Один подход к статистической линеаризации, основанный на квадратической взаимной информации Йенсена-Цаллиса

К. Р. Чернышев
Лаборатория идентификации систем управления
Институт проблем управления им. В.А. Трапезникова РАН
Москва, Россия
myau@ipu.ru

Аннотация. **Представлен подход к статистической линеаризации входо-выходного отображения нелинейных дискретно-временных стохастических систем с белошумным гауссовским входным процессом. Подход основан на применении состоятельной меры зависимости, формируемой использованием энтропии Цаллиса. В его рамках критерий статистической линеаризации представляет собой условие совпадения математических ожиданий выходных процессов системы и модели и условие совпадения данной меры зависимости входного и выходного процессов системы и этой же меры зависимости входного и выходного процессов модели. В результате строятся явные аналитические выражения для коэффициентов весовой функции линеаризованной модели. Рассмотрению предшествует анализ применения состоятельных мер зависимости в задачах идентификации систем.**

Ключевые слова: состоятельные меры зависимости; дивергенция Йенсена-Цаллиса; входо-выходная модель; статистическая линеаризация; идентификация систем; энтропия Цаллиса

I. Теоретико-информационные меры зависимости как математическая основа в идентификации стохастических систем

Решение задачи идентификации систем в рамках стохастической постановки всегда основано на применении мер зависимости случайных величин (СВ), как при представлении исследуемых систем на основе входо-выходного описания, так и при подходах, основанных на методах пространства состояний. Наиболее часто в рамках идентификационных подходов используются традиционные корреляционные/ковариационные меры зависимости. Однако главный недостаток мер зависимости, основанных на линейной корреляции, состоит в возможности их обращения в ноль даже в условиях наличия детерминированной связи между исследуемыми переменными [1, 2]. Именно на преодоление этого недостатка было направлено использование более сложных, состоятельных, мер зависимости в задачах идентификации (в соответствии с терминологией А.Н. Колмогорова, мера зависимости между двумя случайными величинами называется состоятельной, если она обращается в ноль тогда и только тогда, когда данные СВ стохастически независимы).

В 1959 г. А. Реньи сформулировал аксиомы [1], которые в конечном итоге были общепризнанны как наиболее естественные условия для характеризации меры зависимости $\mu(X,Y)$ между двумя СВ X и Y:

A) $\mu(X,Y)$ *определена для любой пары СВ X и Y, если ни одна из них не является с вероятностью 1 константой.*

B) $\mu(X,Y)=\mu(Y,X)$.

C) $0 \le \mu(X,Y) \le 1$.

D) $\mu(X,Y)=0$ *тогда и только тогда, когда X и Y стохастически независимы.*

E) $\mu(X,Y)=1$ *если существует строгая зависимость между X и Y, то есть или* $Y=\varphi(X)$*, или* $X=\psi(Y)$*, где* φ *и* ψ *– некоторые борелевские функции.*

F) Если φ *и* ψ *– некоторые борелевские взаимно-однозначные функции, то* $\mu(\varphi(X),\psi(Y))=\mu(X,Y)$.

G) Если совестное распределение вероятностей X и Y нормально, то $\mu(X,Y)=|r(X,Y)|$*, где* $r(X,Y)$ *– обычный коэффициент корреляции между X и Y.*

Класс состоятельных по Колмогорову мер зависимости (то есть удовлетворяющих хотя бы аксиоме Реньи D) достаточно широк, в то время как класс мер зависимости, удовлетворяющих всем аксиомам Реньи – класс мер зависимости, состоятельных по Реньи, – весьма узок.

Широкий класс состоятельных по Колмогорову мер зависимости может быть построен на основе соответствующих мер сравнения плотностей (непрерывных) многомерных вероятностных распределений, например, $g_1(z)$ и $g_2(z)$, $z \in R^{\nu}$, известных как меры дивергенции, среди которых, по-видимому, наиболее известной и используемой является дивергенция Кульбака-Лайблера. Меры дивергенции можно рассматривать как критерий качества в рамках различных теоретических и прикладных задач. В частности, дивергенция Кульбака-Лайблера приводит к выражению для взаимной информации (ВИ) Шеннона. ВИ Шеннона (1), также часто именуемая как относительная энтропия Шеннона, основана на соответствующих энтропиях Шеннона. В то же время, наряду с определением энтропии по Шеннону известен целый ряд других определений энтропии (многомерной) СВ. Так для некоторой ν-мерной СВ Z с многомерной плотностью распределения $g(z)$ энтропия Цаллиса порядка α определяется следующим образом [3]:

$$T_\alpha(Z) = \frac{1}{\alpha - 1}\left(1 - \int\limits_{R^\nu} (g(z))^\alpha dz\right), \; \alpha > 0, \alpha \neq 1. \qquad (1)$$

При этом по мере того, как α стремится к 1, $T_\alpha(Z)$ стремится к энтропии Шеннона, и, таким образом, энтропия Шеннона может рассматриваться как энтропия Цаллиса «порядка 1».

Также как и энтропия Шеннона, энтропия Цаллиса $T_\alpha(Z)$ ν-мерной СВ Z с многомерной плотностью распределения $g(z)$ может рассматриваться и непосредственно относительно этой плотности распределения $g(z)$, и в этом случае она будет обозначаться как $T_\alpha(g)$. Таким образом, $T_\alpha(Z)$ и $T_\alpha(g)$ следует рассматривать как математически эквивалентные обозначения, оба определяемые формулой (1). Необходимость такого замечания будет объяснена исходя из нижеприводимого рассмотрения, связанного с соответствующей мерой дивергенции вероятностных распределений, которая основана на энтропии Цаллиса (1). А именно, используя свойство выпуклости энтропии Цаллиса, в литературе, например [3], известна дивергенция Йенсена-Цаллиса $D_\alpha^{JT}(g_1 \| g_2)$ порядка α двух плотностей распределения вероятностей $g_1(z)$ и $g_2(z)$ следующим образом:

$$D_\alpha^{JT}(g_1 \| g_2) = T_\alpha\left(\frac{g_1 + g_2}{2}\right) - \frac{T_\alpha(g_1) + T_\alpha(g_2)}{2}, \quad \alpha > 0, \alpha \neq 1. \qquad (2)$$

При этом $D_\alpha^{JT}(g_1 \| g_2)$ неотрицательна и обращается в ноль тогда и только тогда, когда $g_1(z) = g_2(z)$. Соответственно, при $\nu = 2$ и Z, образованной двумя СВ X_1 и X_2, когда одна из плотностей распределения в $D_\alpha^{JT}(g_1 \| g_2)$ в (2), а именно $g_1(z)$, представляет собой совместную плотность распределения этих СВ $p_{12}(x_1, x_2)$, а вторая плотность, $g_2(z)$, является произведением маргинальных плотностей распределения X_1: $p_1(x_1)$, и X_2: $p_2(x_2)$, соответствующая дивергенция $D_\alpha^{JT}(p_{12} \| p_1 p_2)$ Йенсена-Цаллиса порядка α приводит к состоятельной в смысле Колмогорова теоретико-информационной мере зависимости $I_\alpha^{JT}\{X_1, X_2\}$ двух случайных величин X_1 и X_2:

$$D_\alpha^{JT}(p_{12} \| p_1 p_2) = I_\alpha^{JT}\{X_1, X_2\} = \frac{1}{\alpha - 1}\left(1 - \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\left(\frac{p_{12}(x_1, x_2) + p_1(x_1) p_2(x_2)}{2}\right)^\alpha dx_1 dx_2\right) - \\ - \frac{1}{\alpha - 1}\left(1 - \frac{\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}(p_{12}(x_1, x_2))^\alpha dx_1 dx_2 + \int\limits_{-\infty}^{\infty}(p_1(x_1))^\alpha dx_1 \int\limits_{-\infty}^{\infty}(p_2(x_2))^\alpha dx_2}{2}\right). \qquad (3)$$

Выражение (3) определяет одновременно и теоретико-информационный критерий, который может рассматриваться как основа для построения теоретико-информационных методов идентификации систем.

С вычислительной точки зрения, особенно при необходимости построения оценок на основе выборочных данных, энтропию Цаллиса следует рассматривать как существенно более удобную для применения, чем энтропию Шеннона, поскольку энтропия Шеннона включает «интеграл от логарифма», а энтропия Цаллиса вообще не содержит логарифма, что в вычислительном плане существенно проще.

При этом выбор конкретного значения порядка α важен, поскольку, чем больше значение порядка, тем более сложной становится вычислительная процедура. Такого рода рассмотрения вычислительных и аналитических аспектов величины порядка α энтропии Цаллиса приводят к своего рода «компромиссу», в рамках которого значение порядка $\alpha = 2$ выглядит наиболее приемлемым. Значение $\alpha = 2$ в формулах (2) и (3) соответствует квадратической дивергенции и квадратической взаимной информации (КВИ) Йенсена-Цаллиса соответственно. Так, при $\alpha = 2$ из формул (2) и (3) непосредственно следует

$$I_2^{JT}\{X_1, X_2\} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}(p_{12}(x_1, x_2) - p_1(x_1) p_2(x_2))^2 dx_1 dx_2. \qquad (4)$$

В случае стационарных и взаимно-стационарный в узком смысле случайных процессов $y(t)$ и $x(s)$, $I_2^{JT}\{y(t), x(s)\}$, $\tau = t - s$ в формуле (4) становится соответствующей функцией времени:

$$I_2^{JT}\{y(t), x(s)\} = \\ = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}(p_{yx}(y, x, \tau) - p_y(y) p_x(x))^2 dy dx, \quad \tau = t - s. \qquad (5)$$

В (5), $p_{yx}(y,x,\tau)$, $p_y(y)$, $p_x(x)$ – совместная и маргинальные плотности распределения случайных процессов $y(t)$ и $x(s)$, $\tau = t - s$ соответственно.

Задачи идентификации, решение которых существенно определяется характеристиками зависимости входного и выходного процессов исследуемой системы, включают статистическую линеаризацию входо-выходного отображения. При этом известные подходы к статистической линеаризации основаны на применении либо традиционных корреляционных функций, либо дисперсионных функций [2] (другими словами, основанных на корреляционном отношении), что, по причинам, отмеченным выше, может приводить к построению моделей, чей выход будет тождественным нулем. Подход, предложенный в настоящей работе, ориентирован на исключение отмеченных недостатков, связанных с применением корреляционных и дисперсионных мер зависимости при идентификации систем на основе линеаризованных представлений их входо-выходных моделей.

В этих рамках именно функция $I_2^{JT}\{y(t), x(s)\}$ (5) будет использоваться в конечном итоге для построения критерия статистической линеаризации.

II. Статистическая линеаризация: постановка задачи и решение

Пусть в некоторой нелинейной динамической системе $y(t)$ – выходной процесс, рассматриваемый как стационарный в узком смысле эргодический случайный процесс; $x(s)$ – входной процесс, рассматриваемый как белошумный гауссовский случайный процесс. При этом зависимость входного и выходного процессов системы характеризуется плотностью распределения вероятностей (конечно, не известной исследователю)

$$p_{yx}(y,x,\tau),\ \tau = 1, 2, \ldots . \qquad (6)$$

Процессы $y(t)$ и $x(s)$ также предполагаются взаимностационарными в узком смысле (формула (6)). Ради простоты, но без ограничения общности, процессы $y(t)$ и $x(s)$ предполагаются имеющими нулевое среднее и единичную дисперсию, то есть

$$\mathbf{E}\{y(t)\} = \mathbf{E}\{x(s)\} = 0,\ \mathbf{var}\{y(t)\} = \mathbf{var}\{x(s)\} = 1, \qquad (7)$$

где $\mathbf{var}\{\cdot\}$ – символ дисперсии.

Модель системы, описываемой плотностями распределения (6) при условии (7), ищется в виде

$$\hat{y}(t;W) = \sum_{\tau=1}^{\infty} w(\tau)x(t-\tau),\ t = 1,2,\ldots, \qquad (8)$$

где $\hat{y}(t;W)$ – выходной процесс модели, $W = \{w(\tau), \tau \in [1,\infty)\}$, $w(\tau), \tau = 1,2,\ldots$ – коэффициенты весовой функции линеаризованной модели, подлежащие идентификации в соответствии с критерием статистической линеаризации. Такой критерий представляет собой условие совпадения математических ожиданий выходного процесса системы, описываемой плотностями (6), и выходного процесса модели (8), и условие совпадения КВИ Йенсена-Цаллиса (5) выходного и входного процессов системы, описываемой плотностями (6), и выходного и входного процессов модели (8), или, математически,

$$\mathbf{E}\{y(t)\} = \mathbf{E}\{\hat{y}(t;W)\} = 0, \qquad (9)$$

$$I_2^{JT}\{y(t), x(s)\} = I_2^{JT}\{\hat{y}(t;W), x(s)\},\ t - s = \tau = 1,2,\ldots . \qquad (10)$$

Далее, в соответствии с условием нормализации (7), модель (8) ограничена условием $\mathbf{var}\{\hat{y}(t;G)\} = 1$, и, соответственно, весовые коэффициенты модели (8) удовлетворяют условию $\sum_{\tau=1}^{\infty} w^2(\tau) = 1$

Таким образом, выражения (9) и (10) представляют собой критерий статистической линеаризации системы, описываемой плотностями распределения (6).

Таким образом, с учетом введенных обозначений и выражения для модели (8), принимая во внимание свойства линейного преобразования гауссовского вектора и опуская промежуточные выкладки, можно в конечном итоге получить следующее уравнение, чье решение определяет весовые коэффициенты $w(\tau)$, $\tau = 1,2,\ldots$ модели (8):

$$\begin{aligned}&\frac{1}{4\pi\sqrt{1-(w(\tau))^2}} - \frac{1}{\pi\sqrt{4-(w(\tau))^2}} + \frac{1}{4\pi} = \\ &= I_2^{JT}\{y(t), x(s)\},\quad t - s = \tau = 1,2,\ldots .\end{aligned} \qquad (11)$$

Уравнение (11) может быть легко решено численно, и его (положительное) решение определяет значение модуля весовых коэффициентов $w(\tau)$, $\tau = 1,2,\ldots$ модели (8) через соответствующее значение квадратической взаимной информации Йенсена-Цаллиса $I_2^{JT}\{y(t), x(s)\}$, $t - s = \tau = 1,2,\ldots$ (5). Для «раскрытия» модуля решения уравнения (11) следует использовать знак регрессии выходного процесса системы на ее выходной процесс, то есть

$$\mathrm{sign}[reg_{yx}(\tau)] = \begin{cases} 1, & reg_{yx}(\tau) \geq 0 \\ -1, & reg_{yx}(\tau) < 0 \end{cases},$$

где $reg_{yx}(\tau) = \mathbf{E}\left\{ y(t) / x(t-\tau) \right\}$, и $\mathbf{E}\left\{ \cdot / \cdot \right\}$ – символ условного математического ожидания. Таким образом, окончательно:

$$w(\tau) = \mathrm{sign}[reg_{yx}(\tau)]\phi\left(I_2^{JT}\{y(t), x(s)\}\right),\ \tau = 1,2,\ldots, \qquad (12)$$

где функция $\phi\left(I_2^{JT}\{y(t),x(s)\}\right)$ определяет обратное выражение (от $I_2^{JT}\{y(t),x(s)\}$ к $w(\tau)$) к уравнению (11). Таким образом, выражение (12) определяет коэффициенты весовой функции линеаризованной модели (8).

Таким образом, мера зависимости $\phi\left(I_2^{JT}\{y(t),x(s)\}\right)$, определяемая соотношением (5) и формулой (11), удовлетворяет всем аксиомам Реньи [1] для мер зависимости случайных величин за исключением только аксиомы F, однако ее вычисление существенно проще по сравнению с вычислением максимального коэффициента корреляции [1, 5, 6]. Меры зависимости, которые удовлетворяют аксиомам Реньи, за исключением, возможно, аксиомы F, естественно называть мерами зависимости, состоятельными по Реньи. Класс таких мер существенно уже, чем класс мер зависимости, состоятельных по Колмогорову. Поведение функции $\phi\left(I_2^{JT}\{y(t),x(s)\}\right)$ представлено на рис. 1.

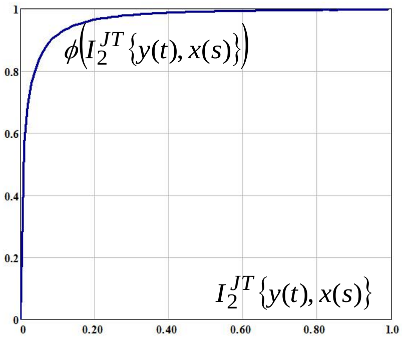

Рис. 1.

III. Явная аналитическая аппроксимация для $\phi\left(I_2^{JT}\{y(t),x(s)\}\right)$

Выражение для аппроксимации точного решения $\phi\left(I_2^{JT}\{y(t),x(s)\}\right)$ уравнения (11), другими словами, обратного выражения (от $I_2^{JT}\{y(t),x(s)\}$ к $w(\tau)$) к уравнению (11), в аналитическом виде, может быть получено за счет подходящей аппроксимации левой части (11). Такая аппроксимация основана на замене члена

$$\frac{1}{\pi\sqrt{4-(w(\tau))^2}}$$

в (11) некоторой константой C_w, причем эта константа C_w должна удовлетворять следующему основному естественному условию

$$\frac{1}{4\pi\sqrt{1-(w(\tau))^2}}-C_w+\frac{1}{4\pi}=I_2^{JT}\{y(t),x(s)\}\equiv 0, \tag{13}$$
$$t-s=\tau=1,2,\ldots. \quad \text{as } w(\tau)=0.$$

Условие (13) непосредственно влечет $C_w=1/(2\pi)$, что в итоге дает

$$\hat{\phi}\left(I_2^{JT}\{y(t),x(s)\}\right)=\sqrt{1-\frac{1}{\left(4\pi I_2^{JT}\{y(t),x(s)\}+1\right)^2}}, \tag{14}$$
$$t-s=\tau=1,2,\ldots,$$

и соответствующе графическое представление поведения $\hat{\phi}\left(I_2^{JT}\{y(t),x(s)\}\right)$ – оценки $\phi\left(I_2^{JT}\{y(t),x(s)\}\right)$– представлено на рис. 2. Такое сравнение позволяет рассматривать выражения (13) и (14) как допустимый метод для построения явной аналитической формулы, которая могла бы адекватно аппроксимировать величину $\phi\left(I_2^{JT}\{y(t),x(s)\}\right)$.

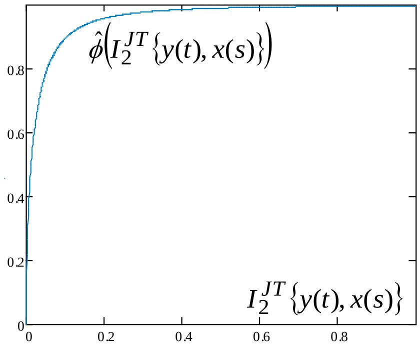

Рис. 2.

IV. Заключение

Рассмотрен подход к статистической линеаризации входо-выходных отображений нелинейных дискретно-временных стохастических динамических систем с белошумным входным процессом. Такой подход основан на применении КВИ Йенсена-Цаллиса (5) при построении критерия статистической линеаризации. В рамках данного подхода критерий статистической линеаризации представляет собой условие совпадения математических ожиданий выходного процесса системы и выходного процесса модели, и условие совпадения КВИ Йенсена-Цаллиса выходного и входного процессов системы и КВИ Йенсена-Цаллиса выходного и входного процессов модели. Получены соотношения (12) для определения коэффициентов весовой функции линеаризованной модели. При этом полученные выражения основаны на КВИ Йенсена-Цаллиса и определяют меру стохастической зависимости случайных величин, являющейся состоятельной в смысле Реньи.

Список литературы

[1] Rényi A. “On measures of dependence” // *Acta Math. Hung.*, 1959, vol. 10, no. 3-4, pp. 441-451.

[2] Rajbman N.S. “Extensions to nonlinear and minimax approaches” // *Trends and Progress in System Identification*, ed. P. Eykhoff, Pergamon Press, Oxford, 1981, pp. 185-237.

[3] Tsallis C. “Possible generalization of Boltzmann-Gibbs statistics” // *Journal of Statistical Physics*, 1988, vol. 52, no. 1, pp. 479–487.

[4] Anguloa J.C., Antolín J., López-Rosa S., and R.O. Esquivel. “Jensen–Tsallis divergence and atomic dissimilarity for ionized systems in conjugated spaces” // *Physica A*, 2011, vol. 390, pp. 769-780.

[5] Sarmanov O.V. “Maximum correlation coefficient (symmetric case)” // *Select. Transl. Math. Stat. Probab*. 1963, vol. 4, pp. 271-275.

[6] Sarmanov O.V. “The maximum correlation coefficient (nonsymmetric case)” // *Sel. Trans. Math. Statist. Probability*, 1963, vol. 4, pp. 207-210.