

ИСПОЛЬЗОВАНИЕ КОРРЕЛЯЦИОННОГО МЕТОДА ДЛЯ УСРЕДНЕНИЯ ХРОМАТОГРАММ

Р. Т. Сайфуллин

Самарский государственный технический университет,
443100, Самара, ул. Молодогвардейская, 244

Рассматривается задача получения усредненной на некотором множестве вводов хроматограммы. Алгоритм оценивания усредненной хроматограммы сводится к решению системы линейных алгебраических уравнений. При определенных вариантах формирования входных последовательностей информационная матрица становится теплоцевой, что упрощает ее обращение. В корреляционной хроматографии (КХ) ввод пробы осуществляется по случайному закону. Произведен анализ входных воздействий, которые могут быть использованы в КХ.

Ключевые слова: усредненная хроматограмма, корреляционная хроматография, ввод пробы по случайному закону

Корреляционная хроматография - перспективный метод получения усредненных на некотором множестве вводов хроматограмм, позволяющий увеличить отношение сигнал/шум при приемлемом суммарном времени проведения эксперимента T_s . В этом методе на вход хроматографа подается некоторая последовательность проб (входное воздействие $x(t)$), а обработка связана с вычислением оценок автокорреляционной функции $R_{xx}(T)$ входного сигнала и взаимокорреляционной функции $R_{xy}(\tau)$ входного $x(t)$ и выходного $y(t)$ сигналов.

Известно, что для любой динамической системы справедливо соотношение [1]:

$$K_{xy}(m) = \int_{-\infty}^{\infty} g(t) W_{xy}(t) dt = \int_0^{T_s} g(t) R_{xx}(T-t) dt, \quad (1)$$

где $g(t)$ - аппаратная функция (искомая хроматограмма); $g(t) = 0$ при $t < 0$ и $t > T_s$. Функция $g(t)$ может быть найдена как решение интегрального уравнения (1), где значения $R_{xy}(\tau)$ и $R_{xx}(T)$ приходится заменять на их оценки $\hat{R}(\tau)$, $\hat{R}_{xx}(T)$, полученные на интервале усреднения $(0, T)$.

Запишем функции $g(t)$, $y(t)$, $x(t)$ в виде последовательности дискретных отсчетов: $g_j = g(j\Delta)$, $j = 0, 1, 2, \dots, N$; $y_k = y(k\Delta)$, $k = 0, 1, 2, \dots, N$; $x_l = x(l\Delta)$, $l = -m, -(m-1), \dots, 0, 1, 2, \dots, N$; где Δ - интервал дискретизации по времени; $m = T_s/\Delta$; $N = T/\Delta$. Отсчет значений входных воздействий $x(t)$ в общем случае может начинаться от момента времени $(-T_s)$ и заканчиваться в момент времени T .

Сайфуллин Раухат Талгатович - профессор кафедры «Информационно-измерительная техника», д.т.н., профессор.

(в дискретном виде соответственно от $-m$ до k).

В дискретной форме уравнение, эквивалентное интегральному уравнению (1), примет следующий вид:

$$\hat{R}_{xy}(j) = \sum_{l=0}^m g_l \hat{R}_{xx}(j-l); \quad j = 0, 1, 2, \dots, m, \quad (2)$$

$$\text{где } \hat{R}_{xy}(j) = \frac{1}{k+1} \sum_{l=0}^k y_l x_{l-j}; \quad \hat{R}_{xx}(j) = \frac{1}{k+1} \sum_{l=0}^k x_l x_{l-j}.$$

Соотношение (2) несложно переписать в виде системы линейных алгебраических уравнений, которая после введения матричных обозначений

$$\bar{G} = \begin{bmatrix} g_0 \\ g_1 \\ \dots \\ g_m \end{bmatrix}; \quad \bar{Y} = \begin{bmatrix} y_0 \\ y_1 \\ \dots \\ y_k \end{bmatrix}; \quad \bar{X} = \begin{bmatrix} x_0 & x_{-1} & x_{-2} & \dots & x_{-m} \\ x_1 & x_0 & x_{-1} & \dots & x_{-(m-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_k & x_{k-1} & x_{k-2} & \dots & x_{-(m-k)} \end{bmatrix}$$

примет вид

$$\bar{X}^T \bar{Y} = (\bar{X}^T \bar{Y}) \bar{G}, \quad (3)$$

где T - знак транспонирования матрицы. Согласно (3) алгоритм получения искоемых значений G эквивалентен методу наименьших квадратов. Учитывая это обстоятельство, можно использовать все результаты, известные в регрессионном анализе для корреляционной хроматографии.

Если для вектора оценок ординат аппаратной функции (усредненной хроматограммы) ввести обозначения: $G = [g_0 \ g_1 \ g_2 \ \dots \ g_m]^T$, то

$$G = (X^T X)^{-1} X^T Y, \quad (4)$$

а точность оценивания может быть охарактеризована с помощью ковариационной матрицы оценок

$$D(G) = \sigma_n^2 (X^T X)^{-1} \quad (5)$$

где σ_n^2 - дисперсия помехи, относительно которой в данном случае предполагается также, что ее значения в соседних отсчетах статистически независимы.

Длиной L последовательности $\{x\}$, полученной с помощью того или иного метода генерации, назовем число, равное количеству элементов последовательности, а фактической длиной L_f - число элементов от первого ненулевого до последнего ненулевого элемента ($L_f \leq L$).

Входная последовательность может занимать различные положения на шкале дискретного времени относительно интервала усреднения $(0-k)$. В зависимости от этого будут получаться матрицы $A = X^T X$ различного вида. Анализ характерных вариантов позволяет выделить два случая, представляющих наибольший интерес [2]:

по

а) первое ненулевое значение входной последовательности совмещается с моментом времени $t = 0$ ($x_n = 0$ при $t < 0$), последовательность длится $B\phi$ тактов (т.е. фактически завершается в момент времени $B\phi - 1$), после чего продолжается фиксация значений выходного сигнала хроматографа еще в течение m тактов;

б) входная последовательность $\{x_n\}$ содержит две части: основную - длительностью B , начинающуюся в момент времени $t = 0$ и заканчивающуюся в момент времени $k = L - 1$, и предысторию, представляющую собой фрагмент основной части, имеющей длину m , значения которой удовлетворяют условию: $x_{-k} = x_{-k+1}$; $k = 1, 2, \dots, m$.

Для обоих вариантов матрица A является теплицевой, т.е. для ее элементов a_{ij} ($i, j = 1, 2, \dots, m+1$, $i < j$) справедливо соотношение:

$$a_{ij} = a_{j-i}; \quad i = 1, 2, \dots, m+1; \quad j = 1, 2, \dots, m-i+2.$$

С учетом симметрии матрицы A это означает, что для ее определения достаточно найти всего $(m+1)$ элемент a_{ij} ($j = 1, 2, \dots, m+1$), т.е. всего одну строку (или столбец). Обратная матрица A^{-1} не является теплицевой, но обладает дополнительной симметрией, когда $a_{ij} = a_{m+1-j, m+1-i}$ так что она полностью определяется $(m+1)/2$ элементами, где $\lfloor x \rfloor$ - ближайшее к x большее целое число. В общем случае для обращения симметричной теплицевой матрицы A , соответствующей вариантам а) и б) формирования входных последовательностей, целесообразно воспользоваться алгоритмом, изложенным в [3].

Однако ниже будут представлены процедуры формирования входных последовательностей и алгоритмы вычисления усредненного сигнала, не требующие обращения матриц большой размерности в процессе вычисления на ЭВМ. Рассмотрим вначале конкретные разновидности последовательностей, позволяющие упростить вычисления в еще большей степени.

Пусть входная последовательность X_n является случайной, подчиняющейся определенным вероятностным закономерностям. Это может быть, например, последовательность, где конкретное значение X_n выбирается равным 1 и 0 с равной вероятностью. Реально в КХ наибольшее распространение получили псевдослучайные бинарные последовательности (ПСБП), достоинством которых является стабильность и полная воспроизводимость как самих сигналов, так и их вероятностных характеристик, в частности, двоичные М-последовательности [4].

М-последовательность представляет собой последовательность из нулей и единиц с периодом (длиной) $B = 2^n - 1$, $n = 2, 3, 4, \dots$, при этом число единиц p на периоде равно $p = 2^{n-1}$. В качестве примера приведена краткая запись М-последовательности с периодом $L=15$: $1^4 0^3 1^1 0^2 1^2 0^1 1^0$, где степени показывают количество следующих подряд единичных или нулевых значений. Использование в качестве входной последовательности исключительно М-последовательностей сужает возможности КХ, поскольку период этой последовательности меняется с ростом n большими скачками, что является нежелательным.

Представляет интерес рассмотрение иных типов ПСБП, позволяющих более гибко изменять их длину. Этому требованию удовлетворяют последовательности с одноуровневой корреляционной функцией [5], в частности, наиболее известные их разновидности - последовательности Лежандра и Якоби. Период последовательности Лежандра определяется соотношением $L - 4/\ell + 3$, где L - простое число (ℓ - положительное число), а период последовательности Якоби - соотношением $L - \ell(\ell + 2)$, где ℓ и $(\ell + 2)$ - простые.

На основе ПСБП могут быть сформулированы новые варианты входных воздействий - псевдослучайные троичные последовательности (ПСТП). Рассмотрим коротко задачу синтеза таких последовательностей. Пусть для некоторого заданного периода L получена некоторая ПСБП: $\{a/\} = a_1, a_2, \dots, a_L$, $/ = 1, 2, \dots, L$, где символы $a/$ принимают значения 0 или 1. Вначале формируется вспомогательная последовательность $\{b/\} = b_1, b_2, \dots, b_L$, в которой символы $b/ = 1$, если $a/ = 1$ и 0 , если $a/ = 0$. Далее формируется новая последовательность $\{c/\} = c_1, c_2, \dots, c_L$, элементы $c/$ которой определяются из соотношения

$$c/ = -(a/ - a_{L+1-}); \quad / = 1, 2, \dots, L.$$

Заметим, что здесь $a_0 = a_L$. Наконец, запишем еще одну последовательность $\{d/\} = d_1, d_2, \dots, d_L$, элементы которой определяются по правилу $d/ = (b/ - c/) \bmod L$.

Поскольку элементы последовательности $\{d/\}$ принимают значения 1, 0, -1, а ввод отрицательной пробы не имеет физического смысла, модифицируем $d/$ следующим образом: $\tilde{d}/ = X/$. В результате будет сформирована искомая ПСТП $\{x/\} = x_1, x_2, \dots, x_L$, принимающие значения 0, 1, 2. Применительно к КХ это означает, что в конкретный момент времени в хроматограф вводится проба единичного объема, двойного объема или, если $X/ = 0$, проба не вводится.

Пример. Пусть $L = 7$, а последовательность Лежандра $\{a/\} = \{1, 1, 1, -1, -1, 1, -1\}$. Тогда $\{b/\} = \{1, 0, 0, -1, 0, 1, -1\}$ и, следовательно, $\{c/\} = \{1, 0, 0, -1, 0, -1, 1\}$. Окончательно, $\{x/\} = \{2, 1, 1, 0, 1, 0, 2\}$.

Пусть ПСТП длительностью L построена на основе М-последовательности и используется вариант б) формирования входного сигнала: с предысторией, длительностью m тактов и основной частью входной последовательности, занимающей временной интервал от 0 до $k > m$. Примем, что основная часть по своей длине совпадает в периодом ПСТП, равным Z , т.е. $k + 1 = Z - 1$.

Определим структуру матрицы A - X размера $\{(L - 1)Z\} \times \{(Z - 1)/2\}$. При выполнении дополнительного соотношения $x_{-}/ = x_{+}/$, $/ = 1, 2, \dots, t$, обеспечивающего повторения фрагмента основной части последовательности на интервале $\{-m, -1\}$, на главной диагонали матрицы $\tilde{A}$ будет величина $\frac{3L + 1}{2}$. В силу

специфических свойств ПСТП все суммы вида $\sum_{j=1}^{2^{n-2}} J \sim U_{2, \dots, m}$ одинаковы $\neq 0$ и равны L . Таким образом, в данном случае матрица A имеет вид

$$\bar{A} = \left\| \begin{array}{ccccc} \frac{3L+1}{2} & L & L & \dots & L \\ L & \frac{3L+1}{2} & L & \dots & L \\ L & L & \frac{3L+1}{2} & \dots & L \\ \vdots & & & \ddots & \vdots \\ L & L & L & \dots & \frac{3L+1}{2} \end{array} \right\|_m.$$

Найдем обратную матрицу $\bar{A}^{-1}$ для этого случая, в связи с чем предварительно определим значения определителя $|\bar{A}|$ и алгебраических дополнений $A[i, j]$:

$$|\bar{A}| = \left(\frac{3L+1}{2} + Lm \right) \left(\frac{L+1}{2} \right)^m, \quad (6)$$

$$A[i, i] = \left(\frac{3L+1}{2} + L(m-1) \right) \left(\frac{L+1}{2} \right)^{m-1}, \quad (7)$$

$$A[i, j] = -L \left(\frac{L+1}{2} \right)^{m-1}. \quad (8)$$

Принимая $(L-1)/2 = m+1$ и используя соотношения (6), (7) и (8), получаем, что диагональные элементы a_{ii}^{-1} матрицы $\bar{A}$ оказываются одинаковыми и могут быть вычислены по формуле

$$a_{ii}^{-1} = \frac{A[i, i]}{|\bar{A}|} = \frac{2}{L+1} \left[1 - \frac{2L}{L^2+1} \right], \quad i = 1, 2, \dots, (L-1)/2. \quad (9)$$

Все недиагональные элементы также равны между собой и могут быть вычислены по формуле

$$a_{ij}^{-1} = \frac{A[i, j]}{|\bar{A}|} = -\frac{4L}{(L+1)(L^2+1)}. \quad (10)$$

Следовательно, в данном случае задача получения усредненной хроматограммы $\bar{G}$ и анализа точности ее оценивания решается просто и фактически не требует выполнения операции обращения матрицы: все элементы матрицы $\bar{A}^{-1}$ могут быть

легко определены заранее с помощью выражений (9) и (10). В результате усредненная хроматограмма определяется соотношением

$$\bar{\tilde{G}} = \frac{2}{L+1} \left[\bar{I} - \left(\frac{2L}{L^2+1} \right) \bar{T} \right] \bar{X}^T \bar{Y}, \quad (11)$$

где $\bar{I}$ - единичная матрица размера $((L-1)/2 \times (B-1)/2)$, $\bar{f}$ - матрица того же размера, все элементы которой равны единице.

Для сравнения запишем матрицу $A = X^T X$ размера $L \times L$ в случае применения ПСБП. Пусть также используется вариант б) формирования входного сигнала. Основная часть по своей длине совпадает с периодом ПСБП, равным $L = 2^r - 1$, предыстория $X_i = X_{n-i}, i, i = 1, \dots, r$ обеспечивает повторение основной части последовательности на интервале $(-r, 1)$. В данном случае элементы главной диагонали матрицы A равны числу единиц в основной части ПСБП, т.е. $a_{ii} = (L+1)/2, i = 1, \dots, L$; все остальные элементы равны $(L+1)/4$.

Вычисляя определитель и алгебраически дополнения этой матрицы аналогично рассмотренному выше, найдем элементы обратной матрицы A^{-1} . В результате получим

$$a_{ii}^{-1} = 4L / (L+1), \quad i = 1, 2, \dots, L;$$

$$a_{ij}^{-1} = -4 / (L+1), \quad i, j = 1, 2, \dots, L.$$

Таким образом, в данном случае усредненная хроматограмма будет определяться соотношением

$$\bar{G} = \frac{1}{(L+1)^{2L}} \bar{r} - \frac{1}{(L+1)^{2L}} \bar{X}^T \bar{Y}, \quad (12)$$

где $\bar{f}, \bar{T}$ - матрицы размерностью $L \times L$.

Оценим эффективность КХ при использовании различных входных последовательностей вводов проб. Как следует из (5), дисперсия $\sigma_{g_i}^2$ оценки любой ординаты

$g_{i,0} / = 0, 1, \dots, A-1$ в КХ одна и та же и составляет величину: $\sigma_{g_i}^2 = \frac{1}{L} \sigma_n^2$. Таким образом,

при использовании ПСТП

$$\frac{\sigma_{g_i}^2}{\sigma_n^2} \cong \frac{2}{L}, \quad (13)$$

при использовании ПСБП

$$\frac{\sigma_{g_i}^2}{\sigma_n^2} \cong \frac{4}{L}. \quad (14)$$

Произведем сопоставление эффективности КХ по сравнению с классическим вариантом, когда усреднение производится по множеству проб, вводимых через интервал времени, не меньший, чем длительность хроматограммы от однократного ввода пробы. В качестве меры эффективности используем отношение дисперсий оценок ординат хроматограмм, полученных с помощью корреляционной и классической хроматографии [2].

Общее время эксперимента T_* для КХ с использованием ПСТП:

$$T_* = (2^n - 1) + m.$$

Принимая для простоты, что $m = 2^q$, имеем:

$$T_* = 2^n (2^q + 1),$$

где $q = n - ju$, ju - целое число. Значение q характеризует величину превышения основной части входной последовательности над длительностью хроматограммы и q принимает значения 1, 2, ...

При фиксированном $T_* = 2^n [2^q + 1]$ в классической хроматографии через интервал $m = 2^q$ может быть введено 2^q проб:

$$p = \frac{2^n}{m} = 2^{n-q} + 1.$$

Следовательно, для классической хроматографии

$$\sigma_{g,1}^2(\text{кл}) = \frac{\sigma_n^2}{2^q + 1}.$$

Для КХ (с использованием ПСТП) из соотношения (13), принимая, что $n = q + ju$, находим

$$S_i = \frac{2^{ju}}{2^{ju} + 1}$$

Введем показатель эффективности ξ , равный отношению $\frac{1}{\sigma_{g,1}^2(\text{кл})} \cdot \frac{1}{\sigma_n^2}$.

$$\xi = \frac{2^{ju} + 1}{2^{ju}}$$

где $\xi_0 = 2^{1/3} / 3$ - показатель эффективности при $q=1$ (минимально допустимой длительности). ξQ зависит только от длины m хроматограммы. При $m = 1024 - m = 4096$ ξ_0 принимает значения 340 - г 1365. На практике этот показатель может принимать и большие значения, что свидетельствует о высокой эффективности КХ. С ростом q возрастание ξ незначительно: $\lim_{q \rightarrow \infty} \xi = 1,5 \xi_0$. Поэтому нецелесообразно увеличивать время экспериментирования по сравнению с минимально

возможным $T_s = 3\text{ м}$, поскольку при этом количество вводов проб растет очень быстро без соответствующего роста показателя эффективности E .

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Современные методы идентификации систем / Под ред. П. Эйхгоффа. - М.: Мир, 1983. - 400 с.
2. Филаретов Н.Ф., Сайфулин Ж.Т. Корреляционный метод измерений в системах автоматизации хроматографического эксперимента // Заводская лаборатория. - 1991. - №10. - С. 59-63.
3. Блейхут Р. Быстрые алгоритмы цифровой обработки сигналов. - М.: Мир, 1989. - 448 с.
4. ГронД. Методы идентификации систем. - М.: Мир, 1979. - 302 с.
5. Свєрдлик М.Б. Оптимальные дискретные сигналы. - М.: Сов. радио, 1975. - 200 С.

Статья поступила в редакцию 26 мая 2009 г.

UDK 681.391:543/545

USE OF THE CORRELATION METHOD FOR AVERAGING OF CHROMATOGRAMS

R. Sayfullin

Samara State Technical University,
244, Molodogvardeyskaya str., Samara, 443100

We consider the receipt problem of chromatogram, which is averaged on some number of inputs. Estimation algorithm of the averaged chromatogram comes to solution of linear algebraic combined equations. By certain variants of input series' formation the information matrix turns into Tepliz's matrix, that simplifies its conversion. In correlation chromatography (CC) the injection is made according to random principle. Input actions, which may be used in CC, are tested.

Keywords: averaged chromatography, correlation chromatography, random injection.