

4.5. Модели редуцированного порядка

Проблема построения методами идентификации приближенных моделей сложных систем тесно связана с хорошо известной проблемой аппроксимации системы с большим числом внутренних переменных, т.е. описываемых дифференциальными уравнениями большой размерности, с помощью более простых систем уравнений существенно меньшей размерности. Она получила название проблема редукции порядка модели для систем большой размерности. Методы малого параметра, которые развивались математиками, позволяли корректно находить приближенные решения сложных задач на основе более простых уравнений, что нашло широкое применение во многих прикладных задачах. Малые параметры выделялись в явном виде, как правило, при переходе к безразмерной форме записи уравнений.

В методах редукции порядка модели, которые будут далее рассмотрены, малые параметры носят скрытый характер и определяются в результате сравнения поведения систем не по внутренним переменным, а по наблюдаемым на выходе при одном и том же произвольном входном воздействии. Близость выходов оценивается по одной из норм, которая задается конкретно решаемой задачей. При этом малые параметры определяются инвариантами наблюдаемой системы, т.е. значениями f_{mp}^{rc} , f_{mp}^{rs} и α_p , β_p (см. раздел 3.1 формулы (3.3), (3.4)). Из общих соображений можно предположить, что им соответствуют плохо наблюдаемые и управляемые моды с малыми значениями f_{mp}^{rc} , f_{mp}^{rs} или группа мод с близкими значениями α_p , β_p .

Редукция порядка модели сложной системы при известном ее описании в виде системы уравнений (4.1) или (4.2) большой размерности может быть осуществлена различными методами, обзор которых можно найти, например, в [34, 35].

Большинство из них группируются вокруг двух подходов: 1) методы, основанные на подпространстве Крылова, выделяемом с помощью практического ранга матриц формируемых по правилу

$$K_j(H, h) = [h, Hh, H^2h, \dots, H^{j-1}h],$$

где H – матрица, а h – вектор;

2) методы, основанные также на установлении практического ранга матриц с помощью SVD-преобразования и последующем переходе к подпространствам меньшей размерности без существенного изменения динамических свойств исходной системы. Это достигается за счет малых сингулярных чисел.

Подпространства Крылова, как правило, связываются с передаточными функциями (матрицами) систем (4.1), имеющими хорошо известный вид

$$Q(s) = C(sI - A)^{-1}B.$$

Функции (матрицы) $Q(s)$ представляются в виде моментных разложений в окрестности некоторой точки s_0 или вблизи нуля. Соответствующие моменты в этих разложениях представляются затем в форме $K_j(H, h)$, по которым находятся подпространства Крылова. Подробности описаны в [34, 35]. В случае систем с одним входом и одним выходом $Q(s)$ есть передаточная функция. Оценку усеченных моделей можно делать по Боде диаграммам. В общем случае используются оценки ошибок аппроксимации по нормам $\|Q_n(s) - Q(s)\|$ в соответствующих пространствах. При этом качество аппроксимации целесообразно оценивать по формулам (1.42) главы 1.

Задачу редукции порядка модели можно свести к оптимальной аппроксимации. Чтобы при этом оставаться во временной области, воспользуемся описанием в форме (3.3) при нулевых начальных условиях, т.е. $f_{mp0}^{rc} = f_{mp0}^{rs} \equiv 0$. Пусть размерность исходной системы определяется достаточно большим значением P в (3.3). Редукцию порядка модели данной системы будем проводить независимо для каждого входа и выхода, т.е. по скалярным уравнениям, связывающим отдельно выход y_m ($m = \overline{1, M}$) со входом u_r ($r = \overline{1, R}$). При этом минимально управляемая и наблюдаемая модель соответствующая этим входу и выходу будет определяться числом ненулевых f_{mp}^{rc} , f_{mp}^{rs} , а для действительных собственных значений еще с учетом того, что $\beta_p = 0$. Если при этом ее размерность окажется малой, то в принципе редукцию можно не проводить. Понижать порядок следует для тех входов и выходов, для которых минимальная модель имеет достаточно большую размерность. В некоторых случаях при больших M и R может оказаться, что для всех входов и выходов получаем минимальные модели небольшой размерности. Это означает, что система распадается на множество слабо связанных подсистем небольшой размерности и нет необходимости ее редуцировать. На тех входах и выходах, для которых редукция желательна, осуществить ее можно с помощью нижеследующих действий. В момент времени t_0 подадим на вход исходной модели и искомой редуцированной входное воздействие в виде δ -функции, т.е. $u(\theta) = \delta(\theta - t_0)$. Для простоты изложения подхода опустим индексы m и r , связанные с конкретным входом и выходом, и будем рассматривать односвязные системы имея в виду одно из уравнений многосвязной системы (3.3). Кроме того, не ограничивая общности, полагаем $t_0 = 0$.

Воспользуемся результатами работы [16], где записаны соотношения вход-выход с

учетом наличия кратных собственных значений, которые имеют вид

$$y(t) = \sum_{p=1}^P \sum_{j=1}^{n_p} \int_0^t \frac{(t-\theta)^{j-1}}{(j-1)!} \cdot [f_{pj}^c \cos \beta_p(t-\theta) + f_{pj}^s \sin \beta_p(t-\theta)] e^{-\alpha_p(t-\theta)} u(\theta) d\theta, \quad (4.27)$$

где n_p – кратность p -ого собственного значения.

После воздействия δ -импульса в асимптотически устойчивой системе происходит процесс релаксации, определяемый формулой

$$y(t) = \sum_{p=1}^P \sum_{j=1}^{n_p} \frac{t^{j-1}}{(j-1)!} (f_{pj}^c \cos \beta_p t + f_{pj}^s \sin \beta_p t) e^{-\alpha_p t}, \quad (4.28)$$

в которой $y(t)$ фактически является импульсной переходной функцией.

Выходной сигнал или просто импульсная переходная функция для моментов времени $t_k = \Delta k$ ($k = 0, 1, 2, \dots, N$) примет выражение

$$y_k = \sum_{p=1}^P \sum_{j=1}^{n_p} \frac{(\Delta k)^{j-1}}{(j-1)!} (f_{pj}^c \cos k \Delta \beta_p + f_{pj}^s \sin k \Delta \beta_p) e^{-k \alpha_p \Delta}. \quad (4.29)$$

При выбранном Δ из (4.29) получаем последовательность выходных данных $\{y_k\}$, которая будет использоваться для построения аппроксимирующей их модели более низкого порядка. Для этого зададимся классом моделей, в котором предполагаем ее искать. Предлагается в качестве такового взять (3.3), в котором отсутствуют кратные корни. В принципе можно не исключать кратные корни, но руководствуясь желанием получить более простые модели, все же их исключим. Тем более, что кратные корни можно аппроксимировать близкими собственными значениями выбирая соответствующим образом другие параметры модели. Редуцированная модель в указанном классе будет давать отклик в тех же точках и при том же воздействии в виде δ -функции согласно формуле

$$\hat{y}_k = \sum_{q=1}^Q (f_q^c \cos k \Delta \beta_q + f_q^s \sin k \Delta \beta_q) e^{-k \alpha_q \Delta}, \quad Q \ll \sum_{p=1}^P n_p. \quad (4.30)$$

Параметры усеченной модели, включая Q , будем определять из решения следующей оптимизационной задачи:

$$\frac{1}{N} \left\| \sum_{k=0}^N (y_k - \hat{y}_k) \right\| \rightarrow \min. \quad (4.31)$$

При этом решается последовательность задач минимизации, в которых величина Q каждый раз увеличивается на единицу, начиная с $Q=1$. В качестве решения задачи редукции берется то, которое при минимальном Q обеспечивает допустимую погрешность аппроксимации, определяемую условием

$$\frac{1}{N} \left\| \sum_{k=0}^N (y_k - \hat{y}_k) \right\| \leq \varepsilon, \quad (4.32)$$

где величина ε задается, а $\|\cdot\|$ – одна из норм, представленных в (4.16).

В общем случае необходимые условия оптимальности в такой задаче сводятся к нелинейным уравнениям. Более того, в ряде случаев может проявиться мультимодальность функции, экстремум которой ищется. Поэтому решение задачи (4.31) особенно при больших значениях Q может вызвать определенные трудности. Однако, принимая во внимание априорное знание распределения на комплексной плоскости собственных значений и инвариантов f_p^c , f_p^s , а также некоторые особенности оптимизационной задачи, можно построить достаточно эффективные упрощенные алгоритмы ее решения.

Рассмотрим, в частности, задачу минимизации с квадратичной нормой $\|\cdot\|_2$. В этом случае (4.31) фактически есть нелинейный метод наименьших квадратов (МНК) [36]. Когда собственные значения нам известны, имеем обычный МНК и инварианты f_p^c и f_p^s определяются из решения стандартной линейной системы алгебраических уравнений классического МНК. Это обстоятельство можно использовать при решении нелинейным МНК для уменьшения числа варьируемых переменных, что особенно важно в итерационной процедуре нахождения размерности Q , обеспечивающего выполнение (4.32). При квадратичной норме в (4.31) условие (4.32) может быть записано для другой нормы, например $\|\cdot\|_\infty$, что допустимо, когда ищется приближенное решение с использованием итераций. А поскольку при этом определяется наименьшая размерность Q , итеративный процесс целесообразно начинать с $Q=1$ и полученное для него решение проверить на выполнимость условия (4.32) в какой бы норме оно не было записано. Если оно не выполняется, то берем следующее значение $Q=2$ и осуществляем те же действия, что и для $Q=1$. Невыполнение условия (4.32) для решения при $Q=2$ означает продолжение итераций с повышением на единицу порядка искомой редуцированной модели. Процесс этот прекращается, когда при очередном повышении порядка выполнится неравенство (4.32), а соответствующая этой размерности полученная модель и будет решением задачи редукции. Когда для каждого Q решение находится нелинейным МНК [36], целесообразно его модифицировать, встраивая в него процедуру вычисления оптимальных f_q^c , f_q^s на каждой итерации собственных значений. Оптимальные значения собственных чисел следует находить в той области комплексной

плоскости, в которой содержатся все собственные значения исходной системы большой размерности.

При мультимодальности целевой функции целесообразно привлекать методы стохастической оптимизации, в частности, метод дифференциальной эволюции (МДЭ) с использованием генетических алгоритмов [37]. Достоинство этих методов прежде всего в том, что они позволяют находить глобальный экстремум и применять их в тех случаях, когда сталкиваемся с недифференцируемостью целевых функций, например, для критериев с нормой $\|\cdot\|_1$. В случаях квадратичного критерия процедуру изменения f_q^c, f_q^s можно осуществлять не псевдослучайным образом, а вычислять их с помощью МНК, т.е. брать оптимальными в каждой варьируемой точке собственных значений. Это позволит вдвое уменьшить число искомых параметров на каждой итерации МДЭ.

Эффективность описанного выше метода падает с увеличением Q , поскольку приходится решать оптимизационную задачу большой размерности. Поэтому при относительно больших Q предлагается иной способ ее решения. Из последовательности $\{y_k\}$ исходной модели большой размерности формируем ганкелеву матрицу

$$Y = \begin{bmatrix} y_0 & y_1 & \dots & y_K \\ y_1 & y_2 & \dots & y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_K & y_{K+1} & \dots & y_{2K} \end{bmatrix}, \quad (4.33)$$

где $K = \frac{N}{2}$, если N – четное и $K = \frac{N-1}{2}$ в противном случае. Размерность K и соответственно N следует выбирать такими, чтобы искомый порядок Q редуцированной модели удовлетворял условию $Q < K$.

Матрица Y в соответствии с теорией не зависит от того, какая реализация в пространстве состояний была у системы в момент подачи δ -импульса: жорданова или какая-то иная, т.е. матрица Y инвариантна относительно неособого преобразования. Это означает, что свободное движение, которое генерировало последовательность $\{y_k\}$ должно удовлетворять не только (4.29), но и системе уравнений

$$\begin{aligned} x_{k+1} &= Ax_k, \\ y_k &= c^T x_k, \quad x(0) = x_0, \end{aligned} \quad (4.34)$$

где вектор-строка c^T и матрица A задают некоторую эквивалентную (4.29) реализацию исходной модели в пространстве состояний соответствующей размерности. Начальное состояние x_0 выражается с помощью неособого преобразования через вектор,

компонентами которого являются инварианты f_{pj}^c, f_{pj}^s . По аналогии с тем, как это делалось при получении (3.26) на основе уравнений (3.25), вытекающих из (4.34), можно получить следующее матричное уравнение:

$$Y = \Gamma \cdot X, \quad (4.35)$$

где Y – матрица (4.33), матрица наблюдаемости Γ и матрица X имеют вид

$$\Gamma = \begin{bmatrix} c^T \\ c^T A \\ \vdots \\ c^T A^{K-1} \end{bmatrix}, \quad X = [x_0 \ x_1 \ \dots \ x_K]. \quad (4.36)$$

Далее производим SVD-разложение матрицы Y , которое записывается как

$$Y = Q \Sigma V^T, \quad (4.37)$$

где Q и V – ортогональные матрицы, а Σ – матрица сингулярных чисел, расположенных на диагонали в невозрастающем порядке. Матрицы Q, Σ, V , полученные при SVD-разложении, разобьем на блоки

$$Q = [Q_r \ Q_u], \quad \Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & \Sigma_u \end{bmatrix}, \quad V = [V_r \ V_u], \quad (4.38)$$

где Q_r, Σ_r, V_r – матрицы, соответствующие редуцированной модели, а Q_u, Σ_u, V_u – матрицы, которые определяют динамику, не попавшую в усеченную модель. Тогда матрица Ганкеля, составленная из выходов редуцированной модели, примет вид

$$Y_r = Q_r \Sigma_r V_r^T. \quad (4.39)$$

Дальнейшие действия аналогичны описанным в разделе 3.1 главы 3. Представим матрицу Y_r в виде произведения двух полноранговых матриц

$$Y_r = Y_{r1} \cdot Y_{r2}, \quad (4.40)$$

где $Y_{r1} = Q_r, Y_{r2} = \Sigma_r V_r^T$.

Матричное уравнение (4.35) для редуцированной модели примет вид

$$Y_r = \Gamma_r \cdot X_r. \quad (4.41)$$

В соответствии с теорией реализаций [20] можно положить

$$\Gamma_r = Y_{r1} = Q_r, \quad X_r = Y_{r2} = \Sigma_r V_r^T, \quad (4.42)$$

задавая таким образом некоторую реализацию системы в пространстве состояний.

Знание матрицы наблюдаемости для полученной в результате реализации, позволяет найти для нее матрицу A . Как видно из выражения матрицы Γ_r в (4.36) она обладает свойством сдвиговой инвариантности, из которого непосредственно следует

следующее матричное приближенное уравнение:

$$Y_{r12} \approx Y_{r11} \cdot A, \quad (4.43)$$

в котором матрица Y_{r12} получается из матрицы Y_{r1} вычеркиванием из нее первой строки, а матрица Y_{r11} – вычеркиванием последней строки. Предполагается, что размерность искомой редуцированной модели меньше размерности матрицы Y . Тогда имеем переопределенную систему уравнений относительно элементов матрицы A . Для ее решения можно использовать МНК или обобщенный МНК [23].

После нахождения матрицы A вычисляются ее собственные значения. Это позволяет построить матрицу наблюдаемости Γ' для жордановой реализации, в которой матрица A' будет иметь вид

$$A' = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_{Q_1}, J_1, J_2, \dots, J_{Q_2}],$$

где λ_q ($q = \overline{1, Q_1}$) действительные собственные значения, а $J_q = \begin{bmatrix} -\alpha_q & \beta_q \\ -\beta_q & -\alpha_q \end{bmatrix}$ ($q = \overline{1, Q_1}$)

представляет диагональные блоки соответствующие комплексным корням. Для этой реализации построим матрицу Γ' так, чтобы она соответствовала наблюдаемой канонической реализации, в которой вектор c' имеет компоненты

$$c'_1 = c'_2 = \dots = c'_{Q_1} = 1, \quad c_1^c = c_2^c = \dots = c_{Q_2}^c = 1, \quad c_1^s = c_2^s = \dots = c_{Q_2}^s = 0.$$

Здесь c'_q ($q = \overline{1, Q_1}$) – компоненты вектора c' соответствующие действительным собственным значениям, а пары $(c_q^c, c_q^s)^T = (1, 0)^T$ – комплексным, т.е. привязаны к блокам J_q . Таким образом, для этой реализации в системе уравнений (4.34) заданы матрица A' и вектор $(c')^T$, а начальное состояние x'_0 согласно (4.30) и (3.4) связано с вектором b' соотношениями

$$\begin{aligned} x'_{01} &= f_1 = b'_1, \quad x'_{02} = f_2 = b'_2, \quad \dots, \quad x'_{0Q_1} = f_{Q_1} = b'_{Q_1}, \\ x'_{0Q_1+1} &= f_1^c = b_{Q_1+1}^c, \quad x'_{0Q_1+1} = f_1^s = -b_{Q_1+1}^s, \quad \dots, \\ x'_{0Q_1+Q_2} &= f_{Q_2}^c = b_{Q_1+Q_2}^c, \quad x'_{0Q_1+Q_2} = f_{Q_2}^s = -b_{Q_1+Q_2}^s. \end{aligned} \quad (4.44)$$

Следовательно, чтобы получить в окончательном виде редуцированную модель заданной размерности для жордановой канонической реализации осталось найти компоненты вектора x'_0 . Для этого найдем сначала из уравнения $\Gamma = \Gamma' \cdot T$ матрицу перехода T от реализации определяемой (4.42) к канонической жордановой реализации, которая определяется формулой

$$T = (\Gamma')^{-1} \cdot \Gamma,$$

где Γ' выражается через матрицу A' и вектор c' . Первый столбец матрицы Y_{r2} согласно

(4.32) равен вектору x_0 соответствующей реализации, в которой была получена матрица A из (4.43). Вектор x'_0 , через который выражаются компоненты вектора b' , определяется равенством

$$x'_0 = Tx_0. \quad (4.45)$$

Завершается данный способ после нахождения из (4.44) и (4.45) вектора b' переходом от канонической жордановой реализации к той, которая необходима для дальнейшего применения.

Знание собственных значений редуцированной модели позволяет применить для определения инвариантов иной подход, в основе которого лежит МНК. Будем рассматривать их нахождение как задачу аналитической конечномерной аппроксимации в виде разложения по заданному (известному) базису табличной функции $\{y_k\}$, определяющей выход исходной модели на интервале наблюдения $[0, N]$. Поскольку $2Q < N$, то имеем задачу, в которой требуется найти решение переопределенной и в общем несовместной системы линейных алгебраических уравнений следующего вида:

$$R_N f = y, \quad (4.46)$$

где матрица R_N имеет размерность $N \times Q$ и состоит из элементов, определяемых значениями базисных собственных функций в точках k ($0, 1, 2, \dots, N$), а f вектор, компонентами которого являются инварианты f_q^c и f_q^s . Правая часть задается значениями y_k в тех же точках k , соответствующих выходу исходной системы большой размерности. Решение (4.46) будет приближенным, которое может быть найдено с помощью МНК, минимизирующего невязку в евклидовой норме. Приближенное решение (4.46), получаемое МНК, записывается как

$$f = (R_N^T \cdot R_N)^{-1} \cdot R_N^T y. \quad (4.47)$$

Здесь в отличие от стандартных задач, в которых несовместность обусловлена приближенно заданной правой частью или самой матрицей R_N , правая часть задана точно. Поэтому речь идет о приближенном решении, минимизирующем невязку $R_N f - y$. В определенном смысле можно говорить о приближенно заданной в (4.46) матрице R_N , хотя в принципе не существует точного значения R_N , при котором система была бы совместной. Приближенное решение такой несовместной системы может быть найдено также с помощью обобщенного МНК, который сводится к оптимизационной задаче

$$\min_F v^T v, \quad (4.48)$$

где $F = \{x, v : y = R_N f + Bv\}$, а B – заданная матрица. В данном случае Bv – есть обобщенная невязка. Решается (4.48) с использованием ортогональных преобразований, в частности, QR-разложения матрицы R_N .

Одним из ключевых моментов обоих вышеописанных подходов является определение размерности Q . Поэтому более подробно опишем алгоритм его нахождения в итеративной форме. Исходным или начальным является выбор размерности матрицы A , когда осуществляется разбиение на блоки сингулярного разложения матрицы Ганкеля. Поскольку сингулярные числа расположены в убывающем (невозрастающем) порядке, то в блоке Σ_u оставляем все несущественные, т.е. достаточно малые значения σ_i . В ряде случаев это сделать легко, например, когда после некоторого значения сингулярного числа на диагонали наблюдается выраженный скачок на порядок и более следующего за ним σ_i , а до этого изменения были малосущественными. В других случаях приходится принимать волевые решения, сравнивая последовательно значения сингулярных чисел с первым значением. Усечение можно делать там, где последующие значения σ_i отличаются более, чем на порядок от первого. Тогда размерность блока Σ_r , состоящего из существенных сингулярных чисел, и будет определять начальное Q_0 . После нахождения из (4.47) вектора f при известных собственных значениях формируем импульсную функцию редуцированной модели и последовательность $\{\hat{y}_k\}$. Проверяем для заданной нормы выполнение условия (4.32). Возможны три исхода. Выполняются равенство или строгое неравенство, неравенство не выполняется. В первом случае редуцированная модель размерности Q_0 построена. В двух других случаях следует продолжить поиск решения, т.е. перейти к следующей итерации. При выполнении строгого неравенства размерность Q_0 следует уменьшить на единицу, а при его невыполнении, наоборот, на единицу увеличиваем Q_0 . Далее повторяются те же действия, которые проводились после выбора Q_0 , а именно: находится матрица A_1 и ее собственные значения, решается задача определения инвариантов f_q^c , f_q^s и оценивается выполнимость неравенства (4.32). В зависимости от полученного результата итерации прекращаются или продолжаются. Они прекращаются в тех случаях, когда получилось равенство или произошла смена знака неравенства (4.32) по отношению к случаю с Q_0 . При равенстве размерность редуцированной модели равна $Q_1 = Q_0 \pm 1$ в зависимости от того, увеличивалось или уменьшалось Q_0 . При смене знака неравенства с «>» на «<» берется значение

$Q_1 = Q_0 + 1$, а при смене знака с « $<$ » на « $>$ » остается значение Q_0 . В остальных случаях размерность соответственно увеличивается или уменьшается на единицу и итерации продолжаются до тех пор, пока не получится равенство или смена знака. Окончательное значение Q берется для последней итерации (« $>$ » поменялось на « $<$ ») или предыдущей (« $<$ » поменялось на « $>$ »).

Одним из важнейших параметров метода является шаг дискретизации Δ . Его следует выбирать согласовано со значениями $\max \alpha_p$ и $\max \beta_p$ ($p = \overline{1, P}$). Будем иметь приемлемую аппроксимацию во всем динамическом диапазоне, если будут выполняться соотношения $\Delta < \frac{\pi}{5\beta_{p\max}}$, $\Delta < \frac{2}{5\alpha_{p\max}}$.

Замечание 4.1. Если в описанном подходе оценка (4.32) делается по норме отличной от $\|\cdot\|_2$, например, по норме $\|\cdot\|_\infty$, то несмотря на эквивалентность норм можем получить редуцированную модель завышенного порядка. В принципе существует возможность с помощью обобщенного МНК, т.е. на основе (4.47), улучшить результат. Для этого, сравнивая поточечно графики импульсной переходной функции исходной и полученной МНК редуцированной моделей, можно в интерактивном режиме подобрать такую матрицу B , с помощью которой уменьшить погрешность аппроксимации по норме $\|\cdot\|_\infty$ по сравнению с МНК.

Для многосвязной системы решать задачу редукции описанным подходом следует в два этапа. Пусть требуется ее решить в следующей постановке. Построить редуцированную модель минимально возможной размерности так, чтобы для каждой импульсной переходной функции, являющейся элементом импульсной переходной матрицы аппроксимирующей системы, выполнялось условие (4.32). Тогда на первом этапе независимо вышеописанным методом решается задача редукции для каждого элемента этой матрицы, т.е. для соответствующей ему передаточной функции. На втором этапе проводится сравнение найденных для них собственных значений. Если все они оказались существенно разными, то по найденным решениям достаточно просто восстанавливается редуцированная модель в пространстве состояний для жордановой реализации многосвязной системы. Процедура ее построения описана в главе 3. Если же для некоторых элементов импульсной переходной матрицы имеются равные или близкие собственные значения, то целесообразно провести их коррекцию с учетом того как вблизи них располагаются собственные значения исходной системы и какие значения принимают соответствующие им инварианты. После этого повторно решить задачу определения

Продemonстрируем некоторые особенности изложенного подхода. Как соотносятся между собой собственные значения исходной и редуцированной модели показано на рис. 4.7 и рис. 4.8. Собственные числа исходной системы представлены в устойчивой области комплексной плоскости в виде кружочков, а редуцированной – в виде крестиков. Результаты, показанные на рис. 4.7, получены для исходной модели одиннадцатого порядка при выбранной размерности редуцированной модели равной трем. А на рис. 4.8 для той же исходной модели показано как располагаются собственные значения усеченной модели четвертого порядка по отношению к реальной.

Figure 1 is a scatter plot showing the relationship between the real part of z ($\text{Re } z$) on the x-axis and the imaginary part of z ($\text{Im } z$) on the y-axis. The x-axis ranges from -0.07 to 0, and the y-axis ranges from -0.2 to 0.2. The plot features a grid with major lines every 0.01 units on the x-axis and every 0.05 units on the y-axis. Data points are represented by open circles and crosses. The points are distributed as follows:

Symbol	$\text{Re } z$	$\text{Im } z$
Circle	-0.065	0.00
Circle	-0.032	0.00
Circle	-0.022	0.09
Circle	-0.022	-0.09
Circle	-0.005	0.15
Circle	-0.005	0.12
Circle	-0.005	0.08
Circle	-0.005	-0.08
Circle	-0.005	-0.12
Circle	-0.005	-0.15
Circle	0.000	0.15
Circle	0.000	0.12
Circle	0.000	0.08
Circle	0.000	-0.08
Circle	0.000	-0.12
Circle	0.000	-0.15
Cross	-0.005	0.15
Cross	-0.005	0.00
Cross	-0.005	-0.15

11

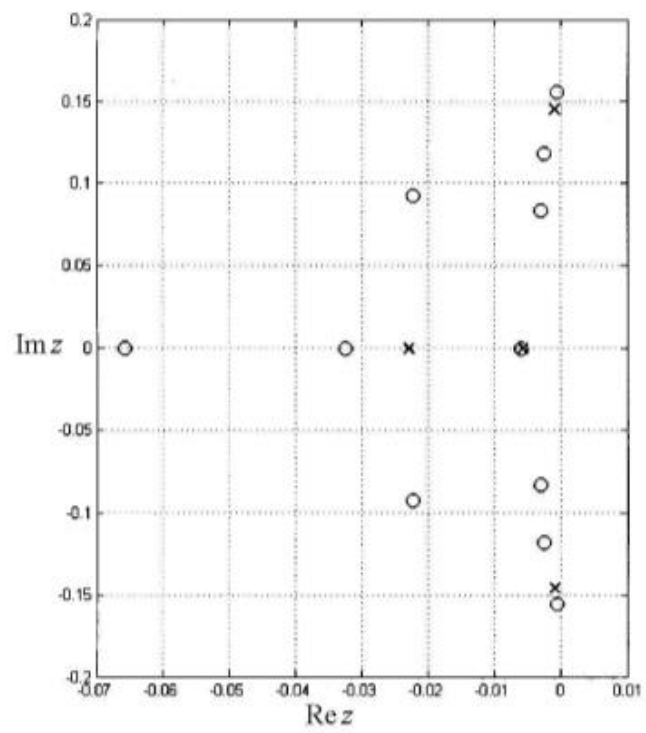


Рис. 4.8

5.2 Непараметрическая идентификация

При эквивалентном представлении моделей в главе 3 было показано, что одна и та же система может иметь разное описание. Зная одно из них всегда с помощью описанных в этой главе преобразований можно перейти к любому другому. В частности, линейную стационарную систему можно описать с помощью передаточных функций или импульсными переходными матрицами. Прямые методы определения этих функций по входным и выходным данным без первоначального выделения классов возможных моделей и составляют непараметрическую идентификацию, поскольку здесь не используются конечномерные вектора параметров при поиске наилучшего описания. Фактически это означает переход от пары данных (y, u) , полученных в экспериментах, к одной функции, которая как и исходные имеет неаналитическое представление. Это возможно, когда описание системы записывается в форме соотношений вход-выход как во временной, так и в частотных областях или в пространстве изображений. Суть метода продемонстрируем на простом примере, когда вход и выход есть скалярные функции и связаны между собой хорошо известным соотношением

$$y(t) = \sum_{k=1}^{\infty} h_k u(t-k) + v(t). \quad (5.6)$$

Предполагаем, что входное воздействие задано точно, т.е. не содержит возмущений. Тогда $v(t)$ есть шум наблюдений. Если в момент $t=0$ на вход такой системы подать входное воздействие в виде единичной функции, а именно

$$u(t) = \begin{cases} 1, & t = 0, \\ 0, & t \neq 0, \end{cases}$$

и при $t < 0$ система находилась в покое, то

$$y(t) = h_t + v(t). \quad (5.7)$$

Из (5.7) получаем оценку импульсной переходной функции

$$\hat{h}_t = y(t)$$

с той же самой ошибкой $v(t)$, что и у выходной переменной. Таким образом, мы перешли от пары исходных данных (y, u) к одной, которая определяет непараметрическое поведение импульсной переходной функции. В том случае, когда шум задан интервальной оценкой $|v(t)| \leq \varepsilon$, для импульсной переходной функции также получаем интервальную оценку $[\hat{h}_t - \varepsilon, \hat{h}_t + \varepsilon]$, гарантировано содержащую точное значение импульсной переходной функции.

Если имеем ступенчатое входное воздействие вида

$$u(t) = \begin{cases} 1, & t \geq 0, \\ 0, & t < 0, \end{cases}$$

и до момента его приложения система была в покое, то выходная переменная принимает значение

$$y(t) = \sum_{k=1}^t h_k + v(t). \quad (5.8)$$

Откуда получаем оценку

$$\hat{h}_t = y(t) - y(t-1)$$

с ошибкой $\xi(t) = v(t) - v(t-1)$.

При той же интервальной погрешности $v(t)$ гарантированный интервал принадлежности точного значения определяется как

$$[\hat{h}_t - 2\varepsilon, \hat{h}_t + 2\varepsilon].$$

При стохастической непараметрической идентификации используется входное воздействие, являющееся квазистационарной последовательностью с заданной функцией ковариации процесса

$$\bar{E}[u(t) \cdot u(t-\tau)] = R_u(\tau).$$

Предполагается также, что шум $v(t)$ есть независимый от входа случайный процесс, так что

$$\bar{E}[u(t) \cdot v(t-\tau)] = 0.$$

Тогда из (5.6) имеем

$$\bar{E}[y(t) \cdot u(t-\tau)] = R_{yu}(\tau) = \sum_{k=1}^{\infty} h_k R_u(k-\tau). \quad (5.9)$$

Уравнение (5.9) позволяет определить импульсную переходную функцию h_k . Когда входное воздействие представляет собой белый шум с

$$R_u(\tau) = \sigma \delta_{\tau_0},$$

где δ_{τ_0} – индекс Кронекера, который равен 0, если $\tau \neq 0$, а σ – дисперсия белого шума, тогда

$$g_{\tau} = \frac{1}{\sigma} R_{yu}(\tau). \quad (5.10)$$

Соответственно, можно из (5.10) получить оценку импульсной переходной функции, если в правой части этого уравнения вместо R_{yu} взять

$$\hat{R}_{yu}^N(\tau) = \frac{1}{N} \sum_{t=\tau}^N y(t) \cdot u(t-\tau). \quad (5.11)$$

Точно так же можно поступить с уравнением (5.8), если в нем кроме (5.10) использовать $\hat{R}_u(\tau)$, определяемую формулой

$$\hat{R}_u(\tau) = \frac{1}{N} \sum_{t=\tau}^N u(t) \cdot u(t - \tau). \quad (5.12)$$

Тогда оценку импульсной переходной функции можно получить, решив уравнение

$$\hat{R}_{yu}^N(\tau) = \sum_{k=1}^N \hat{h}_k \hat{R}_u^N(k - \tau) \quad (5.13)$$

относительно \hat{h}_k .

Более детальное описание метода непараметрической идентификации на основе функций ковариации дано в [42].

Перейдем теперь к рассмотрению более общего случая. Пусть имеем систему с R входами и M выходами, которая в начальный момент времени подачи входного воздействия находилась в покое. Входной сигнал будем подавать отдельно на каждый вход, и на каждом отдельном выходе измеряется отклик на него. В результате проведения R экспериментов с разными входами получим $M \times R$ пар данных для непараметрической идентификации импульсных переходных функций, являющихся элементами импульсной матрицы отклика. Задача идентификации таким образом свелась к скалярному случаю для $M \times R$ отдельных независимых задач.

Когда погрешность выходного сигнала $v(t)$ в (5.6) интерпретируется как нестохастический процесс, т.е. принадлежит ограниченному множеству, это означает, что допустима любая реализация последовательности, в том числе неблагоприятная в статистическом смысле. Оценка функции ковариации погрешности $\hat{R}_v(\tau)$ также как и $R_{uv}(\tau)$ может быть произвольной функцией от τ , зависящей от конкретной реализации. В этом случае ковариационный анализ, основанный на свойствах статистик, может оказаться не эффективным.

Поэтому, когда $v(t) \in V$, где V – ограниченное множество, которому принадлежат значения v при разных t , можно использовать интервальный анализ. При этом получаем не одну импульсную переходную функцию, а целое их множество. Если не исключаются самые неблагоприятные реализации погрешностей, то такое множество будет гарантировано содержать импульсную переходную функцию с точными значениями h_k .

Для асимптотически устойчивых систем при ограниченной неопределенности значение импульсной переходной функции становится, начиная с некоторого момента времени t , сравнимым и даже существенно меньше погрешности. Отношение сигнал/шум

сначала станет меньше единицы, а затем будет продолжать уменьшаться, т.е. h_k в (5.6) при больших k очень мало и вклад информативных членов в этом уравнении в случае ограниченных входов становится меньше помехи $v(t)$.

Следовательно, найдется такое $t = N$, после которого данные не являются информативными и суммирование в уравнении (5.6) следует проводить до этого значения. При нулевых начальных значениях для вычисления h_k получаем простую рекурсивную процедуру, из которой хорошо просматривается данная проблема. Рекуррентная процедура представляется как

$$\begin{aligned} y_1 &= h_1 u_0 + v_1 \\ \vdots & \quad \quad \quad \vdots \\ y_s &= h_1 u_{s-1} + h_2 u_{s-2} + \dots + h_s u_0 + v_s \\ \vdots & \quad \quad \quad \vdots \quad \quad \quad \vdots \end{aligned} \quad (5.14)$$

Система (5.14) всегда разрешима при $u_0 \neq 0$ независимо от того какие u_s реализовывались в последующем. Решение (5.14) в рекуррентной форме записывается формулами

$$\begin{aligned} h_1 &= \frac{1}{u_0} y_1 - \frac{1}{u_0} v_1, \\ h_2 &= \frac{1}{u_0} (y_2 - \frac{u_1}{u_0} y_1) + \frac{1}{u_0} (\frac{u_1}{u_0} v_1 - v_2), \\ h_3 &= \frac{1}{u_0} (y_3 - \frac{u_1}{u_0} y_2 - \frac{u_2 u_0 - u_1^2}{u_0^2} y_1) + \frac{1}{u_0} (\frac{u_2 u_0 - u_1^2}{u_0^2} v_1 + \frac{u_1}{u_0} v_2 - v_3), \\ h_4 &= \frac{1}{u_0} (y_4 - \frac{u_1}{u_0} y_3 - \frac{u_2 u_0 - u_1^2}{u_0^2} y_2 - \frac{u_0^2 u_3 + u_1^3 - 2u_0 u_1 u_2}{u_0^3} y_1) + \\ & \quad + \frac{1}{u_0} (\frac{u_0^2 u_3 + u_1^3 - 2u_0 u_1 u_2}{u_0^3} v_1 + \frac{u_2 u_0 - u_1^2}{u_0^2} v_2 + \frac{u_1}{u_0} v_3 - v_4) \\ \vdots & \quad \quad \quad \vdots \end{aligned} \quad (5.15)$$

Выходной сигнал y_k в силу линейности пропорционален u_0 . Поэтому точные значения h_k инвариантны относительно u_0 . В том случае, когда погрешность v_k не зависит от входного воздействия $u(t)$, целесообразно u_k выбрать так, чтобы минимизировать гарантированный интервал принадлежности точного значения g_k . Принимая во внимание ограниченность реализаций v_k , т.е. интервальную оценку $|v_\varepsilon| \leq \varepsilon_k$, допустимую входную последовательность $\{u_k, k = 1, 2, 3, \dots\}$, обеспечивающую минимальный гарантированный интервал принадлежности точного значения h_k , можно найти задавая некоторые значения u_0 и u_1 , а все последующие однозначно определяются

соотношениями $u_2 = \frac{u_1^2}{u_0}$, $u_3 = \frac{u_1^3}{u_0^2}$ и т.д. Здесь предполагается, что $u_0 > 0$, $u_1 > 0$. В этом

случае гарантированный интервал принадлежности точного значения h_k запишется как

$$\begin{aligned} h_1 \in \left[\hat{h}_1 - \frac{\varepsilon_1}{u_0}, \hat{h}_1 + \frac{\varepsilon_1}{u_0} \right], h_2 \in \left[\hat{h}_2 - \frac{1}{u_0} \left(\frac{u_1}{u_0} \varepsilon_1 + \varepsilon_2 \right), \hat{h}_2 + \frac{1}{u_0} \left(\frac{u_1}{u_0} \varepsilon_1 + \varepsilon_2 \right) \right], \dots, \\ h_k \in \left[\hat{h}_k - \frac{1}{u_0} \left(\frac{u_1}{u_0} \varepsilon_{k-1} + \varepsilon_k \right), \hat{h}_k + \frac{1}{u_0} \left(\frac{u_1}{u_0} \varepsilon_{k-1} + \varepsilon_k \right) \right], \dots \end{aligned} \quad (5.16)$$

При одинаковом для всех k ограничении на погрешность ($\varepsilon = \varepsilon_1 = \varepsilon_2 = \dots$) имеем

$$h_1 \in \left[\hat{h}_1 - \frac{\varepsilon}{u_0}, \hat{h}_1 + \frac{\varepsilon}{u_0} \right], \dots, h_k \in \left[\hat{h}_k - \frac{\varepsilon}{u_0} \left(1 + \frac{u_1}{u_0} \right), \hat{h}_k + \frac{\varepsilon}{u_0} \left(1 + \frac{u_1}{u_0} \right) \right], k = 2, 3, \dots \quad (5.17)$$

Как видим, неопределенность в оценке импульсной переходной функции тем меньше, чем больше значение u_0 и меньше отношение $\frac{u_1}{u_0}$.

В реальной ситуации вряд ли стоит ожидать независимость погрешности v_k от u_k , учитывая особенности измерительных устройств. Тем не менее, когда имеются аддитивные неконтролируемые входные воздействия, увеличение u_0 может повысить отношение сигнал/помеха на выходе. Во всяком случае можно допускать, что существует некоторое u_0 , которое наиболее благоприятно с точки зрения отношения сигнал/помеха, учитывая в том числе ограничения на допустимые входные воздействия.

Приведенные выше последовательности входных воздействий фактически являются различными аппроксимациями функции Дирака, используемых в качестве входных воздействий в активных экспериментах для наиболее эффективного определения импульсной функции отклика системы. Предельной из них можно считать ступенчатую функцию $u_0 = u_1 = u_2 = \dots$, при которой погрешность ограничена величиной $\frac{2\varepsilon}{u_0}$.

5.3 Комбинаторный метод непараметрической идентификации

Не всегда в реальных экспериментах можно обеспечить входное воздействие в виде δ -функции или одной из ее аппроксимаций. Кроме того, предположение нулевых начальных условий в момент подачи импульса входного сигнала снижает общность рассмотрения. При произвольном входном воздействии и нулевых начальных условиях следует ожидать из (5.15), что гарантированный интервал принадлежности h_k будет расти с увеличением k , поскольку кроме v_k и v_{k-1} будут давать вклад в оценку \hat{h}_k

погрешности v_{k-2} , v_{k-3} и т.д. Поэтому необходим метод непараметрической идентификации, который бы не допускал существенного роста гарантированного интервала принадлежности с увеличением k . Для этого следует отказаться от получения гарантированных интервалов, ориентируясь на самые неблагоприятные реализации погрешностей, и в определенном смысле учитывать конкретно реализуемые погрешности в разных комбинациях. Именно такой подход к оценке гарантированных интервалов был разработан в [43] при решении переопределенной системы линейных алгебраических уравнений. Чтобы им воспользоваться, несколько видоизменим постановку задачи непараметрической идентификации. Пусть имеем асимптотически устойчивую дискретную систему, которая при относительно произвольном, но все же информативном входном воздействии на интервале $[1, 2, \dots, N+K]$ сгенерировала измерения $\{y_k\}$. Значение N таково, что в конце интервала $[1, 2, \dots, N]$ импульсная переходная функция практически близка к нулю, а точнее, начиная с $N+1$ и далее h_k по величине меньше погрешности измерения, т.е. меньше ε . При этом последовательность $\{u_k\}$ остается ограниченной. Тогда при условии, что $K > N$, имеем следующую переопределенную систему уравнений

$$\begin{aligned} u_{N-1}h_1 + \dots + u_0h_N &= y_N, \\ u_{N1}h_1 + \dots + u_1h_N &= y_{N+1}, \\ \vdots & \quad \quad \quad \vdots \\ u_{N+k-1}h_1 + \dots + u_kh_N &= y_{N+k}. \end{aligned} \tag{5.18}$$

В результате получили переопределенную систему линейных алгебраических уравнений с неточно заданной правой частью, для которой имеем интервальную оценку. В общем случае она не имеет классического решения. Поэтому находится обобщенное приближенное решение, которое обычно строится с помощью МНК или взвешенного МНК. Однако при интервальной погрешности задания правой части наиболее эффективным оказывается комбинаторный метод, описанный в [43]. Достоинством его является то, что он не ориентирован полностью на наиболее неблагоприятную реализацию погрешностей v_k , а учитываются конкретные ее реализации, которые могут оказаться даже благоприятными, поскольку вероятность самых неблагоприятных реализаций очень мала. Суть комбинаторного метода решения (5.18) сводится к следующему. Из уравнений, входящих в (5.18), формируется множество квадратных систем, выбирая для этого произвольные N уравнений. Таких комбинаций может быть не более C_K^N . На этом конечном множестве квадратных систем оставляем только те, которые невырождены и имеют приемлемую обусловленность (плохо обусловленные системы

исключаются). Их может быть меньше или равно C_K^N . Решение каждый из них дает оценку \hat{h}_k . Кроме этой оценки можно найти покомпонентно гарантированный интервал принадлежности точного значения h_k для всего множества квадратных систем. Для нахождения \hat{h}_k^i и значения δh_k^i , определяющего ширину гарантированного интервала для i -ой квадратной системы, представляемой как

$$y^i = U^i h^i, \quad i = 1, 2, \dots, S, \quad (5.19)$$

где $S \leq C_K^N$, можно использовать правило Крамера.

Согласно этому правилу можно записать

$$h^i = (U^i)^{-1} y^i = \frac{1}{\det U^i} \cdot \check{U}^i \cdot y^i, \quad (5.20)$$

где \check{U}^i – присоединенная матрица алгебраических дополнений. Оценка \hat{h}_k^i находится по формуле

$$\hat{h}_k^i = \frac{1}{\det U^i} \cdot \sum_{j=1}^N \tilde{y}_j^i \cdot \check{U}_{jk}^i, \quad (5.21)$$

где \check{U}_{jk}^i – элементы матрицы \check{U}^i , а значения δh_k^i вычисляются из соотношения

$$\delta h_k^i = \varepsilon \sum_{j=1}^N \left| \frac{\check{U}_{jk}^i}{\det U^i} \right|. \quad (5.22)$$

Значения \tilde{y}_j^i соответствуют реальным (с погрешностью) измерениям, которые оказались в i -ой квадратной системе. Для каждого k отдельно из разных i формируем множество интервалов

$$\begin{aligned} \delta_k^1 &\in [-\delta h_k^1 + \hat{h}_k^1, \hat{h}_k^1 + \delta h_k^1], \\ \delta_k^2 &\in [-\delta h_k^2 + \hat{h}_k^2, \hat{h}_k^2 + \delta h_k^2], \\ &\vdots \\ \delta_k^s &\in [-\delta h_k^s + \hat{h}_k^s, \hat{h}_k^s + \delta h_k^s], \\ &k = \overline{1, N}. \end{aligned} \quad (5.23)$$

Точное значение h_k будет принадлежать интервалу, который образуется пересечением интервалов (5.23), т.е. интервалу

$$\delta_k = [\max_i (-\delta h_k^i + \hat{h}_k^i), \min_i (\hat{h}_k^i + \delta h_k^i)]. \quad (5.24)$$

Середину этого интервала принимаем за окончательную оценку \hat{h}_k , а $\frac{1}{2} \delta_k$ определяет ширину гарантированного интервала, т.е. точное значение h_k в этом случае

отличается от оценки \hat{h}_k не более, чем на $\frac{1}{2}\delta_k$.

Решение этим способом можно находить итеративно, варьируя величину Δ , которая определяется равенством $\Delta N = T$, где T – время переходного процесса, соответствующего информативному интервалу функции $\{h_k\}$. Начинать можно с не очень малых Δ , при которых имеем приемлемое число обусловленности матриц U^i . Затем уменьшаем Δ так, чтобы N оставалось целым числом. Уменьшение Δ продолжаем до тех пор, пока (5.24) дает вполне удовлетворительный гарантированный интервал принадлежности, при котором сохраняется корректность решения, а именно, его практическая пригодность.

Таким образом, при данном способе препятствием для достижения малых значений Δ может оказаться плохая обусловленность матриц U^i при больших N . Продвинуться в область с малыми Δ можно, если использовать другой способ непараметрической идентификации, ориентированный на данные в активных экспериментах, в которых допустимые входные воздействия в виде последовательности прямоугольных импульсов длительностью T чередуются с интервалами релаксации той же длительности T , где вход равен нулю. В таком случае можно применить оригинальный алгоритм, в котором удачно сочетаются достоинства описанных выше алгоритмов. Амплитуду импульсов будем считать единичной, поскольку всегда можно отнормировать выход $y(t)$, поделив его на амплитуду импульсов. При этом может нормироваться и величина ε , задающая ограничение на погрешность. Как это уже говорилось, значение T выбирается таким, что на момент T переходный процесс практически затухает. Характер входного воздействия позволяет констатировать, что описание (5.6) является точным для произвольного шага квантования на интервалах, где $u(t)$ константа. Поэтому будем менять шаг дискретизации, задавая его с помощью соотношения $T = 2^{s+1}\Delta$, $s = 0, 1, 2, \dots$, начиная с $s = 0$. Переход от непрерывного описания, когда (5.6) имеет вид уравнения свертки, к его дискретному аналогу в этом случае соответствует тому, что h_k является интегралом непрерывной функции $h(t)$ на каждом таком шаге

Пусть K_j , $j = 1, 2, \dots, J$ определяет дискретные моменты времени начала прямоугольных импульсов, начиная со второго при $j = 1$ и кончая последним при $j = J$. Ненулевые начальные условия к моменту времени K_1 практически затухают.

Это позволяет для каждого j ($j = \overline{1, J}$) рассматривать отдельно базовый интервал наблюдения с одним прямоугольным импульсом, как это показано на рис. 5. 1.

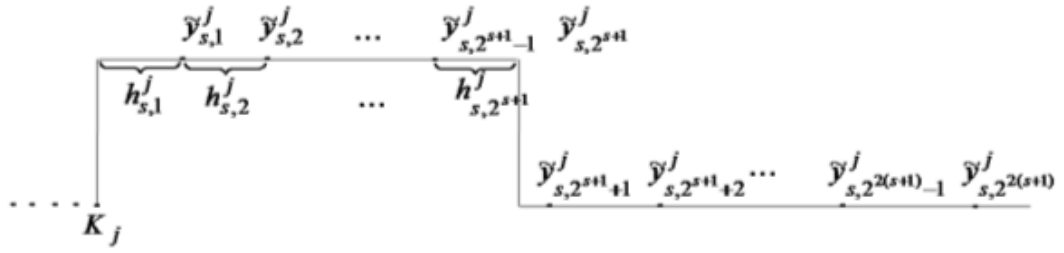


Рис. 5.1

Для данного входного воздействия, как уже говорилось, дискретное описание (5.6) является точным при любом s , если $h_{s,k}$ ($k = 1, 2, \dots$) есть интеграл непрерывной функции $h(t)$ на каждом шаге. Тогда при увеличении s на единицу будут выполняться следующие соотношения:

$$\begin{aligned} h_{s+1,1}^j + h_{s+1,2}^j &= h_{s,1}^j, \\ h_{s+1,3}^j + h_{s+1,4}^j &= h_{s,2}^j, \\ \vdots & \quad \quad \quad \vdots \\ h_{s+1,2^{2(s+1)}-1}^j + h_{s+1,2^{2(s+1)}}^j &= h_{s,2^{s+1}}^j, \end{aligned} \quad (5.25)$$

где $s+1$ соответствует новому шагу дискретизации на единицу больше предшествующего, которое соответствует s .

Особенность используемого входного воздействия, представленного в виде базового элемента K_j , дает возможность для каждого из них записать соотношения вход-выход в форме импульсного отклика как

$$\begin{aligned} \tilde{y}_{s,1}^j &= h_{s,1}^j, \\ \tilde{y}_{s,2}^j &= h_{s,2}^j + h_{s,1}^j, \\ \vdots & \quad \quad \quad \vdots \\ \tilde{y}_{s,2^{s+1}}^j &= h_{s,2^{s+1}}^j + \dots + h_{s,1}^j, \\ \tilde{y}_{s,2^{s+1}+1}^j &= h_{s,2^{s+1}+1}^j + \dots + h_{s,2}^j, \\ \vdots & \quad \quad \quad \vdots \\ \tilde{y}_{s,2^{2(s+1)}-2}^j &= h_{s,2^{2(s+1)}-2}^j + h_{s,2^{2(s+1)}-1}^j, \\ \tilde{y}_{s,2^{2(s+1)}-1}^j &= h_{s,2^{2(s+1)}-1}^j. \end{aligned} \quad (5.26)$$

Совместное использование (5.25) и (5.26) при каждом значении s можно трактовать как переопределенную систему уравнений для нахождения дискретной импульсной переходной функции. Из них несложно формировать множество квадратных систем в рекуррентной треугольной форме. Каждая такая система, входящая в это

множество, позволяет найти оценку $\hat{h}_{s,k}$ и гарантированный интервал принадлежности точного значения дискретной импульсной переходной функции.

Начинать целесообразно с $s=0$, при котором (5.25) не дает уравнений, а (5.26) принимает вид

$$\begin{aligned}\tilde{y}_{01}^j &= h_{01}^j, \\ \tilde{y}_{02}^j &= h_{02}^j + h_{01}^j, \\ \tilde{y}_{03}^j &= h_{02}^j.\end{aligned}\tag{5.27}$$

В этом случае множество состоит из двух рекуррентных систем, которые дают две пары оценок и гарантированные интервалы принадлежности

$$\begin{aligned}h_{01} &\in [\hat{h}_{01}^{1j} - \varepsilon, \hat{h}_{01}^{1j} + \varepsilon], & h_{02} &\in [\hat{h}_{02}^{1j} - \varepsilon, \hat{h}_{02}^{1j} + \varepsilon], \\ h_{01} &\in [\hat{h}_{01}^{2j} - 2\varepsilon, \hat{h}_{01}^{2j} + 2\varepsilon], & h_{02} &\in [\hat{h}_{02}^{2j} - 2\varepsilon, \hat{h}_{02}^{2j} + 2\varepsilon],\end{aligned}\tag{5.28}$$

где $\hat{h}_{01}^{1j} = \tilde{y}_{01}^j$, $\hat{h}_{01}^{2j} = \tilde{y}_{02}^j - \tilde{y}_{03}^j$, $\hat{h}_{02}^{1j} = \tilde{y}_{02}^j$, $\hat{h}_{02}^{2j} = \tilde{y}_{02}^j - \tilde{y}_{03}^j$.

После этого данные действия следует осуществить для всех базовых элементов (рис. 5.1), т.е. для $j = 2, 3, \dots, J$. В результате получим $2J$ интервалов принадлежности для h_{01} и h_{02} . Их пересечение позволит определить уточненные гарантированные множества как для h_{01} , так и h_{02} . Середины этих множеств следует принять за оценки \hat{h}_{01} и \hat{h}_{02} , а половина их ширины даст уточненную интервальную погрешность.

Переход к $s=1$ даст следующие уравнения:

$$\begin{aligned}h_{0,1}^j &= h_{1,1}^j + h_{1,2}^j, & \tilde{y}_{1,1}^j &= h_{1,1}^j, \\ h_{0,2}^j &= h_{1,3}^j + h_{1,4}^j, & \tilde{y}_{1,2}^j &= h_{1,2}^j + h_{1,1}^j, \\ & & \tilde{y}_{1,3}^j &= h_{1,3}^j + h_{1,2}^j + h_{1,1}^j, \\ & & \tilde{y}_{1,4}^j &= h_{1,4}^j + h_{1,3}^j + h_{1,2}^j + h_{1,1}^j, \\ & & \tilde{y}_{1,5}^j &= h_{1,4}^j + h_{1,3}^j + h_{1,2}^j, \\ & & \tilde{y}_{1,6}^j &= h_{1,4}^j + h_{1,3}^j, \\ & & \tilde{y}_{1,7}^j &= h_{1,4}^j.\end{aligned}\tag{5.29}$$

Из уравнений (5.29) можно сформировать четыре рекуррентные системы для определения $h_{1,1}^j$, $h_{1,2}^j$, $h_{1,3}^j$, $h_{1,4}^j$.

Если исходные данные в этих уравнениях также как и в уравнениях для $s=0$ были точные, то все системы давали бы одно и то же решение. При неточных данных в левых частях каждая из этих систем даст свою оценку $\hat{h}_{1,i}^j$, $\hat{h}_{1,2}^j$, $\hat{h}_{1,3}^j$, $\hat{h}_{1,4}^j$, где $i=1, 2, 3, 4$. Зная интервальные оценки для $\tilde{y}_{1,i}^j$ ($i=1, 2, 3, 4$) и $\hat{h}_{0,1}^j$, $\hat{h}_{0,2}^j$, можно построить по аналогии с (5.28) для каждого $h_{1,m}^j$ ($m=1, 2, 3, 4$) гарантированные интервалы принадлежности

точного значения. Далее, как и в случае с $s=0$, осуществляются те же самые действия для всех базовых элементов $j = \overline{2, J}$. В результате получим для каждого h_{1m} $4J$ гарантированных интервалов. Их пересечение дает возможность получить уточненные оценки как \hat{h}_{1m}^j , так и гарантированных интервалов принадлежности точных значений. Приведем еще алгоритм гарантированного оценивания для случая $s=2$, поскольку для него и последующих появляются новые особенности. В этом случае уравнения (5.25) примут вид

$$\begin{aligned} h_{1,1}^j &= h_{2,1}^j + h_{2,2}^j, \\ h_{1,2}^j &= h_{2,3}^j + h_{2,4}^j, \\ h_{1,3}^j &= h_{2,5}^j + h_{2,6}^j, \\ h_{1,4}^j &= h_{2,7}^j + h_{2,8}^j, \end{aligned} \tag{5.30}$$

а (5.26) как и во всех предшествующих случаях дает две рекуррентные системы. Уравнения (5.30) с учетом первой пары уравнений в (5.29) можно записать в эквивалентной форме

$$\begin{aligned} h_{1,1}^j &= h_{2,1}^j + h_{2,2}^j, \\ h_{0,1}^j &= h_{2,1}^j + h_{2,2}^j + h_{2,3}^j + h_{2,4}^j, \\ h_{0,2}^j &= h_{2,5}^j + h_{2,0}^j + h_{2,7}^j + h_{2,8}^j, \\ h_{1,4}^j &= h_{2,7}^j + h_{2,8}^j. \end{aligned} \tag{5.31}$$

В результате уравнения (5.26) при $s=2$ и (5.31) позволяют построить восемь рекуррентных независимых систем для определения $h_{2,m}^j$. Решение каждой из них комбинаторным методом дает оценку $\hat{h}_{2,m}^j$ и гарантированный интервал, которому принадлежит точное значение. Дальнейшее увеличение s будет приводить к расширению числа рекуррентных систем как 2^{s+1} , если при этом использовать (5.25) не только для данного s , но и для предшествующих ему. Для не очень малых s при достаточно больших T можно ограничиться только уравнениями (5.25) и (5.26) без привлечения уравнений (5.25) для меньших s . Тогда множество рекуррентных систем будет расти, но не так быстро.

5.4 Структурно-параметрическая идентификация

Непараметрическая идентификация импульсной переходной функции не решает полностью задачу идентификации, т.е. задачу нахождения модели системы адекватной исходным экспериментальным данным. Полученная в результате решения непараметрической идентификации функция $h(t)$ является, как и $y(t)$, неаналитической, представленной в дискретном случае в виде графической или табличной функции. Ее, как и $y(t)$, следует считать интегральной характеристикой системы, содержащей суммарный вклад всех элементарных подсистем или отдельных мод, из которых состоит всякая сложная система. При решении некоторых прикладных задач можно использовать непосредственно импульсную переходную функцию, заданную неаналитически. Однако, чтобы иметь более глубокое представление об изучаемом объекте, желательно восстановить по этой импульсной функции ее математическую модель, например, в пространстве состояний, включая модель измерительной системы. Другими словами, следует перейти от описания в форме уравнения конволюции к эквивалентному ему в виде системы разностных уравнений и уравнения наблюдения. Такое описание хорошо известно и имеет вид

$$x_{k+1} = Ax_k + bu_k, y_k = c^T x_k. \quad (5.32)$$

Переход от описания (5.32) к импульсной переходной функции определяется соотношением (см. 3.13)

$$h_k = c^T A^{k-1} b, k = 1, 2, \dots \quad (5.33)$$

При этом предполагается, что в начальный момент времени $k=0$ система находилась в покое и $u_k=0$ для любого отрицательного k , т.е. для $k < 0$. Модель в форме (5.32) фактически дает аналитическое описание дискретной импульсной переходной функции в виде разложения

$$h_k = \sum_{p=1}^P \rho_p^k [f_p^c \cos \omega_p(k-1) + f_p^s \sin \omega_p(k-1)], \quad (5.34)$$

где ρ_p , ω_p , f_p^c , f_p^s выражаются через компоненты векторов b , c и собственные значения матрицы A системы (5.32), которые приведены в разделе 3.1 главы 3 (формулы (3.23) и (3.24)). Таким образом, построить модель (5.32) – все равно, что аппроксимировать дискретную функцию h_k в виде разложения (5.34). Особенность такой аппроксимации в том, что параметры базисных функций и их количество также должны определяться при решении соответствующей задачи.

Если в результате решения задачи непараметрической идентификации была

получена оценка \hat{h}_k с гарантированными интервалами принадлежности его точных значений, то модель (5.32) можно построить через решение указанной задачи аналитической аппроксимации в виде разложения (5.34) дискретно заданной функции h_k , полученной при непараметрической идентификации. В этом случае ее решение можно найти методами, которые были описаны при рассмотрении задач редукции порядка модели сложной системы с использованием процедур минимизации норм, характеризующих отклонение аналитического выражения h_1 в форме (5.34) от полученной оценки \hat{h}_k , т.е. из задач оптимизации функций

$$\|\hat{h}_k - h_k\| \rightarrow \min \quad (5.35)$$

по всем неопределенным параметрам, входящих в (5.34). Решение ее целесообразно искать итерационно, задаваясь различными значениями числа P членов разложения в (5.34). Начинать следует с $P=1$ и последующим увеличением на единицу этой размерности. Итерации заканчиваются при том значении P^* , при котором оптимальная аппроксимирующая кривая (5.32) будет целиком принадлежать гарантированному интервалу принадлежности точного значения h_k . Это и будет согласованное по точности регуляризированное решение задачи структурно-параметрической идентификации. После этого можно найти модель в форме (5.32) для одной из канонических жордановых реализаций, управляемой или наблюдаемой. Процедуры их нахождения описаны в разделе 3.1 главы 3. От жордановой реализации с помощью неособого преобразования легко перейти к любой другой эквивалентной реализации, например, управляемой канонической реализации [] широко используемой при синтезе обратной связи.

Неединственность модельного описания (5.32) одной и той же системы позволяет применить для решения задачи структурно-параметрической идентификации другой более оригинальный по реализации способ. По сути он практически содержит те же действия с использованием матриц Ганкеля и SVD-преобразования, которые описаны в разделе 4.5 при решении задачи редукции порядка модели. Поскольку при его реализации потребуется учесть некоторые особенности рассматриваемой задачи, дадим его подробное описание. Из последовательности оценок $\{\hat{h}_k\}$, $k = \overline{1, N}$, полученных одним из непараметрических методов, строим матрицу Ганкеля

$$H_\Gamma = \begin{bmatrix} h_1 & h_2 & \dots & h_K \\ h_2 & h_3 & \dots & h_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ h_K & h_{K+1} & \dots & h_{2K} \end{bmatrix} = \Gamma_K \cdot \Omega_K, \quad (5.36)$$

где $\Gamma_K = \begin{bmatrix} c^T \\ c^T A \\ \vdots \\ c^T A^{K-1} \end{bmatrix}$ и $\Omega_K = [b \ Ab \ \dots \ A^{K-1}b]$ – матрицы наблюдаемости и управляемости, а

$$2K \leq N.$$

Осуществим SVD-преобразование матрицы H_Γ , которое запишем как

$$H_\Gamma = Q \Sigma V^T, \quad (5.37)$$

где Q , V – ортогональные матрицы, а Σ – диагональная матрица сингулярных чисел.

Разобьем матрицы Q , V , Σ на блоки

$$Q = [Q_p \ Q_e], \ V = [V_p \ V_e], \ \Sigma = \begin{bmatrix} \Sigma_p & 0 \\ 0 & \Sigma_e \end{bmatrix}. \quad (5.38)$$

В (5.38) блоки Q_p , V_p , Σ_p соответствуют модели, аппроксимирующей импульсную переходную функцию, а первое сингулярное число Σ_e характеризует уклонение усеченной матрицы $H_{\Gamma P} = Q_p \Sigma_p V_p^T = \Gamma_p \cdot \Omega_p$ от полноранговой матрицы H_Γ по норме $\|\cdot\|_2$. Свойство неединственности описания (5.32) в соответствии с теорией реализации позволяет утверждать, что на множестве всех возможных реализаций, связанных между собой неособым преобразованием, найдется такая, для которой будем иметь

$$\Gamma_p = Q_p \cdot \Sigma_p^{1/2}, \ \Omega_p = \Sigma_p^{1/2} \cdot V_p^T. \quad (5.39)$$

Фактически (5.39) означает представление матрицы $H_{\Gamma P}$ в виде произведения двух полноранговых матриц, одна из которых для некоторой реализации является матрицей наблюдаемости, а другая – матрицей управляемости.

Для выбранной размерности P первая строка матрицы Γ_p равна вектору c^T , а первый столбец матрицы Ω_p равен вектору b модели редуцированной системы (5.32). Остается определить для этой реализации матрицу A . Ее можно найти из решения переопределенной системы уравнений

$$\begin{aligned} \Gamma_{2:P} &\approx \Gamma_{1:P-1} \cdot A, \\ \Omega_{2:P} &\approx A \cdot \Omega_{1:P-1}, \end{aligned} \quad (5.40)$$

где $\Gamma_{1:P-1}$ – матрица наблюдаемости, т.е. матрица $Q_p \cdot \Sigma_p^{1/2}$, у которой вычеркнута последняя строка, а $\Gamma_{2:P}$ есть та же матрица $Q_p \cdot \Sigma_p^{1/2}$, но с вычеркнутой первой строкой. Аналогично строятся матрицы $\Omega_{2:P}$ и $\Omega_{1:P-1}$, у которых вычеркиваются первый и

последний столбцы, соответственно. Решить систему (5.40) можно разными способами. В силу того, что проведенные выше операции приводят к реализации, которая будет балансной в том смысле, что

$$\Gamma_{1:P-1} \cdot \Gamma_{1:P-1}^T = \Omega_{1:P-1} \cdot \Omega_{1:P-1}^T,$$

то для получения расширенной системы идентичных матричных уравнений можно приводить как систему один к виду два в (5.40), так и наоборот, приводить систему два к виду один. Например, система два примет вид аналогичный первой системе после проведения ряда стандартных действий с матрицей управляемости, которые в конечном итоге приведут к следующему результату:

$$\Omega_{2:P}^T \cdot \Omega_{2:P} \Omega_{1:P-1}^T (\Omega_{1:P-1} \cdot \Omega_{1:P-1}^T)^{-1} = \Omega_{2:P}^T A. \quad (5.41)$$

Так как идентифицируется минимальная реализация, которая управляемая и наблюдаемая, то матрица, которую приходится обращать в (5.41), будет невырожденной. Решение переопределенной системы, состоящей из первой системы в (5.40) и (5.41) может быть найдено классическим МНК или обобщенным [23].

Матрицу A можно найти также другим способом, который используется в алгоритме Eigensystem Realization Approach (ERA), описанном в [44]. Согласно этому алгоритму рассматриваются две последовательности $\{h_k\}$. Первая из них состоит из элементов h_k , начиная с $k=1$ и кончая $k=K-1$, а вторая – начинается с $k=2$ и заканчивается $k=K$. Из этих последовательностей формируются две матрицы Ганкеля, H_{Γ_1} и H_{Γ_2} , где H_{Γ_1} соответствует первой последовательности, а H_{Γ_2} – второй. Тогда, исходя из сдвиговой инвариантности, справедливо следующее соотношение

$$H_{\Gamma_2} = \Gamma_{P_1} \cdot A \cdot \Omega_{P_1}, \quad (5.42)$$

где $\Gamma_{P_1} \cdot \Omega_{P_1} = H_{\Gamma_1}$, т.е. Γ_{P_1} – матрица наблюдаемости, а Ω_{P_1} – матрица управляемости усеченной матрицы H_{Γ_1} , образованной из блоков Q_{P_1} , V_{P_1} и Σ_{P_1} , получаемых при разложении SVD-преобразования матрицы H_{Γ_1} на блоки согласно (5.38). Так как $\Gamma_{P_1} = Q_{P_1} \cdot \Sigma_{P_1}^{1/2}$ и $\Omega_{P_1} = \Sigma_{P_1}^{1/2} \cdot V_{P_1}^T$, то

$$H_{\Gamma_2} = Q_{P_1} \cdot \Sigma_{P_1}^{1/2} \cdot A \cdot \Sigma_{P_1}^{1/2} \cdot V_{P_1}^T. \quad (5.43)$$

Введем обозначения

$$H_{\Gamma_2}^+ = \left(Q_{P_1} \cdot \Sigma_{P_1}^{1/2} \right)^T \cdot H_{\Gamma_2} \cdot \left(\Sigma_{P_1}^{1/2} \cdot V_{P_1}^T \right)^T,$$

$$\Gamma_{P_1}^+ = \left(Q_{P_1} \cdot \Sigma_{P_1}^{1/2} \right)^T \cdot Q_{P_1} \cdot \Sigma_{P_1}^{1/2},$$

$$\Omega_{P_1}^+ = \Sigma_{P_1}^{1/2} \cdot V_{P_1}^T \cdot \left(\Sigma_{P_1}^{1/2} \cdot V_{P_1}^T \right)^T.$$

Уравнение (5.43) с использованием введенных обозначений можно привести к виду

$$H_{\Gamma_2}^+ = \Gamma_{P_1}^+ \cdot A \cdot \Omega_{P_1}^+, \quad (5.44)$$

в котором квадратные матрицы $\Gamma_{P_1}^+$ и $\Omega_{P_1}^+$ являются невырожденными, поскольку идентифицируется полностью наблюдаемая и управляемая система.

В результате решение (5.44) запишется как

$$A = \left(\Gamma_{P_1}^+ \right)^{-1} \cdot H_{\Gamma_2}^+ \cdot \left(\Omega_{P_1}^+ \right)^{-1}. \quad (5.45)$$

Целесообразно вместо матрицы H_{Γ_2} в уравнении (5.42) взять матрицу $H_{\Gamma_{P_2}}$, которая как и матрица $H_{\Gamma_{P_1}}$ образуется из блоков Q_{P_2} , V_{P_2} и Σ_{P_2} при разложении SVD-преобразования матрицы H_{Γ_2} на блоки согласно (5.38). Соответственно, в уравнениях (5.43), (5.44), (5.45) матрица H_{Γ_2} замещается матрицей $H_{\Gamma_{P_2}}$.

Одним из ключевых моментов метода – выбор размерности P , т.е. как следует производить разбиение (5.32) на блоки (5.30). Делать это следует итеративно и так, чтобы получить в результате такое P^* , чтобы оценка дискретной функции \hat{h}_k , определяемая соотношением (5.33), целиком принадлежала гарантированным интервалам принадлежности точного h_k , найденного из решения задачи непараметрической идентификации. При этом для $P^* - 1$ функция \hat{h}_k выходит за пределы гарантированного интервала. Такую итеративную схему нахождения P^* можно реализовать следующим образом. По значениям сингулярных чисел матрицы Σ SVD-разложения выбирается P_0 такое, чтобы отношение σ_1 к σ_{P_0} было более одного порядка, а $\sigma_1 / \sigma_{P_0+1}$ было более двух порядков, т.е. между σ_{P_0} и σ_{P_0+1} был существенный разрыв. Для выбранного P_0 производится разбиение на блоки (5.38) и решается задача параметрической идентификации вышеописанным методом. Строится оценка \hat{h}_k , которая затем проверяется на принадлежность гарантированному интервалу. Если она не принадлежит ему, то увеличиваем на единицу размерность P_0 блоков в разложении (5.33), а если принадлежит, то увеличиваем размерность P_0 на единицу. При новом P_1 большем или меньшем P_0 проделываем все то же, что делалось для P_0 вплоть до оценки

принадлежности гарантированному интервалу. В зависимости от результата итерации прекращаются или продолжаются. Заканчиваются они тогда, когда получим значение P^* , обладающее вышеуказанным свойством, а именно, дальнейшее уменьшение или увеличение на единицу P^* в зависимости от того, в каком направлении проводятся итерации, должно приводить либо к попаданию в интервал, либо выходу из него.

Структурно-параметрическая идентификация позволяет найти модель в виде (5.32) или (5.34) согласованную по точности с погрешностью в исходных данных. При желании можно проверить ее на принадлежность гарантированным интервалам для y_k при том входном воздействии, которое использовалось для непараметрической идентификации.