УДК 517.977

Применение метода наименьших квадратов для решения линейных дифференциально-алгебраических уравнений*

В.Ф. Чистяков, Е.В. Чистякова

Институт динамики систем и теории управления Сибирского отделения Российской академии наук, ул. Лермонтова, 134, Иркутск, 664048

E-mails: chist@icc.ru (Чистяков В.Ф.), elena.chistyakova@icc.ru (Чистякова Е.В.)

Чистяков В.Ф., Чистякова Е.В. Применение метода наименьших квадратов для решения линейных дифференциально-алгебраических уравнений // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 2013. — Т. 16, № 1. — С. 81–95.

Рассмотрено применение метода наименьших квадратов для численного решения линейных систем обыкновенных дифференциальных уравнений (ОДУ) с тождественно вырожденной или прямоугольной матрицей перед производной искомой вектор-функции. В работе обсуждается поведение градиентных методов для минимизации функционала квадрата невязки в пространствах Соболева и некоторые другие вопросы. Приведены результаты численных экспериментов.

Ключевые слова: *дифференциально-алгебраические уравнения, индекс, метод наименьших квадратов, градиентные методы.*

Chistyakov V.F., Chistyakova E.V. Application of the least square method to the solving linear differential-algebraic equations // Siberian J. Num. Math. / Sib. Branch of Russ. Acad. of Sci. — Novosibirsk, 2013. — Vol. 16, № 1. — P. 81–95.

We consider application of the least square method to the numerical solution of a linear system of ordinary differential equations (ODEs) with an identically singular matrix multiplied a higher derivative by the desired vector-function. We discuss the behavior of the gradient method for minimizing the functional of the residual square in the Sobolev space and some other issues. The results of the numerical experiments are given.

Key words: *differential-algebraic equations, index, least square method, gradient methods.*

1. Постановка задачи

Рассмотрим задачу

$$\Lambda_1 x := A(t)\dot{x} + B(t)x = f, \quad t \in T = [\alpha, \beta], \tag{1}$$

$$x(\alpha) = a, \tag{2}$$

где $A(t)$, $B(t)$ — $(m \times n)$-матрицы, $x \equiv x(t)$, $f \equiv f(t)$ — искомая и заданная вектор-функции соответственно, a — заданный вектор из $\boldsymbol{R}^n$, $\dot{z} := dz(t)/dt$. Предполагается, что входные данные достаточно гладкие и выполнено условие

$$\operatorname{rank} A(t) < \min\{m, n\} \ \forall\, t \in T. \tag{3}$$

*Работа поддержана грантами междисциплинарного проекта СО РАН (проект № 107) и РФФИ (проекты № 05-08-18160, № 11-01-93005-Вьет-а.)

© Чистяков В.Ф., Чистякова Е.В., 2013

Система (1) называется *замкнутой*, если число уравнений равно числу компонент искомой вектор-функции ($m = n$), *переопределенной*, если $m > n$, и *недоопределенной*, если $m < n$. Для замкнутой системы условие (3) эквивалентно равенству $\det A(t) \equiv 0$, $t \in T$.

Системы вида (1), удовлетворяющие условию (3), принято называть дифференциально-алгебраическими уравнениями (ДАУ) [1] или алгебро-дифференциальными системами (АДС) [2]. Можно использовать также названия "сингулярные системы", "дескрипторные системы".

Под решением задачи (1), (2) на T мы будем понимать вектор-функцию $x(t) \in \boldsymbol{C}^1(T)$, которая обращает уравнение (1) в тождество на T и удовлетворяет условию (2).

Несмотря на то, что численные методы решения дифференциально-алгебраических уравнений исследуются около 40 лет, тематика по-прежнему остается актуальной. Находятся все новые эффекты, затрудняющие применение разностных схем и даже для замкнутых систем. Эти обстоятельства приводят к выводу, что нужно искать другие подходы к построению численных методов решения ДАУ. Ряд авторов предлагает использовать метод наименьших квадратов в том или ином варианте [3–8]. Совсем недавно опубликована работа [9], в которой описан набор подходов к решению ДАУ на основе вариационных принципов. Приведены результаты численных экспериментов. Определенным пробелом статьи является отсутствие оценок, связывающих значения функционалов на приближающих решение ДАУ функциях и отклонения этих функций от решений. В данной работе основное внимание уделено именно этому аспекту проблемы.

В работе используются евклидова и равномерная нормы q-мерного вектора $b \in \boldsymbol{R}^q$, вычисляемые соответственно по правилам

$$\|b\|_q^2 = b_1^2 + b_2^2 + \cdots + b_q^2; \qquad \|b\|_I = \max\{|b_i|,\ i = 1, 2, \ldots, q\},\ \ b = (b_1, b_2, \ldots, b_q)^\top,$$

где $\top$ — символ транспонирования. Включения $b(w)$, $V(w) \in \boldsymbol{C}^\varrho(\boldsymbol{D})$, $w \in \boldsymbol{D} \subseteq \boldsymbol{R}^q$ означают, что все частные производные элементов вектор-функции $b(w)$ или матрицы $V(w)$ непрерывны до порядка ϱ включительно по всем компонентам вектора w в любой точке области $\boldsymbol{D}$. Непрерывности соответствуют обозначения: $b(w)$, $V(w) \in \boldsymbol{C}(\boldsymbol{D})$.

Метод наименьших квадратов в работе формулируется так. Исходная задача заменяется задачей поиска минимума функционала

$$\begin{aligned}\mathcal{Q}_k(x) &= \sum_{j=0}^{k} \int_\alpha^\beta \left\| \left(\frac{d}{dt}\right)^j [\Lambda_1 x - f] \right\|_m^2 dt = \sum_{j=0}^{k} \left\| \left(\frac{d}{dt}\right)^j [\Lambda_1 x - f] \right\|_{\mathcal{H}_m}^2 \\ &= \|D_k[A,\ B](t)\mathrm{d}_{k+1}[x] - \mathrm{d}_k[f]\|_{\mathcal{H}_\mu}^2, \quad x(\alpha) = a, \end{aligned} \tag{4}$$

где k — некоторое целое число, $\mu = (k+1)m$, $\|\cdot\|_{\mathcal{H}_m}$ — норма в гильбертовом пространстве $L_{2,m}(T)$, а именно $h \in \mathcal{H}_m :\ \|h\|_{\mathcal{H},m}^2 = \int_\alpha^\beta \|h(t)\|_m^2 dt$,

$$D_k[A,\ B](t) = \begin{pmatrix} B & A & 0 & \cdots & 0 \\ \dot{B} & \dot{A} + B & A & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ B^{(k)} & C_k^1 B^{(k-1)} + C_k^2 A^{(k)} & C_k^2 B^{(k-2)} + C_k^3 A^{(k-1)} & \cdots & A \end{pmatrix},$$

$\mathrm{d}_k[f]^\top = (f^\top, \dot{f}^\top, \ldots, f^{\top(k)})^\top$ и $(x^\top, \dot{x}^\top, \ldots, x^{\top(k+1)})^\top$; $C_k^\nu = k!/(k-\nu)!\nu!$ суть биномиальные коэффициенты. В дальнейшем нам потребуются и такие представления матрицы $D_k[A,\ B](t)$:

$$D_k[A,\ B](t) = \begin{pmatrix} 0 & \mathcal{M}_k[A(t)] \end{pmatrix} + \begin{pmatrix} \mathcal{M}_k[B(t)] & 0 \end{pmatrix} = \begin{pmatrix} \mathrm{d}_k[B(t)] & \Gamma_k[A(t), B(t)] \end{pmatrix}, \qquad (5)$$

где нулевые блоки имеют размерность $([k+1]m \times n)$,

$$\mathcal{M}_k[A(t)] = \begin{pmatrix} C_0^0 A(t) & 0 & \cdots & 0 \\ C_1^0 A^{(1)}(t) & C_1^1 A(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C_k^0 A^{(k)}(t) & C_k^1 A^{(k-1)}(t) & \cdots & C_k^k A(t) \end{pmatrix}.$$

Замечание 1. Для упрощения записи указание зависимости от t в работе будет иногда опускаться, если это не вызывает путаницы.

2. Некоторые свойства линейных ДАУ

Для того, чтобы определить параметр k в формуле (4), введем такие понятия. Известно [10, с. 335], что для любого пучка постоянных $(m \times n)$-матриц $\lambda A + B$, где λ — параметр (в общем случае комплексный), существуют постоянные матрицы P, Q подходящей размерности со свойствами: $\det P \det Q \neq 0$,

$$P(\lambda A + B)Q = \operatorname{diag}\{\lambda E_d + J, \lambda N_{k_1} + E_{k_1}, \ldots, \lambda N_{k_p} + E_{k_p}, \\ \lambda L_{\eta_1} + M_{\eta_1}, \ldots, \lambda L_{\eta_q} + M_{\eta_q}, \lambda L^\star_{\nu_1} + M^\star_{\nu_1}, \ldots, \lambda L^\star_{\nu_v} + M^\star_{\nu_v}, 0\}, \qquad (6)$$

где E — единичные матрицы размерности, равной индексу, J — некоторый $(d \times d)$-блок,

$$N_{k_j} = \begin{pmatrix} 0 & E_{n_j-1} \\ 0 & 0, \end{pmatrix},\ j = 1, \ldots, p, \quad L^\star_{\nu_1} = \begin{pmatrix} E_{\nu_j - 1} \\ 0 \end{pmatrix}, \quad M^\star_{\nu_j} = \begin{pmatrix} 0 \\ E_{\nu_j - 1} \end{pmatrix},\ j = 1, \ldots, v,$$

$$L_{\eta_1} = \begin{pmatrix} E_{\eta_j - 1} & 0 \end{pmatrix}, \qquad M_{\eta_j} = \begin{pmatrix} 0 & E_{\eta_j - 1} \end{pmatrix},\ j = 1, \ldots, q,$$

и в блоках L, M нулевые подблоки являются либо столбцом, либо строкой. Представление (6) носит название *канонической структуры пучка*, и оно появилось в работах Вейерштрасса и Кронекера. Опираясь на представление (6) в монографии [10], получены необходимые и достаточные условия разрешимости системы (1) с постоянными матрицами коэффициентов. Для нестационарных систем условие приводимости пары матриц $A(t)$, $B(t)$ к виду (6) является слишком жестким. Учтем лишь основные свойства.

Пусть в системе (1), по крайней мере, $A(t)$, $B(t) \in \boldsymbol{C}^l(T)$, $f(t) \in \boldsymbol{C}^l(T)$, $l \geq 1$. Потребуем существования матриц $P(t)$, $Q(t) \in \boldsymbol{C}^l(T)$ подходящей размерности со свойствами: $\det P(t) \det Q(t) \neq 0\ \forall\, t \in T$,

$$P(t)A(t)Q(t)\dot{z} + P(t)[B(t)Q(t) + A(t)\dot{Q}(t)]z \\ = \begin{pmatrix} E_d & 0 & 0 & 0 & 0 \\ 0 & N(t) & 0 & 0 & 0 \\ 0 & 0 & L(t) & 0 & 0 \\ 0 & 0 & 0 & L^\star(t) & 0 \\ 0 & 0 & 0 & 0 & 0_{m_3 \times n_3} \end{pmatrix} \dot{z} + \begin{pmatrix} J(t) & 0 & 0 & 0 & 0 \\ 0 & E_{d_1} & 0 & 0 & 0 \\ 0 & 0 & M(t) & 0 & 0 \\ 0 & 0 & 0 & M^\star(t) & 0 \\ 0 & 0 & 0 & 0 & 0_{m_3 \times n_3} z \end{pmatrix} = Pf, \qquad (7)$$

где $t \in T$, $x = Q(t)z$, $J(t)$ — некоторый $(d \times d)$-блок, $N(t)$ — $(d_1 \times d_1)$-верхнетреугольная матрица с нулевой диагональю (матрицы $P(t)$, $Q(t)$ можно выбрать так, что на диагонали $N(t)$ стоят l нулевых квадратных блоков), для пучков матриц $\lambda L(t) + M(t)$, $\lambda L^\star(t) + M^\star(t)$ размерности $(m_1 \times n_1)$ и $(m_2 \times n_2)$ соответственно выполнены соотношения:

$$m_1 < n_1,\ \operatorname{rank} L(t) = \min\{m_1, n_1\}, \quad m_2 > n_2,\ \operatorname{rank} L^\star(t) = \min\{m_2, n_2\}\ \forall\, t \in T.$$

Здесь $d+d_1+m_1+m_2+m_3 = m$, $d+d_1+n_1+n_2+n_3 = n$. Для ясности уточним, что в случае постоянных матриц A, B в форме (6) блоку $\lambda L(t) + M(t)$ соответствует набор блоков $\lambda L_{\eta_1} + M_{\eta_1}, \ldots, \lambda L_{\eta_q} + M_{\eta_q}$, а блоку $\lambda L^\star(t) + M^\star(t)$ — набор $\lambda L^\star_{\nu_1} + M^\star_{\nu_1}, \ldots, \lambda L^\star_{\nu_v} + M^\star_{\nu_v}$.

Определение 1. Правую часть равенства (7) назовем обобщенной канонической формой (ОКФ) системы (1), а число l — ее индексом.

Нам потребуется такое понятие.

Определение 2 (см., например, [11]). Полуобратной матрицей к произвольной постоянной $(m \times n)$-матрице $\mathcal{A}$ называется $(n \times m)$-матрица, обозначаемая здесь и всюду в дальнейшем как $\mathcal{A}^-$, которая удовлетворяет уравнению $\mathcal{A}\mathcal{A}^-\mathcal{A} = \mathcal{A}$.

Замечание 2. Матричное уравнение $\mathcal{A}X\mathcal{A} = \mathcal{A}$, определяющее полуобратную матрицу, всегда разрешимо относительно матрицы X. Более того, все решения совместной системы $\mathcal{A}\chi = \zeta$ описываются формулой $\chi = \mathcal{A}^-\zeta + (E_n - \mathcal{A}^-\mathcal{A})W$, где W — произвольный вектор из $\boldsymbol{R}^n$.

Ниже мы установим соответствия между формой (7) и различными каноническими формами линейных ДАУ, введенных ранее.

Введем разбиения:

$$\begin{gathered} z = (z_1^\top \quad z_2^\top \quad z_3^\top \quad z_4^\top \quad z_5^\top)^\top, \\ (f_1^\top \quad f_2^\top \quad f_3^\top \quad f_4^\top \quad f_5^\top)^\top = (P_1^\top \quad P_2^\top \quad P_3^\top \quad P_4^\top \quad P_5)^\top f = Pf \end{gathered} \tag{8}$$

и исследуем структуру решений подсистем системы (7). Имеем

$$\begin{gathered} \dot z_1 + J(t) z_1 = f_1, \qquad N(t)\dot z_2 + z_2 = f_2, \\ L(t)\dot z_3 + M(t) z_3 = f_3, \qquad L^\star(t)\dot z_4 + M^\star(t) z_4 = f_4, \qquad 0_{m_3 \times n_3}\dot z_5 + 0_{m_3 \times n_3} z_5 = f_5. \end{gathered}$$

Тогда

$$z_1(t) = Z(t)c + \int_\alpha^t Z^{-1}(t)Z(s)f_1(s)\,ds, \qquad z_2(t) = f_2 - \mathcal{T} f_2 + \cdots + (-1)^l \mathcal{T}^{l-1} f_2, \tag{9}$$

где $Z(t)$ — матрицант системы $\dot v(t) = -J(t)v(t)$, $\mathcal{T}$ — оператор, действующий на вектор-функцию $h(t)$ по правилу: $\mathcal{T}h(t) = N(t)\dot h(t)$.

Далее, матрицы $L(t)$, $L^\star(t)$ имеют по условию постоянный полный ранг для любого $t \in T$. Согласно [12, с. 34], найдутся матрицы $P_*(t)$, $Q_*(t) \in \boldsymbol{C}^l(T)$ со свойствами: $\det P_*(t) \det Q_*(t) \neq 0\ \forall\, t \in T$,

$$\begin{gathered} LQ_*\dot u_3 + (MQ_* + L\dot Q_*)u_3 = \dot u_{3,1} + M_{11}u_{3,1} + M_{12}u_{3,2} = f_3, \quad z_3 = Q_*u_3 = Q_*\begin{pmatrix} u_{3,1} \\ u_{3,2} \end{pmatrix}, \\ z_3(t) = Q_*(t)\begin{pmatrix} Z_1(t)c_1 + \int\limits_\alpha^t Z_1^{-1}(t)Z_1(s)[f_3(s) + w_1(s)]\,ds \\ w_1(t) \end{pmatrix}, \end{gathered} \tag{10}$$

где $Z_1(t)$ — матрицант системы $\dot v(t) = -M_{11}(t)v(t)$, $c_1 \in \boldsymbol{R}^{m_2}$, $w_1(t)$ — произвольная вектор-функция размерности $n_2 - m_2$,

$$P_*(L^\star \dot z_4 + M^\star z_4) = \begin{pmatrix} E_{m_2} \\ 0 \end{pmatrix} \dot z_4 + \begin{pmatrix} M^\star_{11} \\ M^\star_{12} \end{pmatrix} z_4 = P_* f_4, \tag{11}$$

$$z_4(t) = Z_2(t)c_2 + \varphi(t), \qquad \varphi(t) = \int\limits_\alpha^t Z_2^{-1}(t) Z_2(s) P_{*,1}(s) f_4(s)\, ds, \qquad P_* f_4 = \begin{pmatrix} P_{*,1} \\ P_{*,2} \end{pmatrix} f_4, \tag{12}$$

где $Z_2(t)$ — матрицант системы $\dot v(t) = -M^\star_{11}(t)v(t)$, и вектор $c_2 \in \boldsymbol{R}^{n_2}$ удовлетворяет системе с переменными матрицами коэффициентов

$$\mathcal{L}(t)c_2 = \psi(t), \quad t \in T, \qquad \mathcal{L} = M^\star_{12} Z_2, \qquad \psi = P_{*,2} f_4 - M^\star_{12}\varphi. \tag{13}$$

Известно [11, с. 34], что система (13) имеет постоянные решения c_2 тогда и только тогда, когда

$$\psi(t) = \mathcal{L}(t)\mathcal{C}^- \theta, \tag{14}$$

где (с учетом замечания 2)

$$\mathcal{C} = \int\limits_\alpha^\beta \mathcal{L}^\top(s)\mathcal{L}(s)\, ds, \quad \theta = \int\limits_\alpha^\beta \mathcal{L}^\top(s)\psi(s)\, ds, \quad c_2 = \mathcal{C}^-\theta + [E_{n_2} - \mathcal{C}^-\mathcal{C}]W, \tag{15}$$

W — произвольный вектор из $\boldsymbol{R}^{n_2}$, $\mathcal{C}^-$ — полуобратная матрица к матрице $\mathcal{C}$.

Предположим, что $f_5(t) \equiv 0$, $t \in T$, в формулах (7), (8). Тогда

$$z_5(t) = w_2(t), \tag{16}$$

где $w_2(t)$ — произвольная вектор-функция размерности n_3.

Замечание 3. Отметим, что в случае постоянных матриц коэффициентов системы (1) решение системы (14) единственно. Действительно, рассмотрим однородную систему, соответствующую одному из блоков $\lambda L^\star_{\nu_j} + M^\star_{\nu_j}$ в канонической форме (6). Эта система уже имеет структуру, подобную системе (11):

$$M^\star_{11} = \begin{pmatrix} 0 & 0 \\ E_{\nu_j - 1} & 0 \end{pmatrix}, \quad M^\star_{12} = (0 \quad 0 \quad \cdots \quad 0 \quad 1), \quad Z_2(t) = E_{\nu_j} + \sum_{i=1}^{\nu_j - 1} (M^\star_{11})^i (t-\alpha)^i / i!,$$

так как матрица $M^\star_{11}$ — нильпотентная. Система (13) имеет вид:

$$M^\star_{12} Z_2(t) c_2 = (1 \quad (t-\alpha) \quad \cdots \quad (t-\alpha)^{\nu_j - 1}/(\nu_j - 1)!) c_2 = 0.$$

Решение только одно: $c_2 = 0$.

Из вида системы (7) и формул общих решений (9), (10), (12), (16) следует такое утверждение.

Теорема 1. *Пусть для системы* (1) *выполнены условия*:
1) $A(t)$, $B(t) \in \boldsymbol{C}^l(T)$, $f(t) \in \boldsymbol{C}^l(T)$, $l \geq 1$;
2) система приводима к виду (7), *и* l *равно количеству квадратных блоков на диагонали блока* $N(t)$;
3) $f_5(t) \equiv 0$, $t \in T$;
4) выполнено условие (14).

Тогда существуют гладкие в своих областях определения матрицы подходящей размерности

$$X_\nu(t), \quad X_1(t),\ K_1(t,s), \quad C_j(t), \quad j = 0, 1, \cdots, l-1, \quad \tilde C(t), \quad K_2(t,s)$$

такие, что любая линейная комбинация

$$x(t,c) = X_\nu(t)c + X_1(t)\mathcal{C}^-\theta + \int\limits_\alpha^t K_1(t,s)f(s)\,ds + \sum_{j=0}^{l-1} C_j(t)(d/dt)^j f(t) + \tilde{C}(t)w(t) + \int\limits_\alpha^t K_2(t,s)w(s)\,ds,\ t\in T, \tag{17}$$

где $c\in \boldsymbol{R}^\nu$, $\nu = d+n_2+n_4$, $n_4 = \operatorname{rank}[E_{n_2} - \mathcal{C}^-\mathcal{C}]$, $w(t) = (w_1^\top(t)w_2^\top(t))^\top$ — *произвольная вектор-функция размерности* $n_2 - m_2 + n_3$, $\operatorname{rank} X_\nu(t) = \nu\ \forall\, t\in T$ *является решением системы* (1) *и на отрезке* T *нет других решений.*

Более того, начальная задача (1), (2) *имеет решение тогда и только тогда, когда разрешима относительно вектора* c *линейная система*:

$$X_\nu(\alpha)c = a - X_1(\alpha)\mathcal{C}^-\theta - \sum_{j=0}^{l-1} C_j(t)(d/dt)^j f(t)|_{t=\alpha} - \tilde{C}(\alpha)W, \tag{18}$$

где W — *произвольный вектор. Решение задачи* (1), (2) *единственно тогда и только тогда, когда в ОКФ* (7) *отсутствуют блоки* $\lambda L(t) + M(t)$ *и* $0_{m_3\times n_3}$.

Обсудим условия, при выполнении которых система (1) приводима к виду (7). К настоящему времени получены только частные утверждения.

Теорема 2. *Пусть* $A(t), B(t)\in \boldsymbol{C}^A(T)$ — *пространству вещественно-аналитических функций, система* (1) *замкнута* ($m=n$), *и выполнены соотношения*:

$$\operatorname{rank}\Gamma_{r+1}[A(t),B(t)] = \varrho = \operatorname{const} \ge n(r+1)\ \forall\, t\in T, \tag{19}$$

где $r = \max\operatorname{rank}\{A(t),\ t\in T\}$, *и другие обозначения взяты из* (5).

Тогда

1) *существуют матрицы* $P(t)$, $Q(t)\in \boldsymbol{C}^A(T)$ *со свойствами:* $\det\ P(t)Q(t)\ne 0\ \forall\, t\in T$,

$$P(t)A(t)Q(t)\dot z + P(t)[B(t)Q(t) + A(t)\dot Q(t)]z = \begin{pmatrix} E_d & 0\\ 0 & N(t)\end{pmatrix}\dot z + \begin{pmatrix} J(t) & 0\\ 0 & E_{n-d}\end{pmatrix} z = \begin{pmatrix} P_1(t)\\ P_2(t)\end{pmatrix} f(t),\quad t\in T, \tag{20}$$

где обозначения соответствуют обозначениям формулы (7);

2) *существует оператор* $\Lambda_{l,*} = \sum\limits_{j=0}^{l} L_j(t)(d/dt)^j$, $t\in T$, *где* $L_j(t)$ — *матрицы из* $\boldsymbol{C}^A(T)$, *со свойством*

$$(\Lambda_{l,*}\circ\Lambda_1)y = \dot y + \Lambda_{l,*}[B(t)]y\ \forall\, y\in \boldsymbol{C}^{l+1}(T);$$

3) *система разрешима при любой* $f\in \boldsymbol{C}^l(T)$, *и формула* (17) *имеет вид*

$$x(t,c) = X_d(t)c + \omega(t), \omega(t) = \int\limits_\alpha^t K_1(t,s)f(s)\,ds + \sum_{j=0}^{l-1} C_j(t)(d/dt)^j f(t),\quad d = \varrho - n(r+1); \tag{21}$$

4) *задача* (1), (2) *имеет единственное решение тогда и только тогда, когда разрешима система* $X_d(\alpha)c = a - \omega(\alpha)$.

Более того, если для замкнутой системы выполнен любой из пунктов утверждения, то выполнены соотношения (19).

Теорема 2 является сводкой результатов из монографии [2]. Там же даны критерии существования оператора $\Lambda_{l,*}$ и указаны способы построения.

Замечание 4. Правую часть равенства (20) принято называть центральной канонической формой (ЦКФ) системы (4). Это понятие было введено в [13] и сыграло большую роль в развитии теории ДАУ. Точнее говоря, оно появилось в препринте тех же авторов в 1982 г. Для незамкнутых систем в [14] введена некоторая структурная форма, но для целей данной работы она не подходит.

Замечание 5. Если матрицы коэффициентов гладкие: $A(t)$, $B(t) \in \boldsymbol{C}^{2r+3}(T)$, то выполняются только пункты 2), 3) утверждения теоремы 2. ДАУ в этом случае приводима к ЦКФ только локально. А именно, любой отрезок $T_1 = [\alpha_1, \beta_1] \subseteq T$ содержит подотрезок $T_2 = [\alpha_2, \beta_2] \subseteq T_1$, на котором существуют матрицы $P(t)$, $Q(t) \in \boldsymbol{C}^{r+2}(T_2)$, приводящие систему к ЦКФ.

Пример 1. Рассмотрим задачу

$$A(t)\dot{x} + x = f(t),\ t \in T, \qquad A(t) = \begin{pmatrix} 0 & u(t) \\ u(t) & 0 \end{pmatrix} \in \boldsymbol{C}^{\infty}(T),$$

где $u(t)v(t) = 0\ \forall t \in T$, $A^2(t) = 0$. Условия (19) здесь выполнены, существует оператор $\Lambda_{l,*}$ из теоремы 2, где $l = 2$, и решение представимо в виде (21): $x(t) = f(t) - A(t)\dot{f}(t)$. Легко указать конкретный вид матрицы $A(t)$, для которой не существует матрицы $P(t)$: $P(t) \in \boldsymbol{C}(T)$, $\det P(t) \neq 0\ \forall t \in T$, обращающей в нуль строку матрицы $A(t)$.

Пример взят из [1, с. 24]. Аналогичный по свойствам, но не такой прозрачный пример построен в [12, с. 34]. В настоящее время авторы полагают, что при достаточно гладких входных данных любая система (1) локально (в смысле замечания 5) приводима в виду (7). Конечно, размерности блоков ОКФ могут зависеть от выбора отрезка приведения.

3. Применение метода градиентного спуска

Теорема 1 выделяет класс ДАУ, для которых можно в формуле (4) указать значение параметра k, гарантирующее близость приближающей решение функции к множеству решений при малых значениях функционала (4). Следуя [15], сформулируем утверждение.

Лемма 1. *Пусть скалярные функции $h(t) \in \boldsymbol{C}(T)$ и $g(t) \in L_2(T)$. Тогда справедливы включение и оценка $h(t)g(t) \in L_2(T)$, $\|h(t)g(t)\|_{\mathcal{H}} \leq \|h(t)\|_{\boldsymbol{C}}\|g(t)\|_{\mathcal{H}}$, где $\|h(t)\|_{\boldsymbol{C}} = \max\{|h(t)|,\ t \in T\}$.*

Лемма 2. *Если выполнены условия теоремы 1 и для некоторой вектор-функции $x_\epsilon \in \boldsymbol{C}^l(T)$, $x_\epsilon(\alpha) = a$ функционал $\mathcal{Q}_{l-1}(x_\epsilon) < \epsilon$, то найдется решение задачи (1), (2) $x_a(t)$ такое, что*

$$\|x_a(t) - x_\epsilon(t)\|^2_{\mathcal{H}_n} \leq \kappa\epsilon, \quad \kappa = \mathrm{const} > 0.$$

Доказательство. Имеем $\Lambda_1 x = f, \Lambda_1 x_\epsilon = f + f_\epsilon$,

$$\Lambda_1 \boldsymbol{w} = f_\epsilon, \quad \boldsymbol{w}(\alpha) = 0, \tag{22}$$

где $\boldsymbol{w} = x_\epsilon - x$, $f_\epsilon \equiv f_\epsilon(t)$, и согласно условию теоремы

$$\sum_{j=0}^{l-1} \|f_\epsilon^{(j)}(t)\|_{\mathcal{H}}^2 \le \epsilon. \tag{23}$$

Выпишем выражение для $\boldsymbol{w}$, используя формулу (17):

$$\boldsymbol{w}(t,c) = X_\nu(t)c + X_1(t)\mathcal{C}^-\theta_\epsilon + \int_\alpha^t K_1(t,s)f_\epsilon(s)\,ds + \sum_{j=0}^{l-1} C_j(t)(d/dt)^j f_\epsilon(t), \quad t \in T, \tag{24}$$

где принято, что произвольная вектор-функция $w(t) = 0$ и вектор c определяется из выражения

$$X_\nu(\alpha)c = X_1(\alpha)\mathcal{C}^-\theta_\epsilon - \sum_{j=0}^{l-1} C_j(t)(d/dt)^j f_\epsilon(t)\mid_{t=\alpha}. \tag{25}$$

Обратим внимание, что в формуле, аналогичной (9),

$$z_{2,\epsilon}(\alpha) = [f_{2,\epsilon} - \mathcal{T}f_{2,\epsilon} + \cdots + (-1)^l\mathcal{T}^{l-1}f_{2,\epsilon}]\mid_{t=\alpha} = 0,$$

где $Q(t)z_\epsilon(t) = \boldsymbol{w}(t)$, и обозначения соответствуют обозначениям из (7), так как по условию $\boldsymbol{w}(\alpha) = 0$. Следовательно, уравнение (25) и его решение имеют вид

$$X_\nu(\alpha)c = X_1(\alpha)\mathcal{C}^-\theta_\epsilon, \quad c = X_\nu^-(\alpha)X_1(\alpha)\mathcal{C}^-\theta_\epsilon. \tag{26}$$

Здесь в силу полноты ранга матрицы $X_\nu(\alpha)$ и неравенства $\nu \le n$ полуобратная матрица к ней единственна и в формуле из замечания 2: $E_\nu - X_\nu(\alpha)X_\nu^-(\alpha) = 0$. Нам осталось оценить множитель θ_ϵ. С учетом формул (12) имеем

$$\theta_\epsilon = \int_\alpha^\beta \mathcal{L}^\top(s)\psi_\epsilon(s)\,ds,\ \psi_\epsilon = P_{*,2}f_{4,\epsilon} - M_{12}^\star\varphi_\epsilon,\ \varphi_\epsilon(t) = \int_\alpha^t Z_2^{-1}(t)Z_2(s)P_{*,1}(s)f_{4,\epsilon}(s)\,ds. \tag{27}$$

Итак, из формул (27), (26), (24) с учетом оценки (23) и леммы 1 можно последовательно (применяя неравенство Коши–Буняковского) получить соотношения:

$$\|\theta_\epsilon\|_{m_2-n_2}^2 \le \kappa\epsilon, \quad \|c\|_\nu^2 \le \kappa_1\epsilon, \quad \|\boldsymbol{w}(t,c)\|_{\mathcal{H}_n}^2 \le \kappa_2\epsilon, \tag{28}$$

где κ, κ_1, κ_2 — некоторые константы. □

Теперь нам нужно указать методы минимизации функционала (4). Начнем со случая, когда $l = 1$:

$$\mathcal{Q}_0(x) = \|A(t)\dot{x} + B(t)x - f\|_{\mathcal{H}_m}^2, \quad x(\alpha) = a.$$

Для применения метода градиентного спуска преобразуем эту задачу в задачу оптимального управления со свободным правым концом

$$\mathcal{Q}(u) = \|A(t)u + B(t)x - f\|_{\mathcal{H}_m}^2, \quad \dot{x} = u, \quad x(\alpha) = a. \tag{29}$$

Определение 3. Функционал $\Phi(u)$, определенный на выпуклом множестве $\mathcal{U}$ банахова пространства $\mathcal{E}$, называется выпуклым [16, с. 236], если

$$\Phi(\lambda u_1 + (1-\lambda)u_2) \le \lambda\Phi(u_1) + (1-\lambda)\Phi(u_2)$$

при всех $u_1,\ u_2 \in \mathcal{U}$ и всех $\lambda \in [0,\ 1]$. Если же существует константа $\kappa > 0$ такая, что

$$\Phi(\lambda u_1 + (1-\lambda)u_2) \le \lambda\Phi(u_1) + (1-\lambda)\Phi(u_2) - \kappa\lambda(1-\lambda)\|u_1 - u_2\|^2_{\mathcal{X}},$$

то функционал $\Phi(u)$ называется сильно выпуклым [16, с. 57].

Лемма 3. *Функционал $\mathcal{Q}(u)$ является выпуклым, непрерывно дифференцируемым в $\mathcal{H}_n = L_{2,n}(T)$ по Фреше* [16, с. 233] *и его градиент $\mathcal{Q}'(u)$ в точке $u = u(t) \in \mathcal{H}_n$ вычисляется по формулам*:

$$\mathcal{Q}'(u) = 2A^\top(t)[A(t)u + B(t)x - f] - \psi(t), \tag{30}$$

$$\dot\psi(t) = 2B^\top(t)[A(t)u + B(t)x - f], \qquad \psi(\beta) = 0. \tag{31}$$

Кроме того, градиент функционала $\mathcal{Q}'(u)$ удовлетворяет условию Липшица:

$$\|\mathcal{Q}'(u_1) - \mathcal{Q}'(u_2)\|_{\mathcal{H}_n} \le L\|u_1 - u_2\|_{\mathcal{H}_n}, \quad L = \mathrm{const} > 0, \tag{32}$$

для любых u_1, $u_2 \in \mathcal{H}_n$ и справедлива оценка $L \le \sqrt{12\chi}$, где

$$\chi = \max\{\|A^\top(t)A(t)\|,\ \|B^\top(t)A(t)\|,\ \|A^\top(t)B(t)\|t, \|B^\top(t)B(t)\|,\ t \in T\},$$

а под нормой матрицы понимается квадратный корень из суммы квадратов всех ее элементов.

Равенства (30), (31) и неравенство (32), а также оценка для L являются элементарными следствиями [16, гл. 6, § 3, теоремы 4, 5].

Определим для функционала (29), начиная с некоторого $u_0 \in \mathcal{H}$, итерационный процесс

$$u_{i+1} = u_i - \lambda_i\mathcal{Q}'(u_i), \quad i = 0, 1, 2, \dots, \quad \lambda_i = 1/L. \tag{33}$$

Для функционала (29) не выполнено важное свойство, а именно, множество $M(v) = \{u : u \in \mathcal{U},\ \mathcal{Q}(u) \le \mathcal{Q}(v)\}$ не ограничено при любом $v \in \mathcal{U}$ для ДАУ индекса 1 (по этому поводу см. [16, с. 248]).

Пример 2. Рассмотрим начальные задачи:

$$\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\dot h + \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}h = 0, \qquad \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\dot h + \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}h = 0, \quad t \in [0,\ 1], \quad h(0) = 0,$$

которые имеют только нулевые решения. Пусть приближающая решение вектор-функция h_ϵ дает при подстановке в системы малую невязку $f_\epsilon = \begin{pmatrix}0 & \sqrt{\epsilon}\,(1 - \cos\frac{t}{\epsilon})\end{pmatrix}^\top$. Тогда для первой системы

$$h_\epsilon = \begin{pmatrix}0 & \sqrt{\epsilon}(1 - \cos\frac{t}{\epsilon})\end{pmatrix}^\top, \quad \|u(t)\|_{\mathcal{H}_2} \to \infty$$

несмотря на то, что $\mathcal{Q}(u_\epsilon) \to 0$ при $\epsilon \to 0$. Здесь $u(t) = \dot h(t)$.

Отметим, что в первом случае норма самого решения стремится к нулю при стремлении к нулю нормы невязки. Во втором случае нет и этого:

$$h_\epsilon = \Big(-\tfrac{1}{\sqrt{\epsilon}}\sin\tfrac{t}{\epsilon} \quad \sqrt{\epsilon}(1 - \cos\tfrac{t}{\epsilon})\Big)^\top,\ \|h(t)\|_{\mathcal{H}_2} \to \infty,\ \epsilon \to 0.$$

Ниже нам может помочь такое замечательное утверждение из статьи [17].

Теорема 3. *Пусть на гильбертовом пространстве H определен квадратичный функционал $\mathcal{Q}(u) = \|\Lambda u - g\|_G^2$, где Λ — ограниченный линейный оператор из H в гильбертово пространство G, и существует точка минимума этого функционала.*

Тогда при любом u_0 в градиентном методе $\|u_i - u_{0,*}\|_{\mathcal{H}} \to 0$, $i \to \infty$, где $u_{0,*}$ — точка минимума $\mathcal{Q}(u)$, ближайшая к u_0.

Итак, мы имеем все необходимые средства, чтобы получить сведения об эффективности применения метода градиентного спуска при решении ДАУ.

Теорема 4. *Пусть на T выполнены условия теоремы 1 и разрешима система* (18).

Тогда для процесса (33) *справедливы оценки*:

$$\|u_i(t) - u_{0,*}(t)\|_{\mathcal{H}_n}^2 \to 0, \qquad \|x_i(t) - x_*(t)\|_{\boldsymbol{C}(T)} \to 0, \quad i \to \infty, \qquad \mathcal{Q}(u_i) = o\,(1/i)\,,$$

где $u_{0,*}$ — точка минимума $\mathcal{Q}(u)$, ближайшая к u_0, $x_{0,*}(t) = a + \int_\alpha^t u_{0,*}(s)\,ds$.

Доказательство. Запишем функционал (29) в виде

$$\mathcal{Q}(u) = \left\| A(t)u + B(t)\int\limits_\alpha^t u(s)\,ds - (f + B(t)a) \right\|_{\mathcal{H}_m}^2 = \|\Lambda u - g\|_{\mathcal{H}_m}^2, \tag{34}$$

где $g = f(t) + B(t)a$, $\Lambda u = A(t)u + B(t)\int_\alpha^t u(s)\,ds$ — линейный ограниченный оператор, действующий из $\mathcal{H}_n$ в $\mathcal{H}_m$. Ограниченность оператора в (34) вытекает из леммы 1. Согласно теореме 3, можно записать

$$\|u_i(t) - u_{0,*}(t)\|_{\mathcal{H}_n}^2 \to 0, \quad i \to \infty.$$

Сходимость последовательности производных функций в $L_{2,n}(T)$ обеспечивает сходимость последовательности функций в $\boldsymbol{C}(T)$.

Действительно, в силу исходных предположений о задаче $x_{*,0}(\alpha) = x_i(\alpha) = a$ для любого i и

$$\|x_i(t) - x_{*,0}(t)\|_I = \left\| \int\limits_\alpha^t (\dot{x}_i(s) - \dot{x}_{*,0}(s)\,ds \right\|_I \le \int\limits_\alpha^\beta \|\dot{x}_i(t) - \dot{x}_{*,0}(t)\|_I\,dt.$$

Неравенство Коши–Буняковского позволяет записать

$$\|x_i(t) - x_{*,0}(t)\|_I \le \|\dot{x}_i(t) - \dot{x}_{*,0}(t)\|_{\mathcal{H}_n}, \qquad \|x_i(t) - x_{0,*}(t)\|_{\boldsymbol{C}(T)} \to 0, \quad i \to \infty.$$

Перейдем к оценке скорости сходимости по функционалу. Рассмотрим равенство

$$\begin{aligned} \|u_{i+1} - u_{*,0}\|_{\mathcal{H}_n}^2 &= \left\| u_i - \frac{1}{L}\mathcal{Q}'(u_i) - u_{*,0} \right\|_{\mathcal{H}_n}^2 \\ &= \|u_i - u_{*,0}\|_{\mathcal{H}_n}^2 - \frac{2}{L}(\mathcal{Q}'(u_i),\ u_i - u_{*,0}) + \frac{1}{L^2}\|\mathcal{Q}'(u_i)\|_{\mathcal{H}_n}^2. \end{aligned} \tag{35}$$

Аналогично [16, с. 67], выпишем для выпуклого функционала неравенства вида

$$\mathcal{Q}(u_i) - \mathcal{Q}(u_{*,0}) \le (\mathcal{Q}'(u_i),\ u_i - u_{*,0}), \qquad \mathcal{Q}(u_i) - \mathcal{Q}(u_{i+1}) \ge \frac{1}{2L}\|\mathcal{Q}'(u_i)\|^2_{\mathcal{H}_n}. \tag{36}$$

С использованием неравенств (36) равенство (35) можно преобразовать в неравенство

$$\|u_i - u_{*,0}\|^2_{\mathcal{H}_n} - \|u_{i+1} - u_{*,0}\|^2_{\mathcal{H}_n} \ge \frac{2}{L}[\mathcal{Q}(u_{i+1}) - \mathcal{Q}(u_{*,0})]. \tag{37}$$

Суммируя неравенства (37) от нуля до i, получим

$$\|u_0 - u_*\|^2_{\mathcal{H}_n} - \|u_{i+1} - u_{*,0}\|^2_{\mathcal{H}_n} \ge \frac{2}{L}\sum_{j=0}^{i}[\mathcal{Q}(u_{j+1}) - \mathcal{Q}(u_{*,0})]. \tag{38}$$

Из неравенства (38) вытекает, что ряд $\sum_{i=1}^{\infty}[\mathcal{Q}(u_{i+1}) - \mathcal{Q}(u_{*,0})]$, состоящий из монотонно убывающих положительных членов, сходится. Это и доказывает лемму, так как известно [18, с. 289], что, если ряд $\sum_{i=0}^{\infty} a_i$ сходится, его члены положительны и монотонно убывают, то $a_i = o(1/i)$. □

Следствие 1. *Если в условиях теоремы 3 индекс системы равен* 1, *то справедлива оценка*

$$\|x_i(t) - x_{*,0}(t)\|^2_{\mathcal{H}_n} = o\,(1/i)\,.$$

Доказательство вытекает из леммы 2 и теоремы 4. Для замкнутых систем проверка принадлежности к классу ДАУ индекса 1 основывается на таком утверждении.

Теорема 5. *Если* $A(t),\ B(t) \in \boldsymbol{C}^m(T)$, $m \ge 1$, *то система* (4) *имеет индекс* 1 *тогда и только тогда, когда для пучка матриц* $\lambda A(t) + B(t)$ *выполняется критерий "ранг–степень"*

$$\deg \det[\lambda A(t) + B(t)] = \operatorname{rank} A(t) = r = \text{const},$$

причем матрицы $P(t),\ Q(t)$ *можно выбрать из* $\boldsymbol{C}^m(T)$ *и в ЦКФ* $N(t) \equiv 0$, $t \in T$.

Замечание 6. В известных руководствах (см., например, [16, 19]) показано, что для выпуклых функционалов $\Phi(u_i) \le \kappa/i$, причем для константы κ существуют эффективные оценки, и равенство $\Phi(u_i) = o(1/i)$ мало что дает в практическом плане.

Сходимость же по аргументу может быть сколь угодно медленной в случае индекса $l > 1$, так как из (17) следует: малое значение $\mathcal{Q}(u_i) = o(1/i)$ не гарантирует малости $\|x_i(t) - x_{*,0}(t)\|^2_{\mathcal{H}_n}$ (см. пример 2).

В связи с этим замечанием для существования гарантированных оценок нужно выбирать функционал (4), учитывающий индекс ДАУ, и указать способы вычисления градиента для него. Поступим следующим образом. Функционал (4) перепишем в следующем виде:

$$\mathcal{Q}_k(\boldsymbol{u}) = \|\boldsymbol{A}_k(t)\boldsymbol{u} + \boldsymbol{B}_k(t)\boldsymbol{x} - \boldsymbol{f}_k\|^2_{\mathcal{H}_\mu}, \quad \dot{\boldsymbol{x}} = \boldsymbol{u}, \quad \boldsymbol{x}(\alpha) = \boldsymbol{a}, \tag{39}$$

где $k = l - 1$, $\boldsymbol{A}_k(t) = \Gamma_k[A(t), B(t)]$, $\boldsymbol{B}_k(t) = \begin{pmatrix} 0 & \mathrm{d}_k[B(t)] \end{pmatrix}$, $\boldsymbol{f}_k = \mathrm{d}_k[f(t)]$, $\boldsymbol{x} = \mathrm{d}_k[x(t)]$.

Для функционала (39) мы можем выписать выражение для градиента по формулам из леммы 3. Препятствием является то обстоятельство, что нам необходимо для организации вычислительного процесса знать не только вектор $a = x(\alpha)$, но и векторы $a_1 = \dot{x}_a(\alpha),\ \ldots, a_k = x_a^{(k)}(\alpha)$, где $x_a(t)$ — решение задачи (1), (2). В настоящее время эти величины можно вычислить в условиях теоремы 2 по формуле

$$Z = \boldsymbol{A}_{2k}^{-}(\alpha)[\mathcal{B}_{2k}(\alpha)a - \boldsymbol{b}_{2k}(\alpha)],$$

где $\mathcal{B}_{2k}(t) = \mathrm{d}_{2k}[B(t)]$, $\boldsymbol{f}_{2k}(t) = \mathrm{d}_{2k}[f(t)]$. Из [2] известно, что первые kn компонент вектора Z определяются единственным образом и равны соответственно $a_1, \ldots, a_k$.

4. Применение методов конечномерной оптимизации

Сложности, связанные с организацией градиентного метода для ДАУ индекса больше 1, заставляют искать другие подходы к минимизации функционала (4).

Будем искать приближающую вектор-функцию $x_\varepsilon(t)$ в виде векторного полинома

$$\mu_i(t) = \sum_{j=0}^{i} c_j (t-\alpha)^j, \quad i > l, \quad c_0 = a. \tag{40}$$

Подставим (40) в (4) и сведем задачу (1), (2) к минимизации функции

$$\Phi_k(c_1, c_2, \ldots, c_i) = \sum_{j=0}^{k} \left\| \left(\frac{d}{dt}\right)^j (\Lambda_1 \mu_i(t) - f) \right\|_{\mathcal{H}_m}^2, \quad k = l-1. \tag{41}$$

Дифференцируя функцию $\Phi_k(c_1, c_2, \ldots, c_i)$ по ее аргументам с учетом обозначений из формул (4), (39), получим систему линейных алгебраических уравнений относительно вектора $C = (c_1^\top, c_2^\top, \ldots, c_i^\top)^\top$:

$$\left\{ \frac{\partial \Phi_k(c_1, c_2, \ldots, c_i)}{\partial c_j},\ j = 1, 2, \ldots,\ i \right\} = \mathcal{A}C - \boldsymbol{b} = 0, \tag{42}$$

$$\mathcal{A} = \int_\alpha^\beta G^\top(t) G(t)\, dt, \qquad \boldsymbol{b} = \int_\alpha^\beta G^\top(t)[\boldsymbol{B}_k(t)a - \boldsymbol{f}_k(t)]\, dt, \qquad G(t) = D_k[A(t), B(t)]\mathcal{T}(t),$$

$$\mathcal{T}(t) = \begin{pmatrix} tE & t^2E & \cdots & t^iE \\ E & 2tE & \cdots & it^{i-1}E \\ 0 & 2E & \cdots & i(i-1)t^{i-2}E \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & i(i-1)\cdots(i-k-1)t^{i-k-1}E \end{pmatrix}, \quad E = E_n. \tag{43}$$

Лемма 4. *Пусть для для задачи* (1), (2) *выполнены условия теоремы 2.*

Тогда система (42) *имеет единственное решение* $c_1^\star, c_2^\star, \ldots, c_i^\star$ *и*

$$\Phi_k^\star = \Phi_k(c_1^\star, c_2^\star, \ldots, c_i^\star) = \min\{\Phi_k(c_1, c_2, \ldots, c_i),\ c_1, c_2, \ldots, c_i \in \boldsymbol{R}^n\}.$$

Доказательство. Рассмотрим функцию (41) для однородной задачи (1), (2), когда $a = 0$, $f(t) = 0$, $t \in T$. Элементарные преобразования над квадратичными функциями двух многочленов: $\mu_{i,1}(t)$, $\mu_{i,2}(t)$ позволяют записать

$$\Phi_k(\lambda\mu_{i,1} + [1-\lambda]\mu_{i,2}) = \lambda\Phi_k(\mu_{i,1}) + [1-\lambda]\Phi_k(\mu_{i,2}) - \lambda[1-\lambda]\Phi_k(\mu_{i,1} - \mu_{i,2}), \quad \lambda \in [0,1].$$

Из представления решения (21) следует, что

$$\Phi_k(\mu_{i,1}-\mu_{i,2}) \geq \sum_{j=1}^{i} \kappa_j \|c_{j,1}-c_{j,2}\|_n^2, \quad \kappa_j = \text{const}\,.$$

Здесь учтено, что при заданном начальном значении решение задачи (1), (2) единственно. Следовательно, функция $\Phi_k(c_1, c_2, \ldots, c_i)$ является строго выпуклой и имеет единственный минимум [16]. □

Теорема 6. *Пусть для для задачи* (1), (2) *выполнены условия теоремы 2.*
Тогда

$$\Phi_k^\star \leq \frac{\kappa_1}{(i^2\cdot(\ln i)^2)}, \qquad \|\mu_i^\star(t)-x_a(t)\|_{\mathcal{H}_n}^2 \leq \frac{\kappa_2}{(i^2\cdot(\ln i)^2)},$$

где $\mu_i^\star(t)=\sum_{j=0}^{i} c_j^\star (t-\alpha)^j$, $\kappa_1=\text{const}$, $\kappa_2=\text{const}$.

Доказательство. Рассмотрим функционал на многочлене $\boldsymbol{M}_i(t)$, у которого $\boldsymbol{M}_i^{(l)}(t)$ есть многочлен Чебышева для $x_a^{(l)}(t)$. Согласно [20], имеем

$$\max\{\|\mathbf{M}_i^{(l)}(t)-x_a^{(l)}(t)\|_I,\ t\in T\} = O(1/(i\cdot\ln i)), \qquad \Phi_k(\boldsymbol{M}_i)=O(1/(i^2\cdot(\ln i)^2)).$$

Так как на полиноме $\mu_i^\star(t)$ достигается минимум, то $\Phi_k^\star \leq \Phi_l(\boldsymbol{M}_i)$. Из леммы 2 вытекает справедливость второй оценки утверждения. □

Замечание 7. При определенных требованиях на гладкость решения можно брать другие пробные многочлены, например интерполяционный многочлен Лежандра, и получать другие оценки скорости сходимости.

Замечание 8. В лемме 4 и теореме 6 требование выполнения условий теоремы 2 можно заменить на требование отсутствия в ОКФ (7) блоков вида $\lambda L(t)+M(t)$ и $0_{m_3\times n_3}$, что обеспечит единственность решения задачи (1), (2). Но в настоящее время нет утверждения, гарантирующего такую структуру.

При больших i система (42) имеет плохую обусловленность. Для преодоления этого эффекта можно выбирать либо другие координатные функции (например, многочлены Лежандра), либо применить следующий прием: ввести сетку $\Delta=\{t_j : t_j=\alpha+\tau j,\ j=0,1,\ldots,N,\ \tau=(\beta-\alpha)/N\}$ и искать минимум на подотрезках $[t_j,\ t_{j+1}]$, используя сплайны порядка не меньше индекса l. Но здесь возникает задача доказательства устойчивости численного процесса при $\tau\to 0$. Эта проблема пока не решена. Как показали численные эксперименты, на устойчивость процесса оказывают существенное влияние значения дефекта сплайна.

5. Численные эксперименты

Пример 3. Рассмотрим начальную задачу

$$\Lambda_1 x = \begin{pmatrix} \gamma & e^{-t} & 0 \\ 0 & \delta & e^{-t} \\ \gamma & e^{-t} & 1 \\ \gamma e^{t} & 1 & 0 \end{pmatrix} \dot{x}(t) + \begin{pmatrix} e^{t} & 0 & e^{-t} \\ 0 & 1 & 0 \\ e^{t} & 0 & e^{-t} \\ e^{2t} & 0 & 2 \end{pmatrix} x(t) = \begin{pmatrix} -\gamma e^{-t}+4 \\ (\delta+1)e^{t}+2 \\ -\gamma e^{-t}+2e^{t}+4 \\ 6e^{t}-\gamma \end{pmatrix}, \quad t\in[0,1],$$

$$x(0)=\begin{pmatrix}1 & 1 & 2\end{pmatrix}^\top, \qquad x(t)=\begin{pmatrix}e^{-t} & e^{t} & 2e^{t}\end{pmatrix}^\top,$$

где γ, δ — вещественные параметры. Решение единственно при любых значениях параметров. Если $\gamma \neq 0$, $\delta \neq 0$, то система имеет индекс $l = 0$, $\nu = \dim \ker\Lambda_1 = 2$. Если $\gamma \neq 0$, $\delta = 0$, то система имеет индекс $l = 1$, $\nu = \dim \ker\Lambda_1 = 1$. Если $\gamma = 0$, $\delta = 0$, то система имеет индекс $l = 2$, $\nu = \dim \ker\Lambda_1 = 0$. Если γ, δ положительны и малы, то система из нашего примера является жесткой в том смысле, что в соответствующей подсистеме $\dot{z}_1 = -J(t)z_1 + P_1(t)f(t)$ из формулы (7) матрицы $-J(t)$ имеют отрицательные и большие по модулю собственные числа (см., например, [21]).

В таблице 1 приведены результаты решения методом градиентного спуска. Интегрирование в формуле (31) и вычисление функционала проводилось по методу трапеций с шагом $\tau = 0.02$, число итераций $i = 10000$, параметры $\gamma = \delta = 1$, $\lambda = 0.0255$.

В таблице 2 приведены результаты решения методом конечномерной оптимизации. Интегрирование в формуле (43) и вычисление функционала проводилось по методу Симпсона с шагом $\tau = 0.01$ и порядком многочлена $i = 6$.

Таблица 1.

индекс l	$Q(u_i)$	$\|x_*(t) - x_i(t)\|^2_{\mathcal{H}_3}$
0	0.00016	$4.174{\cdot}10^{-9}$
1	0.003277	0.001227
2	0.003985	0.0016

Таблица 2.

индекс l	$\Phi_1^\star$	$\|x_*(t) - \mu_6(t)\|^2_{\mathcal{H}_3}$
0	$1.12{\cdot}10^{-14}$	$6.06{\cdot}10^{-14}$
1	$2.07{\cdot}10^{-13}$	$1.03{\cdot}10^{-13}$
2	$1.86{\cdot}10^{-12}$	$1.08{\cdot}10^{-11}$

Следует отметить, что полученные многочлены хорошо приближают решение и в равномерной метрике. Ошибка не превышает квадратного корня из квадрата нормы ошибки в $\mathcal{H}_3$.

6. Заключение

Как показал анализ результатов численных экспериментов, метод градиентного спуска (в рассмотренной форме) не может в большинстве случаев конкурировать ни по точности, ни по вычислительным затратам с методами решения для замкнутых ДАУ, основанных на применении разностных схем (если, конечно, разностная схема сходится). Метод конечномерной оптимизации дает хорошую точность, но превосходит по трудоемкости методы, основанные на применении разностных схем.

Основной областью применения метода наименьших квадратов можно считать поиск решений ДАУ с прямоугольными матрицами. Авторам не известны обоснованные разностные схемы для таких систем.

Литература

1. **Brenan K.E., Campbell S.L., and Petzold L.R.** Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations (Classics in Applied Mathematics 14). — Philadelphia: SIAM, 1996.
2. **Бояринцев Ю.Е., Чистяков В.Ф.** Алгебро-дифференциальные системы. Методы решения и исследования. — Новосибирск: Наука, 1998.
3. **Чистяков В.Ф.** К методам решения сингулярных линейных систем обыкновенных дифференциальных уравнений // Вырожденные системы обыкновенных дифференциальных уравнений. — Новосибирск: Наука, 1982. — С. 37–66.

4. **Булатов М.В., Чистяков В.Ф.** Решение алгебро-дифференциальных систем методом наименьших квадратов // Тр. XI Междунар. Байкальской школы-семинара, Иркутск, Байкал, 5–12 июля 1998 г. — Т. 4. — С. 72–75.
5. **Горбунов В.К.** Метод нормальной сплайн-коллокации // ЖВМиМФ. — 1989. — Т. 29, № 2. — С. 212–224.
6. **Горбунов В.К., Петрищев В.В.** Развитие метода нормальной сплайн-коллокации для линейных дифференциальных уравнений // ЖВМиМФ. — 2003. — Т. 43, № 8. — С. 1161–1170.
7. **Gorbunov V.K., Lutoshkin I.V.** The parametrization method in optimal control problems and differential algebraic equations // J. Comput. Appl. Mathem. — 2006. — Vol. 185, iss. 2. — P. 377–390.
8. **Gorbunov V.K., Sviridov V.Yu.** The method of normal splines for linear DAEs on the number semi-axis // Appl. Numer. Math. — 2009. — Vol. 59, iss. 3–4. — P. 656–670.
9. **Булатов М.В., Горбунов В.К., Мартыненко Ю.В., Нгуен Дин Конг.** Вариационные подходы к численному решению дифференциально-алгебраических уравнений // Вычислительные технологии. — 2010. — Т. 15, № 5. — С. 3–13.
10. **Гантмахер Ф.Р.** Теория матриц. 3-е изд. — М.: Наука, 1966.
11. **Бояринцев Ю.Е.** Регулярные и сингулярные системы линейных обыкновенных дифференциальных уравнений. — Новосибирск: Наука, 1980.
12. **Чистяков В.Ф.** Алгебро-дифференциальные операторы с конечномерным ядром. — Новосибирск: Наука, 1996.
13. **Campbell S.L., Petzold L.R.** Canonical forms and solvable singular sys of differential equations // SIAM J. Alg. and Discrete Methods. — 1983. — № 4. — P. 517–521.
14. **Kunkel P., Mehrmann V.** Differential-Algebraic Equations: Analysis and Numerical Solution. — European Mathematical Society, 2006.
15. **Маслов В.П.** Операторные методы. — М.: Наука, 1973.
16. **Васильев Ф.П.** Методы решения экстремальных задач. — М.: Изд-во МГУ, 1974.
17. **Поляк Б.Т.** Итерационные методы решения некорректных вариационных задач // Вычислительные методы и программирование. — М.: Изд-во МГУ, 1969. — С. 95–108.
18. **Фихтенгольц Г.М.** Курс дифференциального и интегрального исчислений. Т. II. — М.: Физматгиз, 1962.
19. **Нестеров Ю.Е.** Введение в невыпуклую оптимизацию. — М.: МЦМНО, 2010.
20. **Бахвалов Н.С., Жидков Н.П., Кобельков Г.М.** Численные методы. — М.: Наука, 1987.
21. **Федоренко Р.П.** Введение в вычислительную физику. 2-е изд., испр. и доп. — Долгопрудный: Издательский дом “Интеллект”, 2008.

Поступила в редакцию 18 марта 2011 г.,
в окончательном варианте 26 сентября 2011 г.