

ПРОБЛЕМА РЕДУКЦИИ ПОРЯДКА МОДЕЛИ ЛИНЕЙНОЙ СТАЦИОНАРНОЙ СИСТЕМЫ БОЛЬШОЙ РАЗМЕРНОСТИ

В.Ф. Губарев

Институт космических исследований НАН Украины и ГКА Украины (Киев)

Рассмотрена и описана очень важная для приложений проблема редукции модели линейной стационарной системы большой размерности к более простой невысокого порядка. Предложен аппроксимационный подход к решению этой проблемы, который позволил свести проблему редукции к решению оптимизационной задачи. Разработаны и обоснованы несколько оригинальных способов ее решения.

Ключевые слова: редукция моделей, аппроксимация, оптимизация, соответствие моделей, модель в пространстве состояний.

Розглянуто та описано дуже важливу для застосувань проблему редукції моделі лінійної стаціонарної системи великої розмірності до більш простої невисокої розмірності. Запропоновано аппроксимативний підхід до вирішення цієї проблеми, який дозволив звести проблему редукції до розв'язання оптимізаційної задачі. Розроблено та обґрунтовано декілька оригінальних способів її вирішення.

Ключові слова: редукція моделей, аппроксимация, оптимізація, відповідність моделей, модель в просторі станів.

ВВЕДЕНИЕ

В последнее время большой интерес специалистов по моделированию и управлению проявляется к возможности аппроксимации линейных стационарных систем 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. Ее решение позволяет на основе более простых уравнений эффективно решать задачи управления и прогнозирования по наблюдаемым выходным переменным. 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 [1]. 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. собственными значениями и некоторыми другими характеристиками. Из общих соображений можно предположить, что им соответствуют плохо

наблюдаемые и управляемые моды или группа мод с близкими собственными значениями.

ЦЕЛЬ РАБОТЫ

Основная цель, которая ставилась при выполнении данной работы, состояла в том, чтобы предложить и развить подход альтернативный существующим, позволяющий создавать эффективные методы и конструктивные алгоритмы построения моделей не очень высокого порядка, которые дают отклик на выходе близкий к тому, что получается на том же выходе реальной системы очень большой размерности при произвольном допустимом входном воздействии. Практическая важность постановки такой проблемы очевидна, поскольку упрощенное описание сложной системы позволяет синтезировать более простые, но достаточно эффективные законы управления. В основу предлагаемого подхода положен вариационный принцип построения усеченных моделей, который широко используется при идентификации систем.

ОПИСАНИЕ ПРОБЛЕМЫ

Исходной является математическая модель с описанием линейной стационарной системы в пространстве состояний в виде

$$\frac{dx}{dt} = Ax + Bu, \quad y = Cx, \quad (1)$$

где x — вектор внутреннего состояния системы, u — вектор входного воздействия, а y — вектор наблюдаемых выходных переменных. Размерность недоступного для наблюдения вектора x , интерпретируемого как внутреннее состояние системы, достаточно большая также как и соответствующие размерности матриц A , B , C . При этом вход u и выход y имеют существенно меньшую размерность.

Ставится задача: в пространстве состояний существенно меньшей размерности построить математическую модель:

$$\frac{d\hat{x}}{dt} = \hat{A}\hat{x} + \hat{B}u, \quad \hat{y} = \hat{C}\hat{x}, \quad (2)$$

порядок которой гораздо меньше исходной (1), обладающую следующим свойством. Отклик системы (2) на произвольное допустимое входное воздействие мало отличался бы от выхода y системы (1). Обычно малость отклонения $y - \hat{y}$ оценивается по некоторой норме, т.е. в виде неравенства:

$$\|y - \hat{y}\| \leq \varepsilon, \quad (3)$$

где ε — достаточно малая задаваемая величина.

Систему (2) называют редуцированной моделью системы большой размерности, т.е. \hat{x} , \hat{A} , \hat{B} , \hat{C} имеют меньшую размерность по отношению к

исходному математическому описанию (1). Норма, входящая в (3), задается при постановке проблемы редукции.

КРАТКИЙ ОБЗОР СУЩЕСТВУЮЩИХ МЕТОДОВ

Редукция порядка модели сложной системы при известном ее описании в виде системы уравнений (1) большой размерности может быть осуществлена различными методами, обзор которых можно найти, например, в [2, 3].

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 (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 находят подпространства Крылова. Подробности описаны в [2, 3]. 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.

ВАРИАЦИОННЫЙ СПОСОБ РЕШЕНИЯ ЗАДАЧИ РЕДУКЦИИ

Задачу редукции сложной модели к более простой можно рассматривать как аппроксимационную, если сравнивать согласно (3) их поведение по выходным переменным. Сами модели при этом существенно отличаются друг от друга, а их выходы могут быть очень близкими, что и является основным аргументом, позволяющим не считать задачу редукции бессмысленной. Хотя с чисто математической точки зрения ей трудно придать необходимую для этого строгость. Основным мотивом является практическая потребность такой постановки проблемы в тех случаях, когда приходится иметь дело со сложными системами. Исходя из такой

интерпретации задачи редукции, можно сформулировать ее в вариационной постановке, т.е. свести к оптимизационной. Для этого необходимо от представления моделей в виде (1), (2) перейти к эквивалентному им описанию в форме вход-выходных соотношений. Такие соотношения были получены в работах [4, 5]. В [4] они записаны для более общего случая с кратными корнями, а в [5] представлены результаты, когда кратные корни отсутствуют.

Не ограничивая общности и сути предлагаемого метода, ограничимся для простоты случаем без кратных корней. Тогда эквивалентная (1), (2) модель в виде соотношения вход-выход при нулевых начальных условиях запишется как:

$$y_m(t) = \sum_{r=1}^R \sum_{p=1}^P \int_0^t [f_{mp}^{rc} \cdot \cos \beta_p(t-\theta) + f_{mp}^{rs} \cdot \sin \beta_p(t-\theta)] e^{-\alpha_p t} u_r(\theta) d\theta, \quad (4)$$

где

$$f_{mp}^{rc} = c_{mp}^c b_{rp}^c + c_{mp}^s b_{rp}^s, \quad f_{mp}^{rs} = c_{mp}^s b_{rp}^c - c_{mp}^c b_{rp}^s, \quad m = \overline{1, M} \quad (\alpha_p > 0, \beta_p > 0). \quad (5)$$

Здесь за начало процесса принято $t=0$, а переходу к (4), (5) предшествовало преобразование (1), (2) к жордановой реализации, при которой матрица A имеет блочно-диагональную форму; R — число входов, а M — выходов системы.

$$A = \text{diag}(J_1, J_2, \dots, J_P), \quad (6)$$

где $J_p = \begin{pmatrix} -\alpha_p & \beta_p \\ -\beta_p & -\alpha_p \end{pmatrix}$, а $c_{mp}^c, c_{mp}^s, b_{rp}^c, b_{rp}^s$ — соответствующие этим

блокам элементы матриц C и B . Кроме того, в (4), (5) принята унифицированная форма записи, объединяющая действительные и комплексно-сопряженные собственные значения. Действительным соответствуют параметры $\beta_p = 0, c_{mp}^s = b_{rp}^s = 0$ с вытекающим из этого удалением некоторых столбцов и строк у матриц A, B, C для жордановой реализации. Собственные значения и параметры f_{mp}^{rc}, f_{mp}^{rs} являются инвариантами управляемой и наблюдаемой системы, т.е. не зависящие от того какая реализация была у исходных систем (1) или (2).

Пусть размерность исходной системы определяется достаточно большим значением P в (4). 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. $u(\theta) = \delta(\theta)$. 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 (4).

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

$$y(t) = \sum_{p=1}^P (f_p^c \cdot \cos \beta_p t + f_p^s \cdot \sin \beta_p t) \cdot e^{-\alpha_p t}, \quad (7)$$

в которой $y(t)$ является ни чем иным как импульсной переходной функцией.

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

$$y_k = y(\Delta k) = \sum_{p=1}^P (f_p^c \cdot \cos k\Delta\beta_p + f_p^s \cdot \sin k\Delta\beta_p) \cdot e^{-k\Delta\alpha_p}. \quad (8)$$

При выбранном Δ из (8) 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. В качестве такового возьмем также случай, в котором отсутствуют кратные корни. Редуцированная модель в указанном классе будет давать отклик в тех же точках и при том же воздействии в виде δ -функции согласно формуле:

$$\hat{y}_k = \sum_{q=1}^Q (f_q^c \cos k\Delta\beta_q + f_q^s \sin k\Delta\beta_q) e^{-k\alpha_q \Delta}, \quad Q \ll \sum_{p=1}^P n_p. \quad (9)$$

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

$$\frac{1}{N} \left\| \sum_{k=0}^N (y_k - \hat{y}_k) \right\| \rightarrow \min. \quad (10)$$

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:

$$\frac{1}{N} \left\| \sum_{k=0}^N (y_k - \hat{y}_k) \right\| \leq \varepsilon, \quad (11)$$

где ε — достаточно малая величина, а $\|\cdot\|$ — заданная норма.

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 (10), 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 (10) фактически есть нелинейный метод наименьших квадратов (МНК). 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 (11). При квадратичной норме в (10) условие (11) 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 (11) 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 (11) 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 (11), а соответствующая этой размерности полученная модель и будет решением задачи редукиции. Когда для каждого Q решение находится нелинейным МНК, 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 [6]. 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.

МЕТОД ВЫДЕЛЯЕМОГО ПОДПРОСТРАНСТВА

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

$$Y = \begin{bmatrix} y_0 & y_1 & \dots & y_K \\ y_1 & y_2 & \dots & y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_K & y_{K+1} & \dots & y_{2K} \end{bmatrix}, \quad (12)$$

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 , как уже говорилось, 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 (8), но и системе уравнений:

$$\begin{aligned} x_{k+1} &= Ax_k, \\ y_k &= c^T x_k, \quad x(0) = x_0, \end{aligned} \quad (13)$$

где вектор-строка c^T и матрица A задают некоторую эквивалентную (8) реализацию исходной модели в пространстве состояний соответствующей размерности. Начальное состояние x_0 выражается с помощью неособого

преобразования через вектор, компонентами которого являются инварианты f_p^c, f_p^s . На основе уравнений (13) можно получить следующее матричное уравнение:

$$Y = \Gamma \cdot X, \quad (14)$$

где Y — матрица (12), а матрица наблюдаемости Γ и матрица X имеют вид:

$$\Gamma = \begin{bmatrix} c^T \\ c^T A \\ \vdots \\ c^T A^{K-1} \end{bmatrix}, \quad X = [x_0 \ x_1 \ \dots \ x_K]. \quad (15)$$

Далее производим SVD-разложение матрицы Y , которое записывается как:

$$Y = U \Sigma V^T, \quad (16)$$

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

$$U = [U_r \ U_u], \quad \Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & \Sigma_u \end{bmatrix}, \quad V = [V_r \ V_u], \quad (17)$$

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

$$Y_r = U_r \Sigma_r V_r^T. \quad (18)$$

Теперь представим матрицу Y_r в виде произведения двух полноранговых матриц:

$$Y_r = Y_{r1} \cdot Y_{r2}, \quad (19)$$

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

Матричное уравнение (14) для редуцированной модели примет вид:

$$Y_r = \Gamma_r \cdot X_r. \quad (20)$$

В соответствии с теорией реализаций [6] можно положить

$$\Gamma_r = Y_{r1} = U_r, \quad X_r = Y_{r2} = \Sigma_r V_r^T, \quad (21)$$

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 (15), она обладает свойством сдвиговой инвариантности, из которого непосредственно следует следующее матричное приближенное уравнение:

$$Y_{r12} \approx Y_{r11} \cdot A, \quad (1)$$

в котором матрица Y_{r12} получается из матрицы Y_{r11} 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 [7].

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

$$J_q = \begin{bmatrix} -\alpha_q & \beta_q \\ -\beta_q & -\alpha_q \end{bmatrix} \quad (q = \overline{1, Q_2})$$

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 (13) заданы матрица A' и вектор $(c')^T$, а начальное состояние x'_0 , согласно (9), связано с вектором b' соотношениями:

$$\begin{aligned} x'_{01} &= f_1 = b'_1, \quad x'_{02} = f_2 = b'_2, \quad \dots, \quad x'_{0Q_1} = f_{Q_1} = b'_{Q_1}, \\ x'^c_{0Q_1+1} &= f_1^c = b'^c_{Q_1+1}, \quad x'^s_{0Q_1+1} = f_1^s = -b'^s_{Q_1+1}, \quad \dots, \\ x'^c_{0Q_1+Q_2} &= f_{Q_2}^c = b'^c_{Q_1+Q_2}, \quad x'^s_{0Q_1+Q_2} = f_{Q_2}^s = -b'^s_{Q_1+Q_2}. \end{aligned} \quad (23)$$

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 (21) к канонической жордановой реализации, которая определяется формулой:

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

где Γ' выражается через матрицу A' и вектор c' . Первый столбец матрицы $Y_{r,2}$, согласно (15), равен вектору x_0 соответствующей реализации, в которой была получена матрица A из (22). Вектор x'_0 , через который выражаются компоненты вектора b' , определяется равенством

$$x'_0 = Tx_0. \quad (24)$$

После нахождения из (23) и (24) вектора b' можно перейти от канонической жордановой реализации к той, которая необходима для дальнейшего применения.

Знание собственных значений редуцированной модели позволяет применить иную процедуру определения инвариантов, в основе которой 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:

$$R_N f = y, \quad (25)$$

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 большей размерности. Решение (25) будет приближенным, которое может быть найдено с помощью МНК, минимизирующего невязку в евклидовой норме. Приближенное решение (25), получаемое МНК, записывается как:

$$f = (R_N^T \cdot R_N)^{-1} \cdot R_N^T y. \quad (26)$$

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 (25) 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:

$$\min_F v^T v, \quad (27)$$

где $F = \{x, v : y = R_N f + Bv\}$, а B — заданная матрица. В данном случае Bv — есть обобщенная невязка. Решается (27) 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 малозначительными. 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 (26) вектора f при известных собственных значениях формируем импульсную функцию редуцированной модели и последовательность $\{\hat{y}_k\}$. Проверяем для заданной нормы выполнение условия (11). 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 (11). В зависимости от полученного результата итерации прекращаются или продолжаются. Они прекращаются в тех случаях, когда получилось равенство или произошла смена знака неравенства (11) 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 удовлетворительная 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. Если в описанном подходе оценка (11) 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 (26), 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.

РЕДУКЦИЯ МНОГОСВЯЗНОЙ СИСТЕМЫ

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 (11). 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 статье [8]. 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 инвариантов f_q^c и f_q^s для

соответствующих передаточных функций, для которых были скорректированы собственные значения. Проверка выполнимости (11) позволяет оценить, насколько удачно произведена коррекция и в случае необходимости провести повторную коррекцию.

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Продemonстрируем некоторые особенности изложенного подхода. Как соотносятся между собой собственные значения исходной и редуцированной модели показано на рис. 1 и рис. 2.

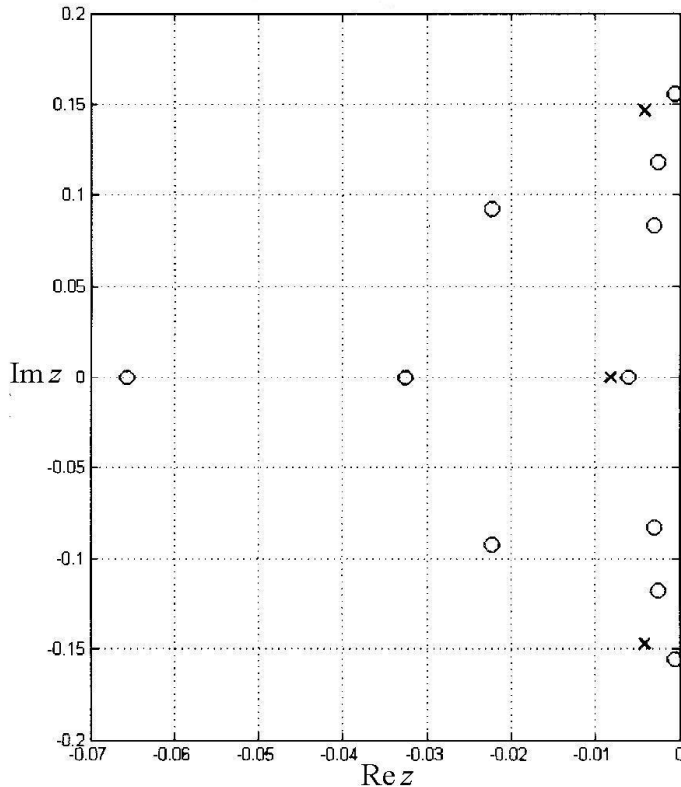


Рис. 1. Расположение на комплексной плоскости собственных значений исходной системы — « \circ » и редуцированной модели — « \times » для $Q = 3$

Собственные числа исходной системы представлены в устойчивой области комплексной плоскости в виде кружочков, а редуцированной — в виде крестиков. Результаты, показанные на рис. 1, получены для исходной модели 11-го порядка при выбранной размерности редуцированной модели

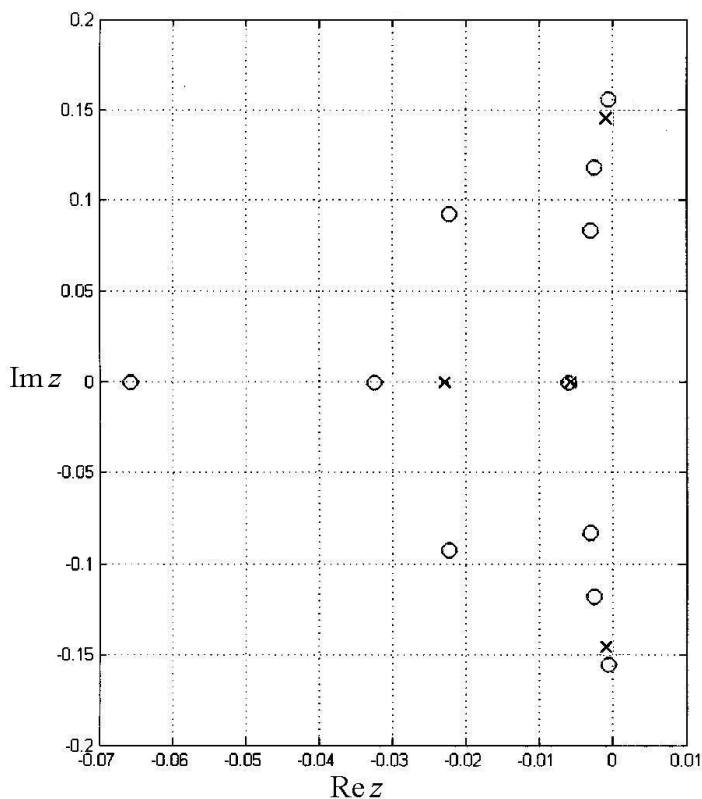


Рис. 2. Расположение на комплексной плоскости собственных значений исходной системы — «○» и редуцированной модели — «х» для $Q = 4$

равной трем. А на рис. 2 для той же исходной модели показано как располагаются собственные значения усеченной модели четвертого порядка по отношению к реальной.

Из рисунков видно, что каждое собственное значение модели уменьшенного порядка аппроксимирует некоторый кластер корней исходной системы. Другими словами, при редукации порядка модели получаем некоторые усредненные оценки параметров исходной системы, и ни о каких несмещенных оценках при этом не может идти речь. Особенно это проявляется при больших размерностях исходной модели.

1. Мищенко Е.Ф., Розов Н.Х. Дифференциальные уравнения с малым параметром и релаксационные колебания. М.: Наука. 1975. 248 с.
2. Antoulas A.C., Sorenson D.C., Gugercin S.A. A survey of model reduction methods for large-scale systems. *Contemporary Mathematics*. 2001. Vol. 280. pp. 193–219.
3. Reis T., Stykel T. A survey on model reduction of coupled systems. *Model order Reduction. Theory, Research Aspects and Applications*. 2008.
4. Губарев В.Ф. Метод итеративной идентификации многомерных систем по неточным данным. Часть 1. теоретические основы. *Проблемы управления и информатики*. 2006. № 5. С. 16–31.

5. Губарев В.Ф., Мельничук С.В. Идентификация многомерных систем по параметрам установившегося режима. *Проблемы управления и информатики*. 2012. № 5. С. 26–42.
6. Rainer S., Price K. Differential evolution — a simple and efficient adaptive scheme for global optimization over continuous spaces. *Journal of Global Optimization*. 1997. N 11. pp. 341–359.
7. Голуб Дж., Ван Лоун Ч. Матричные вычисления. М.: Мир. 1999. 548 с.
8. Мельничук С.В. Метод структурно-параметрической идентификации многосвязных систем по частотным характеристикам. *КБТ*. 2015. № 181. С. 66–79.

UDC 519.71

PROBLEM OF MODEL ORDER REDUCTION FOR LINEAR LARGE-SCALE TIME-INVARIANT SYSTEM

Gubarev V.F.

Space Research Institute NAS Ukraine and SSA Ukraine (Kyiv)

Introduction. Very significant for application model reduction problem of large-scale time-invariant system to more simple small order is considered and developed in the paper. Real and approximate models fitting is determined by norms which establish the difference between impulse response of these two models.

The purpose of the article is to propose a new approach of setting the model reduction problem and to develop methods based on variational principle of its solving.

Methods. It is proposed to set model reduction problem as optimization. For this initial state space model was transformed to equivalent description in form of input-output relation using analytical expression for impulse response. Such form allows to apply conception of fit between real system and its low-order approximation widely used in identification. Parameters of approximate model and its dimension are determined from optimization problem with different measure of fit writing as norm. Algorithms of numerical solving the optimization problems and needed for this data are considered in the paper. Besides the modified subspace method that permits to construct the observability matrix directly from output data using SVD factorization is proposed and described. Singular values of SVD-decomposition indicate as the best way to truncate full model.

Results. Some results dealing with mutual disposition of eigenvalues of real model and reduced one are demonstrated.

Conclusion. Developed methods may be used both for systems with scalar input and output and for multi-input and multi-output system as well. Results obtained by modelling show efficiency of all worked out methods.

Keywords: model reduction, approximation, optimization, model fit, state-space model.

1. Mischenko E.F. and Rozov N.H. Differential equations with small parameter and relaxation oscillations. M.: Nauka. 1975. 248 p. (in Russian).
2. Antoulas A.C., Sorenson D.C., Gugercin S.A. A survey of model reduction methods for large-scale systems. *Contemporary Mathematics*. 2001. Vol. 280. pp. 193–219.
3. Reis T., Stykel T. A survey on model reduction of coupled systems. *Model order Reduction. Theory, Research Aspects and Applications*. 2008.

© В.Ф. Губарев, 2016

4. Gubarev V.F. Method of iterative identification of multivariable systems over inexact data. Part 1. Theoretical aspects. *Journal of Automation and Information Sciences*. 2006. № 5. pp. 16–31 (in Russian).
5. Gubarev V.F. and Melnichuk S.V. Identification of multivariable systems using steady-state parameters. *Journal of Automation and Information Sciences*. 2012. № 5. pp. 26–42 (in Russian).
6. Rainer S., Price K. Differential evolution – a simple and efficient adaptive scheme for global optimization over continuous spaces. *Journal of Global Optimization*. 1997. № 11. pp. 341–359.
7. Golub G.H. and Van Loan Ch.F. *Matrix Computations*. Baltimore and London: John Hopkins University Press. 1999. 550 p.
8. Melnichuk S.V. Method of structural parametric multivariable systems identification using frequency characteristics. *Cybernetics and Computer Engineering*. 2015. № 181. p. 66–79 (in Russian).

Получено 15.09.16