УДК 519.71

СТАТИСТИЧЕСКАЯ ЛИНЕАРИЗАЦИЯ МНОГОМЕРНЫХ СТОХАСТИЧЕСКИХ СИСТЕМ ПО ИНФОРМАЦИОННОМУ КРИТЕРИЮ

К. Р. Чернышев,
канд. физ.-мат. наук
Институт проблем управления им. В. А. Трапезникова РАН

Предложена конструктивная процедура построения линейной входо-выходной модели, которая представляет собой статистический эквивалент некоторой нелинейной многомерной динамической стохастической системы с гауссовым входным процессом в виде белого шума. Ключевым моментом такой процедуры является использование в качестве критерия статистической линеаризации условия покомпонентного совпадения взаимной информации входного и выходного процессов системы и взаимной информации входного и выходного процессов модели. Данный подход позволяет получить явные соотношения, определяющие элементы весовых матриц линеаризованной модели.

Ключевые слова *— взаимная информация, входо-выходная модель, гауссова плотность распределения, информационный критерий, меры зависимости, многомерная система, статистическая линеаризация.*

Введение

Решение задачи идентификации систем всегда основано на применении тех или иных мер зависимости случайных величин (процессов), идет ли речь о представлении исследуемых систем в виде входо-выходного соотношения или в пространстве состояний. Наиболее часто в качестве такой меры выступают традиционные линейные ковариационные или корреляционные меры зависимости, использование которых непосредственно вытекает из самой постановки задачи идентификации на основе среднеквадратического критерия. Их основным достоинством является удобство использования, включая как возможность построения явных аналитических выражений для определения искомых характеристик, так и относительную простоту построения их оценок, в том числе и на основе наблюдения зависимых данных. Однако главным недостатком мер зависимости, основанных на линейной корреляции, является, как известно, возможность их обращения в нуль даже в случае существования детерминированной зависимости между парой исследуемых переменных [1–3].

Именно на преодоление этого недостатка направлено использование в задачах идентификации более сложных, нелинейных, мер зависимости, таких как дисперсионная функция, являющаяся аналогом известного в литературе дисперсионного отношения, максимальная корреляция, взаимная информация по Шеннону. При этом две последние меры, как известно, являются *состоятельными*, по терминологии А. Н. Колмогорова, мерами зависимости, т. е. обращающимися в нуль тогда и только тогда, когда случайные процессы (величины) в данной паре являются стохастически независимыми. В этом, в первую очередь, состоит привлекательность применения максимальной корреляции и взаимной информации в задачах идентификации, особенно в случае рассмотрения нелинейных систем.

К задачам нелинейной идентификации, решение которых существенно определяется характеристиками зависимости входных и выходных процессов системы, относится статистическая линеаризация входо-выходного отображения исследуемых систем. При этом известные подходы к статистической линеаризации основаны на применении либо обычных корреляционных функций, либо дисперсионных функций, что, в силу указанных выше причин, может приводить к построению моделей, выход которых тождественен нулю. В частности, возможность такой ситуации иллюстрируется в последнем разделе примером. Предлагаемый в настоящей работе подход направлен на исключение отмеченных недостатков, связанных с применением корреляционных и дисперсионных (основанных на корреляционном отношении) мер зависимости при идентификации систем на основе линеаризованных представлений их входо-выходных моделей. В его рамках рассматривается постановка задачи статистической линеаризации многомерных систем с дискретным временем по информационному критерию, обобщающая подход, представленный в работе [4] для одномерных систем.

Предварительные замечания

Применение состоятельных мер зависимости имеет свои особенности и ограничения. В этих рамках шенноновская взаимная информация выглядит предпочтительнее максимальной корреляционной функции, вычисление которой сопряжено с необходимостью использовать сложную итеративную процедуру определения первого собственного числа и пару первых собственных функций стохастического ядра $p_{yw}(y,w,\tau)\Big/\sqrt{p_w(w)p_y(y)}$ [1, 5], где $p_w(w)$, $p_y(y)$, $p_{yw}(y, w, \tau)$ представляют собой маргинальные и совместную плотности распределения случайных процессов $w(s)$ и $y(t)$ соответственно, $\tau = t - s$. К использованию взаимной информации приводит выбор в качестве критерия идентификации теоретико-информационного критерия. Примером такого подхода является работа [6], в которой постановка задачи идентификации ограничена рассмотрением класса *линейных гауссовых* систем и естественным образом приводит к использованию следующего соотношения для взаимной информации $I(Y, X)$ многомерного нормального распределения:

$$I(Y,X) = -\frac{1}{2}\ln\left(\frac{\det(Q_{\mathbf{ZZ}})}{\det(Q_{\mathbf{YY}})\det(Q_{\mathbf{XX}})}\right). \quad (1)$$

В формуле (1) приняты следующие обозначения: $\mathbf{Z}$ — нормально распределенный случайный вектор с ковариационной матрицей $Q_{\mathbf{ZZ}}$, $\dim\mathbf{Z} = n + m$, причем $\mathbf{Z} = (\mathbf{X}^T\ \mathbf{Y}^T)^T$, где $\dim\mathbf{X} = n$, $\dim\mathbf{Y} = m$; $Q_{\mathbf{XX}}$, $Q_{\mathbf{YY}}$ — ковариационные матрицы случайных векторов $\mathbf{X}$ и $\mathbf{Y}$ соответственно. При этом целью работы [6] является демонстрация эквивалентности ряда критериев идентификации и управления для *линейных гауссовых* систем. В то же время нельзя не отметить, что в этих рамках в принципе исчезает сам смысл обращения к подобному информационному критерию, поскольку в данном случае достаточно использовать обычный среднеквадратический критерий (как хорошо известно, в случае нормальности совместного распределения максимальная корреляция линейна и совпадает с обычной).

Постановка задачи

Пусть в некоторой многомерной (MIMO — multi input / multi output) нелинейной динамической стохастической системе $Y(t) = (y_1(t), ..., y_n(t))^T$ — n-мерный выходной случайный процесс системы, предполагающийся стационарным и эргодическим; $W(s) = (w_1(s), ..., w_m(s))^T$ — m-мерный входной случайный процесс системы, предполагающийся в данной постановке задачи белым гауссовым шумом с известной ковариационной матрицей C_W, а зависимость компонент входных и выходных процессов системы характеризуется (конечно, неизвестными исследователю) плотностями распределения

$$p_{y_i,w_j}(y,w,\tau),\ i = 1, ..., n,\quad j = 1, ..., m,\ \tau = 1, 2, ... \quad (2)$$

Ради простоты построения, но без потери общности, компоненты данных процессов $Y(t)$ и $W(s)$ предполагаются имеющими нулевые средние и единичные дисперсии

$$\mathbf{M}\{y_i(t)\} = \mathbf{M}\{w_j(s)\} = 0,\ \mathbf{D}\{y_i(t)\} = \mathbf{D}\{w_j(s)\} = 1,\quad i = 1, ..., n,\ j = 1, ..., m, \quad (3)$$

где $\mathbf{M}\{\cdot\}$, $\mathbf{D}\{\cdot\}$ — символы математического ожидания и дисперсии соответственно. В сделанных предположениях

$$C_W = \begin{pmatrix} 1 & c_{12} & \cdots & c_{1m} \\ c_{12} & \ddots & & \vdots \\ \vdots & & \ddots & c_{(m-1)m} \\ c_{1m} & \cdots & c_{(m-1)m} & 1 \end{pmatrix}. \quad (4)$$

Также процессы $Y(t)$ и $W(s)$ предполагаются стационарно связанными в строгом смысле.

Линейная *входо-выходная* модель системы, характеризуемой плотностями распределения (2), ищется в виде

$$\hat{Y}(t;\mathbf{G}) = \sum_{k=1}^{\infty} G(k)W(t-k),\ t = 1, 2, ..., \quad (5)$$

где $\hat{Y}(t;\mathbf{G}) = (\hat{y}_1(t;\mathbf{G}),\ldots,\hat{y}_n(t;\mathbf{G}))^T$ — выходной процесс модели, $\mathbf{G} = \{G(k), k\in[1, \infty)\}$, $G(k)$, $k = 1, 2, ...$ — матричнозначные (размерностью $n \times m$) коэффициенты весовой функции линеаризованной модели, подлежащие идентификации в соответствии с условием совпадения взаимной информации i-й компоненты выходного процесса $y_i(t)$ и j-й компоненты входного процесса $w_j(s)$ системы, характеризуемой плотностями распределения (2), и взаимной информации i-й компоненты выходного процесса $\hat{y}_i(t,\mathbf{G})$ и j-й компоненты входного процесса $w_j(s)$ модели (5) для всех $i = 1, ..., n$, $j = 1, ...,$

m. Аналитически данный критерий имеет следующий вид:

$$I_{y_i w_j}(\tau)=I_{\hat{y}_i(\mathbf{G})w_j}(\tau),\ \tau=1,\,2,\,\ldots. \qquad (6)$$

Безусловно, с точки зрения задачи статистической линеаризации условие (6) необходимо дополнить условием совпадения математических ожиданий выходных процессов системы и модели

$$\mathbf{M}\{y_i(t)\}=\mathbf{M}\{\hat{y}_i(t;\mathbf{G})\}=0,\ i=1,\,\ldots,\,n. \qquad (7)$$

Очевидно, что в рамках данной постановки задачи условие (7) выполняется автоматически.

Далее, следуя условию нормировки (3), на компоненты выходного процесса модели (5) налагается условие единичности дисперсии

$$\mathbf{D}\{\hat{y}_i(t;\mathbf{G})\}=1,\ i=1,\,\ldots,\,n, \qquad (8)$$

и, как следствие, строки матричнозначных коэффициентов модели (5) должны удовлетворять условию

$$\sum_{k=1}^{\infty}\vec{g}_i(k)C_W\vec{g}_i^{\,T}(k)=1,\ i=1,\,\ldots,\,n, \qquad (9)$$

где $\vec{g}_i(k)=(g_{i1}(k),\ldots,g_{im}(k))$ — i-я строка матрицы $G(k)$ из (5).

Соотношение (9) очевидным образом определяется цепочкой

$$1=\mathbf{D}\{\hat{y}_i(t;\mathbf{G})\}=\mathbf{D}\left\{\sum_{k=1}^{\infty}\vec{g}_i(k)W(t-k)\right\}=$$

$$=\sum_{k=1}^{\infty}\vec{g}_i(k)\mathbf{M}\{W(t-k)W^T(t-k)\}\vec{g}_i^{\,T}(k)+$$

$$+\sum_{p\neq q}\vec{g}_i(p)\mathbf{M}\{W(t-p)W^T(t-q)\}\vec{g}_i^{\,T}(q)=\sum_{k=1}^{\infty}\vec{g}_i(k)C_W\vec{g}_i^{\,T}(k)$$

в силу описания модели (5) и условий нормировки (3), (4), (8).

Выражения (6) и (7) представляют собой, таким образом, критерий статистической линеаризации системы, характеризуемой плотностями распределения (2). В терминах плотностей распределения условие (6) записывается в виде

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\ln\frac{p_{y_i,w_j}(y,w,\tau)}{p_{y_i}(y)p_{w_j}(w)}\right)p_{y_i,w_j}(y,w,\tau)\mathrm{d}y\mathrm{d}w=$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\ln\frac{p_{\hat{y}_i(\mathbf{G}),w}(\hat{y}(\mathbf{G}),w,\tau)}{p_{\hat{y}_i(\mathbf{G})}(\hat{y}(\mathbf{G}))p_{w_j}(w)}\right)\times$$

$$\times\, p_{\hat{y}_i(\mathbf{G}),w_j}(\hat{y}(\mathbf{G}),w,\tau)\mathrm{d}\hat{y}(\mathbf{G})\mathrm{d}w,$$

$$i=1,\,\ldots,\,n,\ j=1,\,\ldots,\,m,\ \tau=1,\,2,\,\ldots,$$

где $p_{y_i,w_j}(y,w,\tau)$, $p_{\hat{y}_i(\mathbf{G}),w_j}(\hat{y}(G),w,\tau)$ — соответственно совместные плотности распределения i-й компоненты выходного и j-й компоненты входного процессов системы, характеризуемой плотностями распределения (2), и i-й компоненты выходного и j-й компоненты входного процессов модели (5); $p_{y_i}(y)$, $p_{\hat{y}_i(\mathbf{G})}(\hat{y}(\mathbf{G}))$ и $p_{w_j}(w)$ — соответственно маргинальные плотности распределения i-х компонент выходных процессов $Y(t)$ системы, характеризуемой плотностями распределения (2), модели $\hat{Y}(t;\mathbf{G})$ (5) и j-й компоненты входного процесса системы, характеризуемой плотностями распределения (2), равно как и модели (5), $W(s)$, $\tau=t-s$.

Метод решения

Пусть

$$v_i\langle-\tau;t\rangle=\sum_{j=1}^{\tau-1}\vec{g}_i(j)W(t-j)+\sum_{j=\tau+1}^{\infty}\vec{g}_i(j)W(t-j),$$

$$\tau=1,\,2,\,\ldots$$

— последовательность случайных величин, которые очевидно являются гауссовыми с нулевыми средними и дисперсиями, имеющими в силу (9) вид

$$\mathbf{D}\{v_i\langle-\tau;t\rangle\}=\sum_{j=1}^{\tau-1}\vec{g}_i(j)C_W\vec{g}_i^{\,T}(j)+$$

$$+\sum_{j=\tau+1}^{\infty}\vec{g}_i(j)C_W\vec{g}_i^{\,T}(j)=1-\vec{g}_i(\tau)C_W\vec{g}_i^{\,T}(\tau),$$

$$\tau=1,\,2,\,\ldots\,.$$

Тогда, в рамках введенных обозначений, $(m+1)$-мерный случайный вектор

$$V_i(t,\tau)=\left(v_i\langle-\tau;t\rangle,w_1(t-\tau),\ldots,w_m(t-\tau)\right)^T$$

является гауссовым с ковариационной матрицей

$$C_{V_i(t,\tau)}=$$

$$=\begin{pmatrix}1-\vec{g}_i(\tau)C_W\vec{g}_i^{\,T}(\tau) & 0 & \cdots & \cdots & 0\\ 0 & 1 & c_{12} & \cdots & c_{1m}\\ \vdots & c_{12} & \ddots & & \vdots\\ \vdots & \vdots & & \ddots & c_{(m-1)m}\\ 0 & c_{1m} & \cdots & c_{(m-1)m} & 1\end{pmatrix},$$

а для двумерного случайного вектора $(\hat{y}_i(t;\mathbf{G})$ $w_j(t-\tau))^T$ можно записать следующее соотношение:

$$\begin{pmatrix}\hat{y}_i(t;\mathbf{G})\\ w_j(t-\tau)\end{pmatrix}=A_{ij}(\tau)V_i(t,\tau),$$

где $(2\times(m+1))$-мерная матрица $A_{ij}(\tau)$ имеет вид

$$A_{ij}(\tau)=\begin{pmatrix}1 & g_{i1}(\tau) & \cdots & g_{ij}(\tau) & g_{i(j+1)}(\tau) & \cdots & g_{im}(\tau)\\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0\end{pmatrix},$$

и, как и всюду, $g_{i1}(\tau)$, ..., $g_{im}(\tau)$ — элементы i-й строки матрицы $G(\tau)$ из (5).

Следовательно, случайный вектор $(\hat{y}_i(t;\mathbf{G})\ w_j(t-\tau))^T$ является гауссовым с ковариационной матрицей $C_{(\hat{y}_i w_i)}(\tau)$, определяемой соотношением

$$C_{(\hat{y}_i w_i)}(\tau) = A_{ij}(\tau) C_{V_i(t,\tau)} A_{ij}^T(\tau).$$

Вычисление произведения трех матриц в правой части дает

$$C_{(\hat{y}_i w_i)}(\tau) = \begin{pmatrix} 1 & \gamma_{ij}(\tau) \\ \gamma_{ij}(\tau) & 1 \end{pmatrix},$$

где $\gamma_{ij}(\tau)$ — j-я компонента вектор-столбца $C_W \vec{g}_i^{\,T}(\tau)$.

Таким образом, в силу формулы (1) следует, что взаимная информация $I_{\hat{y}_i(\mathbf{G})w_j}(\tau)$ выходного и входного процессов модели (5) имеет вид

$$I_{\hat{y}_i(\mathbf{G})w_j}(\tau) = -\frac{1}{2}\ln(1-\gamma_{ij}^2(\tau)),\ \tau = 1, 2, \ldots.$$

Тогда из условия (6) следуют искомые выражения для строк $\vec{g}_i(\tau)$ матричнозначных весовых коэффициентов $G(\tau)$, $\tau = 1, 2, \ldots$ модели (5):

$$\vec{g}_i^{\,T}(\tau) = C_W^{-1}\ \mathbf{I}_{y_i W}(\tau),\ \tau = 1, 2, \ldots, \qquad (10)$$

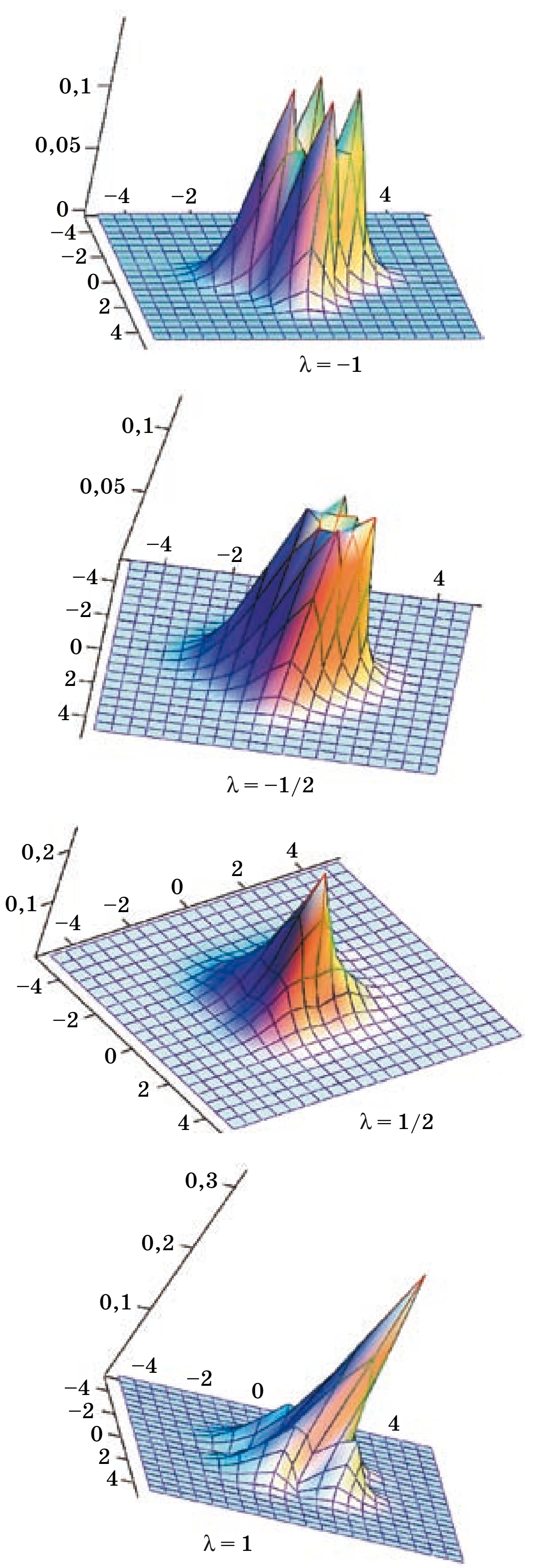
где

$$\mathbf{I}_{y_i W}(\tau) = \begin{pmatrix} \operatorname{sign}(\mathbf{m}_{y_i|w_1}(\tau)) \times \sqrt{1-\exp(-2I_{y_i w_1}(\tau))} \\ \vdots \\ \operatorname{sign}(\mathbf{m}_{y_i|w_j}(\tau)) \times \sqrt{1-\exp(-2I_{y_i w_j}(\tau))} \\ \vdots \\ \operatorname{sign}(\mathbf{m}_{y_i|w_m}(\tau)) \times \sqrt{1-\exp(-2I_{y_i w_m}(\tau))} \end{pmatrix}.$$

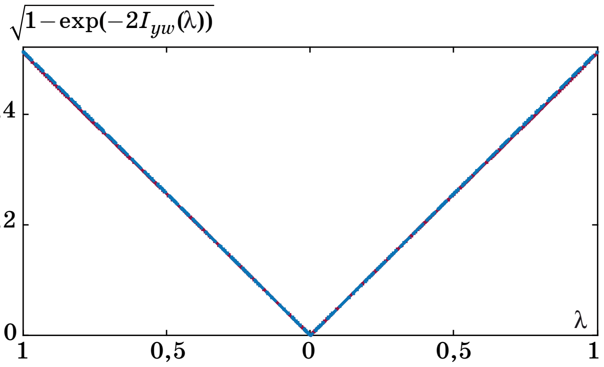
В формуле (10) $\mathbf{m}_{y_i|w_j}(\tau)$ — регрессия $y_i(t)$ на $w_j(t-\tau)$; $\operatorname{sign}(x) = 1$ при $x \geq 0$, $\operatorname{sign}(x) = -1$ при $x < 0$ — знак соответствующей функции регрессии соответствует «взаимной ориентации» входного и выходного процессов; подкоренное выражение всегда неотрицательно, поскольку взаимная информация принимает значения на положительной полуоси $[0, +\infty)$.

Таким образом, обращение в нуль весовых коэффициентов линеаризованной модели (5) системы (2) эквивалентно обращению в нуль взаимной информации выходного и входного процессов системы (2). В свою очередь, последнее возможно лишь тогда, когда данные процессы стохастически независимы. При этом, как отмечено выше, существуют примеры, когда традиционные меры зависимости обращаются в нуль при наличии стохастической зависимости между переменными.

Так, можно рассмотреть следующую плотность распределения, принадлежащую классу распределений О. В. Сарманова [7, 8], которая имеет вид



■ *Рис. 1. Форма плотности (11) при различных значениях параметра λ*



■ ***Рис. 2.*** *Близость значений выражения (12) и $S_{yw}(\lambda)$ при различных значениях параметра λ в плотности (11)*

$$p_\lambda(y,w)=\frac{e^{-\frac{w^2+y^2}{2}}}{2\pi}\times$$
$$\times\left(1+\lambda\left(2e^{-\frac{3}{2}w^2}-1\right)\left(2e^{-\frac{3}{2}y^2}-1\right)\right),$$
$$-1\le\lambda\le1. \qquad (11)$$

Ее маргинальные плотности являются стандартными гауссовыми. Для плотности (11) коэффициент корреляции и дисперсионное отношение равны нулю, а максимальный коэффициент корреляции имеет вид

$$S_{yw}(\lambda)=\left(\frac{4}{\sqrt{7}}-1\right)|\lambda|.$$

Значение параметра λ оказывает существенное влияние на форму плотности (11) (рис. 1).

При этом величина $\sqrt{1-\exp(-2I_{yw}(\lambda))}$ из выражения (10), соответствующая плотности $p_\lambda(y, w)$ в (11), зависит от модуля параметра λ строго монотонно (рис. 2) и обращается в нуль только при $\lambda = 0$, что эквивалентно независимости случайных величин.

На рис. 2 показана зависимость значений

$$\sqrt{1-\exp(-2I_{yw}(\lambda))} \qquad (12)$$

от параметра λ в плотности (11) в сравнении с соответствующими значениями максимального коэффициента корреляции $S_{yw}(\lambda)$, наглядно демонстрирующая практически полное их совпадение.

Таким образом, если, например, стохастическая зависимость (2) между компонентами выходного процесса, $y_i(t)$, и входного процесса, $w_j(s)$, некоторой нелинейной системы определяется плотностью распределения (конечно, предполагаемой неизвестной исследователю) вида (11) с параметром $\lambda = \lambda_{ij}(\tau)$, $\tau = t - s$, то применение как традиционных корреляционных, так и дисперсионных методов статистической линеаризации привело бы, при построении модели (5), к представлению компонент выходного процесса системы как тождественного нуля, что исключается при использовании данного теоретико-информационного подхода.

Заключение

В настоящей работе для многомерных нелинейных систем с белым гауссовым векторнозначным входным шумом рассмотрена задача определения статистически эквивалентных линейных входо-выходных моделей из условия покомпонентного совпадения взаимной информации входного и выходного процессов системы и входного и выходного процессов модели. Получаемые в конечном итоге уравнения для элементов весовых матриц линеаризованной модели определяют их как функции взаимной информации входного и выходного процессов системы, обращающиеся в нуль только при обращении в нуль взаимной информации, т. е. при стохастической независимости данных компонент входного и выходного процессов системы.

Литература

1. **Сарманов О. В.** Максимальный коэффициент корреляции (несимметричный случай) // Докл. АН СССР. 1958. Т. 121. № 1. С. 52–55.
2. **Rényi A.** On measures of dependence // Acta Math. Hung. 1959. Vol. 10. N 3–4. P. 441–451.
3. **Дисперсионная** идентификация / Под ред. Н. С. Райбмана. — М.: Наука, 1981. — 320 с.
4. **Чернышев К. Р.** Информационные меры зависимости в статистической линеаризации // Автоматика и телемеханика. 2002. № 9. С. 74–84.
5. **Сарманов О. В., Захаров Е. К.** Меры зависимости между случайными величинами и спектры стохастических ядер и матриц // Математический сборник. 1960. Т. 52(94). № 4. С. 953–990.
6. **Stoorvogel A. A., van Schuppen J. H.** System identification with information theoretic criteria // Identification, adaptation, learning / Ed. by S. Bittanti and G. Picci. — Berlin: Springer-Verlag, 1996. P. 289–338.
7. **Сарманов О. В.** Замечания о некоррелированных гауссовских зависимых случайных величинах // Теория вероятностей и ее применения. 1967. Т. 12. № 1. С. 141–143.
8. **Kotz S., Balakrishnan N., Johnson N. L.** Continuous Multivariate Distributions. Vol. 1. Models and Applications. Second ed. — N. Y.: Wiley, 2000. — 752 p.