Информационная дивергенция в статистической линеаризации

К. Р. Чернышев

Институт проблем управления им. В.А. Трапезникова РАН

myau@ipu.ru

Аннотация. **В рамках настоящей работы развивается обобщённая концепция статистической линеаризации входо-выходных отображений нелинейных дискретных стохастических систем, функционирующих под управлением гауссовского белошумового процесса. Предлагаемая концепция опирается на привлечение произвольных информационно-дивергентных функционалов к процедурам статистической линеаризации и ориентирована на преодоление ограничений, присущих классическим корреляционным и дисперсионным (выстроенным через корреляционные соотношения) характеристикам, возникающих при формировании линеаризованных входо-выходных моделей. Детально исследуются новые разновидности информационных расхождений, ассоциированные с симметризованным расхождением Кульбака–Лейблера и параметрическим расхождением Реньи; в частности, при значении параметра, равном двум, анализируется расхождение Коши–Шварца.**

Ключевые слова: информационная дивергенция, критерий на основе совпадения информационных дивергенций, системы со многими входами / многими выходами, непараметрическая идентификация

I. Введение

В данной работе рассматривается часть, посвященная информационной дивергенции. В целом, исследования в этой области ведутся уже много лет, можно отметить некоторые недавние результаты. Настоящая работа посвящена методике статистической линеаризации, основанной на альтернативном подходе к традиционным корреляционным методам, а именно на информационной дивергенции. Задача статистической линеаризации входо-выходных отображений принадлежит к тому классу задач идентификации, качество решения которых в критической степени обусловлено характером статистической взаимосвязи между входными и выходными сигналами идентифицируемого объекта. Вместе с тем известные подходы к решению данной задачи опираются либо на привлечение стандартных корреляционных функций, либо на использование дисперсионных характеристик (через корреляционные соотношения), что способно порождать модели, выходной процесс которых оказывается тождественно нулевым. Корень этой проблемы заключается в том, что корреляционные характеристики способны обращаться в нуль даже в присутствии функциональной (детерминированной) связи между переменными. Соответствующие примеры можно найти в [1, 2].

Большинство литературных источников по корреляционной статистической линеаризации можно найти в монографии [3], публикациях [4, 5], монографии [6] и многочисленных смежных источниках. Необходимо подчеркнуть, что статистическая линеаризация сохраняет свой статус как объекта теоретического изучения, так и действенного практического инструмента, востребованного при решении широкого спектра прикладных задач и в качестве аналитического средства в различных теоретических построениях.

Путь, позволяющий преодолеть отмеченные недостатки корреляционных и дисперсионных характеристик при конструировании линеаризованных входо-выходных моделей, состоит в привлечении информационных расхождений произвольного типа.

II. Информационная дивергенция

Как известно, для k −многомерной случайной величины Z ее дивергенции $D_Z(g_1||g_2)$ называются информационным расстоянием между k −многомерными плотностями распределения $g_1(\boldsymbol{z})$ и $g_2(z)$. В этом случае, если $g_1(\boldsymbol{z}) = g_2(\boldsymbol{z})$ почти всюду, то $D_Z(g_1||g_2) = 0$, и, в общем случае, дивергенция принимает значения на всей неотрицательной числовой полуоси.

Рассмотрим случай, когда одна из плотностей, $g_2(\boldsymbol{z})$, допускает факторизацию в виде произведения двух маргинальных плотностей: $g_{21}(\boldsymbol{z}_1)$, заданной в -мерном пространстве $(n < k)$, и $g_{22}(\boldsymbol{z}_2)$, заданной в m-мерном пространстве ($m < k$, причем $n + m = k$). В таком случае другая плотность, $g_1(\boldsymbol{z})$, естественным образом трактуется как совместное распределение случайной величины Z, представимой в виде конкатенации -мерного и m-мерного случайных векторов Z_1 и Z_2:

$$Z = (Z_1^T \; Z_2^T)^T. \tag{1}$$

В этом случае расхождение $D_Z(g_1||g_2)$ становится *информационной* дивергенцией,

$$D_Z(g_1||g_2) = D_Z(g_1||g_{21} \cdot g_{22}). \tag{2}$$

Поскольку в общем случае $D_Z(g_1||g_2) \neq D_Z(g_2||g_1)$ естественно рассматривать симметричную дивергенцию, $S_Z(g_1||g_2)$, определяемую выражением

$$S_Z(g_1||g_2) = \frac{1}{2}\big(D_Z(g_1||g_2) + D_Z(g_2||g_1)\big). \tag{3}$$

Поскольку значения дивергенции лежат на неотрицательной числовой полуоси, естественно возникает вопрос о нормализации её значений, скажем, в диапазоне от 0 до 1. В противном случае трудно сказать, какое значение информационной дивергенции является большим или малым с точки зрения количественной характеристики зависимости между случайными векторами. Такой подход осуществим следующим образом. Поскольку, как известно, обнуление коэффициента корреляции для гауссова распределения вероятностей эквивалентно независимости соответствующих случайных величин, любую информационную дивергенцию необходимо вычислить для стандартного гауссова распределения, то есть информационную дивергенцию необходимо построить как функцию модуля коэффициента корреляции. Затем эту функцию инвертировать. Полученное выражение определит соответствующую информационную дивергенцию. Но следует отметить, что аналитической инверсии такой функции может и не существовать; её существование определяется исходной аналитической формой самой информационной дивергенции.

III. Постановка задачи статистической линеаризации по критерию, основанному на информационной дивергенции

Пусть в некоторой устойчивой дискретной системе MIMO (много входов/много выходов) с размерностью nдля выхода и mдля входа соответствующее отображение вход/выход описывается бесконечным множеством плотностей распределения вероятностей (разумеется, неизвестных исследователю).

$$p_{y_i,w_j}(y,w,\tau), i = 1,\dots,n, j = 1,\dots,m, \quad \tau = 1,2,\dots. \tag{4}$$

В системе наблюдаемый выходной процесс, строго стационарный по отношению к входному сигналу, описывается как $Y(t) = \left(y_1(t),\dots,y_n(t)\right)^T$, а наблюдаемый входной процесс $W(s) = (w_1(s),\dots,w_m(s))^T$ представляет собой гауссовский процесс белого шума с известной ковариационной матрицей C_W.

$$C_W = \begin{pmatrix} 1 & c_{12} & \cdots & c_{1m} \\ c_{12} & \ddots & & \vdots \\ \vdots & & \ddots & c_{(m-1)m} \\ c_{1m} & \cdots & c_{(m-1)m} & 1 \end{pmatrix}. \tag{5}$$

Здесь предполагается, что компоненты процесса $W(s)$ имеют единичную дисперсию. Компоненты процесса $Y(t)$ имеют нулевое среднее значение и единичную дисперсию, и $W(s)$ предполагается, что компоненты также имеют нулевое среднее значение. Очевидно, что такие предположения не уменьшают общность постановки задачи, поскольку любой случайный процесс может быть сведен к форме компонентов с нулевым средним значением и единичной дисперсией. Все эти предположения можно выразить следующим образом:

$$\begin{gathered} \mathbf{E}\{y_i(t)\} = \mathbf{E}\{w_j(s)\} = 0, \\ \mathbf{var}\{y_i(t)\} = \mathbf{var}\{w_j(s)\} = 1, \\ i = 1,\dots,n, j = 1,\dots,m, \end{gathered} \tag{6}$$

где символы $\mathbf{E}\{\cdot\}$, $\mathbf{var}\{\cdot\}$ используются для обозначения математического ожидания и дисперсии соответственно.

Линеаризованная модель системы естественным образом представляется в виде:

Таким образом, естественно представить линеаризованную модель системы в следующей форме:

$$\hat{Y}(t;\mathbf{G}) = \sum_{k=1}^{\infty} G(k)W(t-k), t = 1,2,\dots, \tag{7}$$

с выходным вектором $\hat{Y}(t;\mathbf{G}) = (\hat{y}_1(t;\mathbf{G}),\dots,\hat{y}_n(t;\mathbf{G}))^T$ и матричными $(n \times m)$ весовыми коэффициентами $G(k)$ линейного отображения, подлежащими нахождению в соответствии с выбранным критерием статистической линеаризации.

Роль такого критерия играет требование согласования информационных дивергенций: расхождение между i-й компонентой $y_i(t)$ выхода $Y(t)$ и j-й компонентой $w_j(s)$ входа системы, вычисленное по совместным плотностям (4), должно равняться аналогичному расхождению для соответствующих компонент выхода $\hat{y}_i(t,\mathbf{G})$ и входа $w_j(s)$ модели (7) при всех допустимых значениях индексов $i = 1,\dots,n$, $j = 1,\dots,m$.

Аналитическая запись этого критерия имеет вид:

$$\begin{gathered} D_{y_i w_j}\left(p_{y_i,w_j} \middle\| p_{y_i} p_{w_j}\right) = \\ = D_{\hat{y}_i(\mathbf{G}) w_j}(p_{\hat{y}_i(\mathbf{G}),w_j} \| p_{\hat{y}_i(\mathbf{G})} p_{w_j}), \tau = 1,2,\dots. \end{gathered} \tag{8}$$

Критерий (8) параметризован дискретным временным сдвигом $\tau = 1,2,\dots$. Хотя формулировка (8) эксплицитно не содержит условия равенства математических ожиданий выходных компонент системы и модели, в рассматриваемой постановке данное условие оказывается автоматически выполненным:

$$\mathbf{E}\{y_i(t)\} = \mathbf{E}\{\hat{y}_i(t;\mathbf{G})\} = 0, \tag{9}$$

Столь же закономерным является покомпонентное совпадение дисперсий выходных процессов модели (7) и исходной системы:

$$\mathbf{var}\{\hat{y}_i(t;\mathbf{G})\} = 1, i = 1,\dots,n. \tag{10}$$

Таким образом, уравнение (10) устанавливает условие на $\vec{g}_i(k) = (g_{i1}(k),\dots,g_{im}(k))$, $i - 1$ — это строки матрицы $G(k)$ из (7). Такое условие имеет вид

$$\sum_{k=1}^{\infty} \vec{g}_i(k) C_W \vec{g}_i^T(k) = 1, i = 1,\dots,n. \tag{11}$$

Соотношение (11) можно получить непосредственно, используя цепочку равенств.

$$1 = \mathbf{var}\{\hat{y}_i(t;\mathbf{G})\}$$
$$= \mathbf{var}\left\{\sum_{k=1}^{\infty}\vec{g}_i(k)W(t-k)\right\} =$$
$$= \sum_{k=1}^{\infty}\vec{g}_i(k)\mathbf{E}\{W(t-k)W^T(t-k)\}\vec{g}_i^T(k) +$$
$$+ \sum_{p\neq q}\vec{g}_i(p)\mathbf{E}\{W(t-p)W^T(t-q)\}\vec{g}_i^T(q) =$$
$$= \sum_{k=1}^{\infty}\vec{g}_i(k)\,C_W\vec{g}_i^T(k). \quad (12)$$

Это следует непосредственно из описания модели (7) и условий нормализации (9), (10) и (11).

Таким образом, выражение (8) вместе с условием (9) представляет собой информационно-теоретический критерий статистической линеаризации системы, описываемой неизвестными плотностями распределения вероятностей (9).

IV. Построение универсального подхода (для любых информационных дивергенций)

Используя выражение (11), мы можем ввести следующий набор случайных значений.

$$v_i\langle -\tau; t\rangle = \sum_{j=1}^{\tau-1}\vec{g}_i(j)W(t-j) + \sum_{j=\tau+1}^{\infty}\vec{g}_i(j)W(t-j)\,, \qquad \tau = 1, 2, \ldots. \quad (13)$$

Они явно имеют гауссово распределение с нулевыми средними значениями и дисперсиями вида

$$\mathbf{var}\{v_i\langle -\tau; t\rangle\} = \sum_{j=1}^{\tau-1}\vec{g}_i(j)C_W\vec{g}_i^T(j) + \sum_{j=\tau+1}^{\infty}\vec{g}_i(j)C_W\vec{g}_i^T(j) = 1 - \vec{g}_i(\tau)C_W\vec{g}_i^T(\tau), \qquad \tau = 1, 2, \ldots. \quad (14)$$

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

$$V_i(t,\tau) = (v_i\langle -\tau; t\rangle, w_1(t-\tau), \ldots, w_m(t-\tau))^T \quad (15)$$

имеет гауссово распределение вероятностей, для которого ковариационная матрица равна

$$C_{V_i(t,\tau)} = \begin{pmatrix} 1-\vec{g}_i(\tau)C_W\vec{g}_i^T(\tau) & 0 & \cdots & \cdots & 0 \\ 0 & 1 & c_{12} & \cdots & c_{1m} \\ \vdots & c_{12} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & c_{(m-1)m} \\ 0 & c_{1m} & \cdots & c_{(m-1)m} & 1 \end{pmatrix}, \quad (16)$$

А для двумерного случайного вектора $(\hat{y}_i(t;\mathbf{G}) \quad w_j(t-\tau))^T$ мы можем записать следующее соотношение

$$\begin{pmatrix}\hat{y}_i(t;\mathbf{G}) \\ w_j(t-\tau)\end{pmatrix} = A_{ij}(\tau)V_i(t,\tau), \quad (17)$$

где $A_{ij}(\tau)$ – матрица размерности $(2\times(m+1))$. Ее форма следующая:

$$A_{ij}(\tau) = \begin{pmatrix} 1 & g_{i1}(\tau) & \cdots & g_{ij}(\tau) & g_{i(j+1)}(\tau) & \cdots & g_{im}(\tau) \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}, \quad (18)$$

Как и ранее, $g_{i1}(\tau), \ldots, g_{im}(\tau)$суть элементы -й строки матрицы $G(\tau)$ из представления (7).

Следовательно, $(\hat{y}_i(t;\mathbf{G}) \quad w_j(t-\tau))^T$ является случайным вектором с гауссовым распределением вероятностей; его ковариационная матрица равна $C_{(\hat{y}_i w_i)}(\tau)$,

$$C_{(\hat{y}_i w_i)}(\tau) = A_{ij}(\tau)C_{V_i(t,\tau)}A_{ij}^T(\tau). \quad (19)$$

Вычисление произведения этих матриц дает следующее выражение:

$$C_{(\hat{y}_i w_i)}(\tau) = \begin{pmatrix} 1 & \gamma_{ij}(\tau) \\ \gamma_{ij}(\tau) & 1 \end{pmatrix}, \quad (20)$$

где $j-1$-й компонент вектор-столбца $C_W\vec{g}_i^T(\tau)$ выражается как $\gamma_{ij}(\tau)$.

Для строк $i-$х, $\vec{g}_i(\tau), = 1, 2, \ldots$ весовых коэффициентов с матричным значением модели (7) $G(\tau)$ непосредственно следует:

$$\vec{g}_i^T(\tau) = C_W^{-1}\cdot\mathbf{D}_{y_i W}(\tau), \quad (21)$$

где

$$\mathbf{D}_{y_i W}(\tau) = \begin{pmatrix} \mathrm{sign}\left(\mathbf{m}_{y_i|w_1}(\tau)\right)\times\phi^{-1}\left(D_{y_i w_1}\left(p_{y_i,w_j}||p_{y_i}p_{w_j}\right)\right) \\ \vdots \\ \mathrm{sign}\left(\mathbf{m}_{i|w_j}(\tau)\right)\times\phi^{-1}\left(D_{y_i w_j}\left(p_{y_i,w_j}||p_{y_i}p_{w_j}\right)\right) \\ \vdots \\ \mathrm{sign}\left(\mathbf{m}_{я_i|w_m}\tau)\times\right)\times\phi^{-1}\left(D_{y_i w_m}\left(p_{y_i,w_j}||p_{y_i}p_{w_j}\right)\right) \end{pmatrix}. \quad (22)$$

В формуле (22) $\mathbf{m}_{y_i|w_j}(\tau)$ — регрессия $y_i(t)$ на $w_j(t-\tau)$; $\mathrm{sign}(x) = 1$ для $x \geq 0$, $\mathrm{sign}(x) = -1$ для $x < 0$. Она служит для учета координации компонентов входных и выходных процессов системы. Также выражение $\phi^{-1}\left(D_{y_i w_j}\left(p_{y_i,w_j}||p_{y_i}p_{w_j}\right)\right)$ в формуле (22) всегда неотрицательно, информационная дивергенция всегда неотрицательна,$[0, +\infty)$.

Таким образом, для системы, описываемой неизвестными плотностями распределения (4), строится модель (7), матричные весовые коэффициенты которой ищутся на основе теоретико-информационного критерия (8). Это гарантирует, что процесс выхода модели отличается от нуля, если хотя бы одна компонента этих весовых коэффициентов отличается от нуля. Но если все весовые коэффициенты построенной модели (9) равны нулю, то такая ситуация эквивалентна неидентифицируемости системы. Кроме того, если все весовые коэффициенты равны нулю, то процесс выхода системы не будет удовлетворять условию покомпонентной единичности дисперсии. То есть, решение задачи статистической линеаризации вообще не существует в указанной формулировке задачи.

V. Системы с нулевой корреляцией вход/выход

Как указывалось во Введении, известен обширный класс ситуаций, в которых привлечение классических корреляционных подходов к конструированию моделей не приводит к удовлетворительным результатам. Среди систем подобного рода особое место занимают те, чья входо-выходная зависимость описывается вероятностными распределениями, принадлежащими семейству, введённому О. В. Сармановым [7, 8]. К данному семейству относится, в частности, следующая двумерная плотность:

$$\begin{gathered} p(x,y) = \frac{e^{-\frac{x^2+y^2}{2}}}{2\pi} \times \\ \times \left(1 + \lambda\left(2e^{-\frac{3}{2}x^2} - 1\right)\left(2e^{-\frac{3}{2}y^2} - 1\right)\right), \\ -1 \le \lambda \le 1. \end{gathered} \tag{23}$$

Значения параметра λ существенно влияют на форму плотности (23).

Для распределения (23) как обычный коэффициент корреляции $r_{X,Y}$, так и корреляционное отношение $\theta_{X,Y}$ тождественно обращаются в нуль, тогда как максимальный коэффициент корреляции задаётся формулой:

$$S_{X,Y}(\lambda) = \left(\frac{4}{\sqrt{7}} - 1\right)|\lambda|. \tag{24}$$

В качестве иллюстрации рассмотрим ситуацию, в которой стохастическая связь между выходным процессом $y(t)$ и входным процессом $w(s)$ нелинейной системы задаётся плотностью вида (23) с параметром $\lambda = \lambda(\tau)$, где $\tau = t - s$. Применение как стандартных корреляционных методов, так и дисперсионной статистической линеаризации в этом случае неминуемо привело бы к тождественно нулевому выходному процессу построенной модели (7). Предлагаемый подход, опирающийся на состоятельные меры зависимости, позволяет исключить подобное вырождение. Вместе с тем необходимо констатировать, что при нулевой дисперсии выходного сигнала задача в рассмотренной формулировке вообще лишена решения, поскольку это вступает в противоречие с исходным ограничением (10).

VI. Выводы

В статье представлено решение задач идентификации, опирающееся на критерий, сформулированный через информационное расхождение совместной плотности вероятности входных и выходных сигналов системы и произведения соответствующих маргинальных плотностей. Принципиальной чертой развиваемого подхода является его универсальность: информационное расхождение используется в обобщенной форме, не обусловленной ни конкретным видом совместных плотностей распределения, ни конкретной разновидностью дивергенции. На базе данной методологии разработана процедура статистической линеаризации многомерных входо-выходных отображений стохастических дискретных систем, управляемых многомерным гауссовским белошумовым процессом. Критерий линеаризации объединяет покомпонентное согласование математических ожиданий выходных сигналов системы и модели с покомпонентным согласованием информационных дивергенций входо-выходных пар системы и модели. Получены замкнутые аналитические выражения для коэффициентов весовой функции линеаризованной модели как функций информационного расхождения (формирующего критерий линеаризации) входных и выходных сигналов идентифицируемого объекта. Установлено, что указанная функция определяет тип нелинейного преобразования, обеспечивающего построение выражения для информационной дивергенции, нормированной на единичном отрезке.

Список литературы

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

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

[3] Roberts J.B. and P.D. Spanos. *Random Vibration and Statistical Linearization*, Dover, New York, 464 p., 2003.

[4] Socha L. “Linearization in Analysis of Nonlinear Stochastic Systems: Recent Results – Part I: Theory”, *Applied Mechanics. Reviews*, vol. 58, no. 3, pp. 178-205, 2005.

[5] Socha L. “Linearization in Analysis of Nonlinear Stochastic Systems: Recent Results –Part II: Applications”, *Applied Mechanics. Reviews*, no. 5, pp. 303-315, 2005.

[6] Socha L. *Linearization Methods for Stochastic Dynamic Systems*, Lect. Notes Phys. 730, Springer Berlin, Heidelberg, 383 p., 2008.

[7] Balakrishnan N. and Chin-Diew Lai. *Continuous Bivariate Distributions* / Second Edition, Springer, 712 p., 2009.

[8] Sarmanov O.V. “Remarks on uncorrelated Gaussian dependent random variables”, *Theory Probab. Appl.*, vol. 12, no. 1, pp. 124-126, 1967.