. 237, Dept. of Mathematics, Univ. of Manchester.
- [18] Higham, N.J. (1996). *Accuracy and Stability of Numerical Algorithms*. SIAM, Philadelphia.
- [19] Kailath, T. (1980). *Linear Systems*. Prentice–Hall, Englewood Cliffs, N.J.
- [20] Maciejowski, J.M. (1989). *Multivariable Feedback Design*. Addison-Wesley, Wokingham, U.K.
- [21] Petkov, P.Hr., N.D. Christov and M.M. Konstantinov (1991). *Computational Methods for Linear Control Systems*. Prentice Hall, Hemel Hempstead, U.K.
- [22] Skogestad, S. and I. Postlethwaite (1996). *Multivariable Feedback Control*. John Wiley & Sons, Chichester, U.K.
- [23] Stewart, G.W., J.–G. Sun (1990). *Matrix Perturbation Theory*. Academic Press, Boston.
- [24] Stoorvogel, A.A. (1992). *The  $\mathcal{H}_\infty$  Control Problem: A State Space Approach*. Prentice–Hall, Englewood–Cliffs, N.J.
- [25] Zhou, K., J.C. Doyle and K. Glover (1996). *Robust and Optimal Control*. Prentice–Hall, Upper Saddle River.