

# Статистическая линейаризация на основе максимальной коррэнтропии

К. Р. Чернышев

*Институт проблем управления им. В.А. Трапезникова РАН*

myau@ipu.ru

**Аннотация.** Настоящая работа развивает теоретический аппарат построения линейных приближений для операторов «вход–выход» нелинейных дискретных стохастических систем, функционирующих под действием гауссовского белого шума. В основу предлагаемого подхода положен принцип согласования коррэнтропийных характеристик: требуется одновременное выполнение двух условий — тождественность первых моментов выходных сигналов исследуемого объекта и его линейного аналога, а также равенство максимальных коррэнтропийных коэффициентов, вычисленных для пар «вход–выход» обеих систем. В рамках изложенного подхода выведены замкнутые аналитические соотношения, явно задающие коэффициенты импульсной характеристики линейаризованной модели. Дополнительно разработана процедура ослабления влияния аддитивных ненаблюдаемых помех в выходном канале при условии, что закон распределения этих помех известен заранее.

**Ключевые слова:** *состоятельные меры зависимости, статистическая линейаризация, максимальный коррэнтропийный коэффициент, системы с одним входом / одним выходом, непараметрическая идентификация*

## I. ВВЕДЕНИЕ

Задача аппроксимации сложного нелинейного оператора его линейным эквивалентом в вероятностном смысле принадлежит к числу фундаментальных проблем теории случайных процессов и систем автоматического управления. Такая аппроксимация, традиционно именуемая статистической линейаризацией, позволяет заменить труднообозримое нелинейное преобразование случайных сигналов его ближайшим линейным аналогом, сохраняющим наиболее существенные статистические свойства оригинала.

Исторически основной инструментарий статистической линейаризации формировался вокруг корреляционного анализа: взаимная корреляционная функция между входом и выходом исследуемой системы служила и мерой связи, и основой для построения линейной модели. Подобный подход получил широкое распространение в инженерных дисциплинах и продолжает активно развиваться в прикладных контекстах — от анализа вибрационных процессов в системах сбора энергии на основе пьезоэлектрических материалов [1] до планирования робастных траекторий [2], оценки нестационарного отклика систем с дробными производными [3], оптимизации полуактивного управления [4], моделирования случайных колебаний плавучих ветроэнергетических установок [5] и проектирования экспериментов для идентификации робототехнических манипуляторов [6].

Вместе с тем корреляционный подход обладает принципиальным ограничением, которое неоднократно отмечалось в литературе и хорошо иллюстрируется классическими контрпримерами. Коэффициент корреляции Пирсона, лежащий в основе взаимнокорреляционного анализа, способен фиксировать лишь линейную составляющую стохастической связи. Существуют детерминированные функциональные зависимости (например, квадратичные или периодические), при которых корреляция между величинами обращается в нуль, несмотря на полную функциональную обусловленность одной величины другой. Данное обстоятельство свидетельствует о том, что корреляция не является состоятельной мерой зависимости, а значит, линейаризация, опирающаяся исключительно на корреляционный критерий, может не отражать истинной структуры связей в нелинейной системе.

Осознание этого факта мотивирует обращение к более общим характеристикам стохастической зависимости. Математически строгий подход к определению «хороших» мер зависимости восходит к работам А.Н. Колмогорова, который сформулировал представление о том, какими свойствами должен обладать числовой индикатор связи между случайными объектами, — так называемая состоятельность (непротиворечивость) меры зависимости. В дальнейшем А. Реньи [7] формализовал и существенно уточнил это представление, предложив систему из семи аксиом, которым должна удовлетворять мера зависимости двух случайных величин. Эти аксиомы, выдержавшие многолетнюю проверку, стали общепризнанным стандартом в данной области. Меры, удовлетворяющие всей аксиоматике Реньи, образуют довольно узкий подкласс в совокупности всех колмогоровских состоятельных мер.

Одним из центральных результатов Реньи [7] стало доказательство того, что максимальная корреляция (supremum коэффициента корреляции по всем измеримым преобразованиям исходных случайных величин) удовлетворяет всем семи аксиомам, то есть является состоятельной мерой зависимости в наиболее сильном смысле. Этот факт делает максимальную корреляцию теоретически идеальным инструментом для задач идентификации и линейаризации. Однако непосредственное вычисление максимальной корреляции по выборке наталкивается на серьезные трудности: необходимость оптимизации по бесконечномерному пространству измеримых функций не допускает, как правило, эффективной численной реализации.

В связи с этим актуален поиск таких приближений максимальной корреляции, которые, с одной стороны, наследовали бы ее ключевые аксиоматические свойства (или по крайней мере большинство из них), а с другой — допускали бы вычисление по конечным выборочным данным с приемлемой трудоемкостью. Именно такой характеристикой является максимальный коррэнтропийный коэффициент, введенный в рамках теории информационного обучения [8]. Этот коэффициент основан на замене произвольных измеримых преобразований аффинными и на использовании ядерных оценок плотности вместо истинных распределений. Результатом является «почти состоятельная» мера зависимости: она удовлетворяет первым пяти аксиомам Реньи при симметричном ядре, а шестая аксиома (для аффинных преобразований) и седьмая (предельный переход к точной зависимости) выполняются в определенных режимах выбора параметров ядра.

В настоящей статье указанный коррэнтропийный критерий систематически применяется к задаче построения линейного приближения отображения «вход–выход» нелинейной дискретной стохастической системы. Предлагаемая схема опирается на требование совпадения максимальных коррэнтропийных коэффициентов между входом и выходом исходного

объекта, с одной стороны, и между входом и выходом линейной модели — с другой, при одновременном согласовании первых моментов.

Дальнейшее изложение организовано следующим образом. Раздел 2 содержит необходимые сведения о максимальном коррэнтропийном коэффициенте, его определение, связь с классическими информационными функционалами и выборочную форму, удобную для вычислений. В разделе 3 сформулирован и решен основной результат — получены явные выражения для коэффициентов импульсной характеристики линейаризованной модели на основе коррэнтропийного критерия согласования. Раздел 4 посвящен обобщению на ситуацию, когда наблюдаемый выход системы искажен аддитивным ненаблюдаемым шумом с известным законом распределения, и описана процедура восстановления «бесшумной» совместной плотности, допускающая последующее применение основного результата.

## II. МАКСИМАЛЬНЫЙ КОРРЭНТРОПИЙНЫЙ КОЭФФИЦИЕНТ

Следуя [8], для произвольной пары невырожденных случайных величин  $X$  и  $Y$ , обладающих непрерывными распределениями, определим максимальный коррэнтропийный коэффициент  $\Gamma^*(X, Y)$  соотношением

$$\Gamma^*(X, Y) = \sup_{a, b \in \mathbb{R}} \frac{\iint \kappa(y, ax+b)(p_{XY}(x, y) - p_X(x)p_Y(y)) dx dy}{\int \kappa(ax+b, ax+b)(p_{XY}(x, x) - p_X(x)p_Y(x)) dx \int \kappa(y, y)(p_{XY}(y, y) - p_X(y)p_Y(y)) dy}$$

где  $\kappa(\cdot, \cdot)$  — положительно определенное ядро.

Термин «коррэнтропия» объясняется тем фактом, что если в выражении

$$\iint \kappa(y, x) p_{XY}(x, y) dx dy$$

положить (не требуя положительной определенности ядра  $\kappa(y, x)$ )

$$\kappa(y, x) = -\ln p_{XY}(x, y),$$

тогда мы получим выражение для дифференциальной энтропии; если мы положим

$$\Gamma_N^*(X, Y) = \sup_{a, b \in \mathbb{R}} \frac{\frac{1}{N} \sum_{i=1}^N G_\sigma(y_i, ax_i + b) - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G_\sigma(y_j, ax_i + b)}{\sqrt{\frac{1}{2\pi\sigma} - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G_\sigma(ax_j, ax_i)}} \sqrt{\frac{1}{2\pi\sigma} - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G_\sigma(y_j, y_i)}$$

Здесь принимается во внимание, что  $G_\sigma(ax_j + b, ax_i + b) = G_\sigma(ax_j, ax_i)$  and  $G_\sigma(0) = \frac{1}{\sqrt{2\pi\sigma}}$ .

Величина  $|\Gamma_N^*(X, Y)|$  удовлетворяет первым пяти аксиомам Реньи (см. [7] ( $A - E$  (при условии выбора симметричного ядра)). Аксиома  $G$  выполняется (приближенно) для очень больших ядер [8]. Для преобразований аффинного типа выполняется аксиома  $F$ . Аксиома  $F$  не играет существенной роли в задачах идентификации, но использование максимального

$$\kappa(y, x) = \frac{1}{\alpha - 1} (1 - p_{YX}^{\alpha-1}(x, y)),$$

где  $\alpha > 0, \alpha \neq 1$  — числовой параметр, следует выражение для энтропии Цаллиса порядка  $\alpha$ .

Далее, когда  $\kappa(\cdot, \cdot)$  симметричен,  $\Gamma^*(X, Y)$  симметричен,  $\Gamma^*(X, Y) = \Gamma^*(Y, X)$ . В дальнейшем мы будем рассматривать только симметричные ядра, основное внимание уделяя ядру Гаусса

корреляционного коэффициента значительно предпочтительнее максимальной корреляции.

Условия оптимальности для определения  $a$  и  $b$  имеют, в конечном счете, простой вид:  $\frac{d|\Gamma_N^*(X, Y)|}{da} = 0$ ;  $\frac{d|\Gamma_N^*(X, Y)|}{db} = 0$ .

Естественно, эта система из двух уравнений с двумя переменными  $(a, b)$  может быть решена только численно, но эффективно.

### III. ПОСТРОЕНИЕ ЛИНЕЙНОЙ МОДЕЛИ НА ОСНОВЕ КОРРЭНТРОПИЙНОГО КРИТЕРИЯ

Рассмотрим нелинейную динамическую систему с дискретным временем, выходной процесс которой  $y(t)$  предполагается строго стационарным и эргодическим. Входом системы служит случайный процесс  $w(t)$ , представляющий собой гауссовский белый шум. Совместное распределение компонент входного и выходного процессов описывается семейством плотностей (в общем случае неизвестных исследователю):

$$p_{y,w}(y, w, \tau), \tau = 1, 2, \dots \quad (1)$$

Оба процесса нормированы: математические ожидания равны нулю, дисперсии равны единице:

$$\begin{aligned} \mathbf{E}\{y_i(t)\} &= \mathbf{E}\{w_j(s)\} = 0, \\ \mathbf{var}\{y(t)\} &= \mathbf{var}\{w(t)\} = 1. \end{aligned} \quad (2)$$

Случайные процессы  $y(t)$  и  $w(t)$  являются совместно стационарными в строгом смысле.

Линейный аналог системы (1) ищется в форме бесконечной свертки:

$$\hat{y}(t; G) = \sum_{k=1}^{\infty} g(k)w(t-k), t = 1, 2, \dots, \quad (3)$$

где  $\hat{y}(t; G)$  – выход модели,  $G = \{g(k), k \in [1, \infty)\}$ ,  $g(k), k = 1, 2, \dots$  – последовательность подлежащих определению коэффициентов импульсной характеристики. Критерий подбора линейной модели формулируется как совокупность двух требований: тождество первых моментов выходов объекта и модели и покомпонентное равенство максимальных коррентропийных коэффициентов для пар «выход–задержанный вход»:

$$\mathbf{E}\{y(t)\} = \mathbf{E}\{\hat{y}(t; G)\} = 0, \quad (4)$$

$$|\Gamma_N^*(y(t), w(t-k))| = |\Gamma_N^*(\hat{y}(t; G), w(t-k))|. \quad (5)$$

Введем вспомогательные случайные величины:

$$v_t(-k) = \sum_{j=1}^{k-1} g(j)w(t-j) + \sum_{j=k+1}^{\infty} g(j)w(t-j), \\ k = 1, 2, \dots.$$

Тогда в рамках введенных обозначений и в силу описания модели (2), (3) могут быть записаны следующие матричные равенства:

$$\begin{pmatrix} \hat{y}(t; G) \\ w(t-k) \end{pmatrix} = \begin{pmatrix} 1 & g(k) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_t(-k) \\ w(t-k) \end{pmatrix},$$

$$\begin{pmatrix} v_t(-k) \\ w(t-k) \end{pmatrix} = \begin{pmatrix} 1 & -g(k) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{y}(t; G) \\ w(t-k) \end{pmatrix}.$$

Таким образом, эти две формулы представляют собой линейное преобразование гауссова случайного вектора и, следовательно,

$$p_{\hat{y}(G), w}(\hat{y}(y), w, k) = \\ = \frac{1}{2\pi\sqrt{1-g^2(k)}} e^{-\frac{\begin{bmatrix} \hat{y}(G) \\ w \end{bmatrix}^T \begin{bmatrix} 1 & -g(k) \\ -g(k) & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{y}(G) \\ w \end{bmatrix}}{2}}, k = 1, 2, \dots$$

то есть плотность является гауссовой. Следовательно, в силу аксиомы Реньи  $G$  [7], из этого непосредственно следует, что

$$S_{\hat{y}(G)w}(k) = |g(k)|, k = 1, 2, \dots$$

и, следовательно,

$$|\Gamma_N^*(\hat{y}(t), w(t-k))| = S_{\hat{y}(G)w}(k), k = 1, 2, \dots$$

и при соблюдении условия (5)

$$|g(k)| = |\Gamma_N^*(y(t), w(t-k))|, k = 1, 2, \dots \quad (6)$$

Чтобы «убрать» знак модуля в (5), следует применить знак регрессии выходного процесса к входному, т. е.

$$\mathit{sign}[\mathit{reg}_{yw}(k)] = \begin{cases} 1, & \mathit{reg}_{yw}(k) \geq 0 \\ -1, & \mathit{reg}_{yw}(k) < 0 \end{cases}, k = 1, 2, \dots,$$

здесь

$$\mathit{reg}_{yw}(k) = \mathbf{E}\left\{\frac{y(t)}{w(t-k)}\right\}, k = 1, 2, \dots$$

Итак, наконец,

$$g(k) = \mathit{sign}[\mathit{reg}_{yw}(k)] |\Gamma_N^*(y(t), w(t-k))|, \\ k = 1, 2, \dots \quad (7)$$

Последнее уравнение определяет коэффициенты весовой функции линеаризованной модели (3).

### IV. ШУМОПОДАВЛЕНИЕ

Пусть теперь выходной процесс исходной нелинейной системы, описываемый плотностями вида (1), возмущается ненаблюдаемым строго стационарным шумом  $\xi(t)$  с нулевым средним значением и известной функцией плотности вероятности  $p_\xi(\xi)$ , т. е.

$$p_{\tilde{y}}(\tilde{y}) = \int p_y(z) p_\xi(y-z) dz. \quad (8)$$

Предполагается также, что процессы  $\xi(t)$  и  $w(t)$  стохастически независимы.

В этих условиях постановка задачи линейной аппроксимации должна быть модифицирована: статистическое согласование модели (3) осуществляется не с зашумленным, а с «очищенным» выходом системы,

то есть с процессом, соответствующим условию  $\xi(t) \equiv 0$  почти наверное:

$$y(t) \equiv \tilde{y}(t) \text{ при } \xi(t) \equiv 0 \text{ почти наверное.}$$

Критерий (4), (5) остается в силе, но формулируется относительно ненаблюдаемого «бесшумного» процесса  $\tilde{y}(t)$ .

Для практической реализации результатов раздела 3 требуется восстановить совместную плотность  $p_{y,w}(y, w, k)$  пары  $(y(t), w(t))$ . Обозначим через  $p_{\tilde{y},w}(\tilde{y}, w, k)$  совместную плотность наблюдаемого (зашумленного) выхода  $\tilde{y}(t)$  и входа  $w(t)$ . Благодаря стохастической независимости  $\xi(t)$  и  $y(t)$  искомая плотность связана с наблюдаемой интегральным соотношением

$$p_{y,w}(y, w, k) = \int_{-\infty}^{\infty} p_{\tilde{y},w}(\tilde{y} + \xi, w, k) p_{\xi}(\xi) d\xi.$$

Таким образом, применяя непараметрические методы оценивания к совместной плотности  $p_{\tilde{y},w}(\tilde{y}, w, k)$  на основе доступных выборочных данных и решая обратную задачу деконволюции, можно использовать формулу (6), а затем определить коэффициенты импульсной характеристики (7) линеаризованной модели (3) для системы с зашумленным выходом (8). Дополнительно отметим, что при наличии аддитивного шума функция регрессии сохраняется:  $reg_{yw}(k) = reg_{\tilde{y}w}(k)$ .

## V. Выводы

В статье предложен и теоретически обоснован подход к задаче линейной аппроксимации нелинейных стохастических систем, принципиально отличающийся от традиционной корреляционной статистической линеаризации. Основу подхода составляет критерий согласования, при котором максимальный коррэнтропийный коэффициент между выходом объекта и произвольной задержанной компонентой его входа должен совпадать с аналогичным коэффициентом, вычисленным для линейной модели. Это условие дополняется стандартным требованием тождества первых моментов.

Ключевое преимущество коррэнтропийного критерия перед корреляционным заключается в его «почти состоятельности» в смысле Реньи. Максимальный коррэнтропийный коэффициент удовлетворяет пяти из семи аксиом Реньи при любом выборе симметричного положительно определенного ядра, а оставшиеся аксиомы выполняются в предельных режимах. Это означает, что предложенный критерий способен улавливать существенно более широкий спектр стохастических зависимостей по сравнению с обычной

взаимной корреляцией, которая, как известно, нечувствительна к нелинейным связям.

Важно подчеркнуть, что, несмотря на параметрическую форму итогового результата – коэффициенты  $g(k)$  импульсной характеристики, – сама задача носит непараметрический характер. Непараметричность проявляется двояко. Во-первых, число определяемых коэффициентов не фиксировано заранее и может быть произвольным. Во-вторых, для нахождения каждого коэффициента  $g(k)$  достаточно решить двумерную оптимизационную задачу (по переменным  $a$  и  $b$ ), вычислительная сложность которой составляет  $O(N)$  при объеме выборки  $N$ , что делает метод практически пригодным для обработки больших массивов данных.

Важно подчеркнуть, что, несмотря на параметрическую форму итогового результата – коэффициенты  $g(k)$  импульсной характеристики, – сама задача носит непараметрический характер. Непараметричность проявляется двояко. Во-первых, число определяемых коэффициентов не фиксировано заранее и может быть произвольным. Во-вторых, для нахождения каждого коэффициента  $g(k)$  достаточно решить двумерную оптимизационную задачу (по переменным  $a$  и  $b$ ), вычислительная сложность которой составляет  $O(N)$  при объеме выборки  $N$ , что делает метод практически пригодным для обработки больших массивов данных.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Feng Qian, da Silva L.S.P., Yabin Liao, and Lei Zuo (2023). “Statistical linearization for random vibration energy harvesting with piezoelectric material nonlinearity”, Mechanical Systems and Signal Processing, vol. 188, art. 109985.
- [2] Leparoux, C., Bonalli, R., Hérisse, B., and F. Jean (2024). “Statistical linearization for robust motion planning”, Systems & Control Letters, vol. 189, art. 105825.
- [3] Pomaro, B. and P. D. Spanos (2023). “Extended statistical linearization approach for estimating non-stationary response statistics of systems comprising fractional derivative elements”, Probabilistic Engineering Mechanics, vol. 74, art. 103471.
- [4] Viet Duc La and Ngoc Tuan Nguyen (2024). “Using statistical linearization to optimize a class of semi-active on-off control in a general state space system”, Probabilistic Engineering Mechanics, vol. 75, art. 103555.
- [5] Yifei Li, Shujin Li, Dacheng Zheng, and Renjie Han (2024). “Random vibration analysis of nonlinear floating offshore wind turbine system via statistical linearization process under coupled nonlinear wave excitation”, Ocean Engineering, vol. 296, art. 116894.
- [6] Zimmermann, S. A., Moberg, S., Gunnarsson, S., and M. Enqvist (2024). “Using statistical linearization in experiment design for identification of robotic manipulators.” Control Engineering Practice, vol. 150, art. 106008.
- [7] Rényi, A. (1959). “On measures of dependence”, Acta Acad. Sci. Math. Hung., vol. 10, no. 3-4, pp. 441-451.
- [8] Principe, J. C. (2010). Information Theoretical Learning. Rényi’s Entropy and Kernel Perspectives, Springer, New York, Dordrecht, Heidelberg, London, 515 p.