

УДК 621.391; 519.21; 519.245

© Г. Ф. Малыхина, А. В. Меркушева

АДАПТИВНЫЕ МЕТОДЫ И АЛГОРИТМЫ РАЗДЕЛЕНИЯ СМЕСИ СИГНАЛОВ С НЕЗАВИСИМЫМИ КОМПОНЕНТАМИ

Рассмотрены методы и алгоритмы, предназначенные для разделения смеси сигналов из независимых компонент (первичных сигналов, образующих смесь). Реализация методов ориентирована на нейронные сети. Приведены модели нейронных сетей, правила для адаптивного метода их обучения.

Кл. сл.: смеси сигналов, первичные компоненты, независимость, адаптивные методы, алгоритмы

ВВЕДЕНИЕ

Основы методов разделения смеси сигналов (РСС) и восстановления первичных сигналов (ПС), входящих в состав смеси с неизвестным составом, проанализированы в [1–5]. Рассмотрено несколько методов РСС: декорреляция сигналов и использование статистик второго порядка; последовательное выделение ПС; извлечение ПС из смеси типа свертки; использование критерия независимости и вида распределения сигналов, а также модифицированного градиента в алгоритме РСС [4–9].

Многообразие подходов к РСС делает целесообразным анализ системы адаптивных алгоритмов для нейронных сетей (НС), используемых в этой задаче. Алгоритмы имеют преимущество по эффективности, простоте и форме реализации. Некоторые из них удобны для РСС в случае, когда анализируемые сигналы зашумлены, существенно различаются по величине или когда количество ПС и их статистические свойства неизвестны.

ОСНОВЫ ИСПОЛЬЗОВАНИЯ НС ДЛЯ РСС

Для нейронной сети (НС) с рекуррентной структурой и при линейном смешивании ПС адаптивное правило обучения использовано Коммоном, Джуттен и Хераулт (Common, Jutten, Herault) [10]. Сигналы $x_i(t)$ на сенсорах информационно-измерительной сети (ИИС) определяются соотношением

$$x_i(t) = \sum_j^n h_{ij} s_j(t), \quad i = 1, 2, \dots, n, \quad (1)$$

где

h_{ij} — неизвестные параметры; $s_j(t)$ — неизвестные ПС, для которых предполагается независимость¹⁾ и нулевое среднее. Рекуррентная НС (рис. 1) описывается соотношениями, имеющими (в векторно-матричной форме) вид

$$\mathbf{y}(t) = \mathbf{x}(t) - \hat{\mathbf{W}}(t)\mathbf{y}(t), \quad \mathbf{x}(t) = \mathbf{H}\mathbf{s}(t). \quad (2)$$

Из (2) следует, что $\mathbf{y}(t) = [\mathbf{I} + \hat{\mathbf{W}}(t)]^{-1} \mathbf{x}(t)$,

$$\text{где } \hat{\mathbf{W}} = \begin{bmatrix} 0 & \hat{w}_{12} & \dots & \hat{w}_{1n} \\ \hat{w}_{21} & 0 & \dots & \hat{w}_{2n} \\ & & \ddots & \\ \vdots & \vdots & \ddots & \vdots \\ \hat{w}_{n1} & \hat{w}_{n2} & \dots & 0 \end{bmatrix}.$$

При этом параметры сети $\hat{w}_{ij}$ (при $\hat{w}_{ii} = 0$) адаптируются так, чтобы наблюдаемые сигналы $x_i(t)$ образовали оптимальную оценку ПС: $\hat{s}_j(t) = y_j(t) = x_j - \sum_{i=1}^n \hat{w}_{ij} y_i(t)$. Оптимальные веса соответствуют статистической независимости сигналов $y_j(t)$ на выходе НС.

¹⁾ Единственной априорной информацией является независимость ПС. Это означает, что плотности распределения вероятности $p(y_i, y_j)$ сигналов y_i и y_j являются мультипликативными: $p(y_i, y_j) = p(y_i) \cdot p(y_j)$, т. е. все ковариации $E\{f(y_i)g(y_j)\} - E\{f(y_i)\} \cdot E\{g(y_j)\}$ равны нулю, а все дисперсии $E\{f(y_i)g(y_i)\} - E\{f(y_i)\} \cdot E\{g(y_i)\}$ — не нулевые.

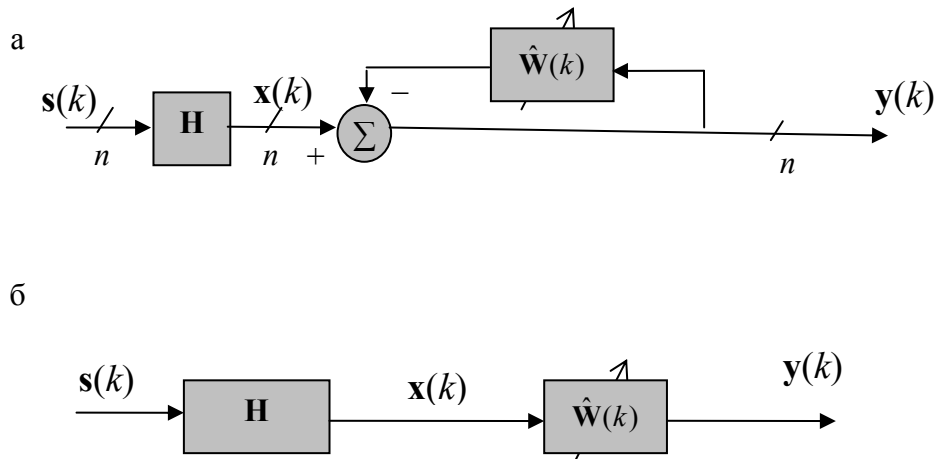


Рис. 1. Блок-схемы рекуррентной НС (а) и НС с прямым распространением сигнала (б)

При нулевых средних значениях ПС обобщенная корреляционная матрица для $f(y_i)$ и $g(y_i)$ (различных передаточных функций нейронов (ПФН) сети) является диагональной матрицей:

$$\mathbf{R}_{fg} = E\{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})\} - E\{\mathbf{f}(\mathbf{y})\} \cdot E\{\mathbf{g}^T(\mathbf{y})\} =$$

$$= \begin{bmatrix} E\{f(y_1)g(y_1)\} - E\{f(y_1)\} \cdot E\{g(y_1)\} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & E\{f(y_n)g(y_n)\} - E\{f(y_n)\} \cdot E\{g(y_n)\} \end{bmatrix}. \quad (3)$$

Следует отметить, что если плотность распределения каждого ПС является четной²⁾, то для нечетных функций $f(\cdot)$ и $g(\cdot)$ элементы типа $E\{f(y_i)\} \cdot E\{g(y_j)\}$ в ковариационной матрице (3) равны нулю. Вместе с тем общим условием для статистической независимости сигналов являются равные нулю (или очень малые) значения кросс-кумулянт высокого порядка.³⁾

²⁾ Четная функция $f(x)$ не изменяет своего значения при замене знака аргумента (т. е. $f(-x) = f(x)$), а нечетная функция при этом изменяет знак: $f(-x) = -f(x)$.

³⁾ Кумулянты являются статистиками высокого порядка, которые формализованно определяются как коэффициенты разложения в ряд Тейлора совместной характеристической функции произвольного набора вероятностных переменных, в том числе временных отсчетов (по крайней мере локально) стационарного случайного сигнала [11]. Так, для набора переменных $\{z_1, z_2, \dots, z_n\}$ их (общая) кумулянта k -го порядка ($k = k_1 + k_2 + \dots + k_n$) определяется соотношением

На основе критерия независимости ПС предложено адаптивное правило обучения НС [10]:

$$\frac{d\hat{w}_{ij}(t)}{dt} = \mu(t) f(y_i(t)) \cdot g(y_j(t)), \quad (4)$$

$$\text{Кум.} [z_1^{k_1}, z_2^{k_2}, \dots, z_n^{k_n}] \equiv$$

$$\equiv (-j)^k \frac{\partial^k \psi(\omega_1, \omega_2, \dots, \omega_n)}{\partial z_1^{k_1} \partial z_2^{k_2} \dots \partial z_n^{k_n}} \Big|_{\omega_1 = \omega_2 = \dots = \omega_n = 0},$$

где

$\psi(\omega_1, \omega_2, \dots, \omega_n) \equiv E\{\exp[j \cdot (\omega_1 z_1 + \omega_2 z_2 + \dots + \omega_n z_n)]\}$ — характеристическая функция и E — символ математического ожидания. Для набора n вероятностных (случайных) стационарных переменных $\{z_i(t)\}_{i=1,2,\dots,n}$ кросс-кумулянта (С) n -го порядка представляется в виде

$$C_{z_1 z_2 \dots z_n}(\tau_1, \tau_2, \dots, \tau_{n-1}) \equiv$$

$$\equiv \text{Кум.} [z_1(t), z_2(t + \tau_1), z_3(t + \tau_2), \dots, z_n(t + \tau_{n-1})].$$

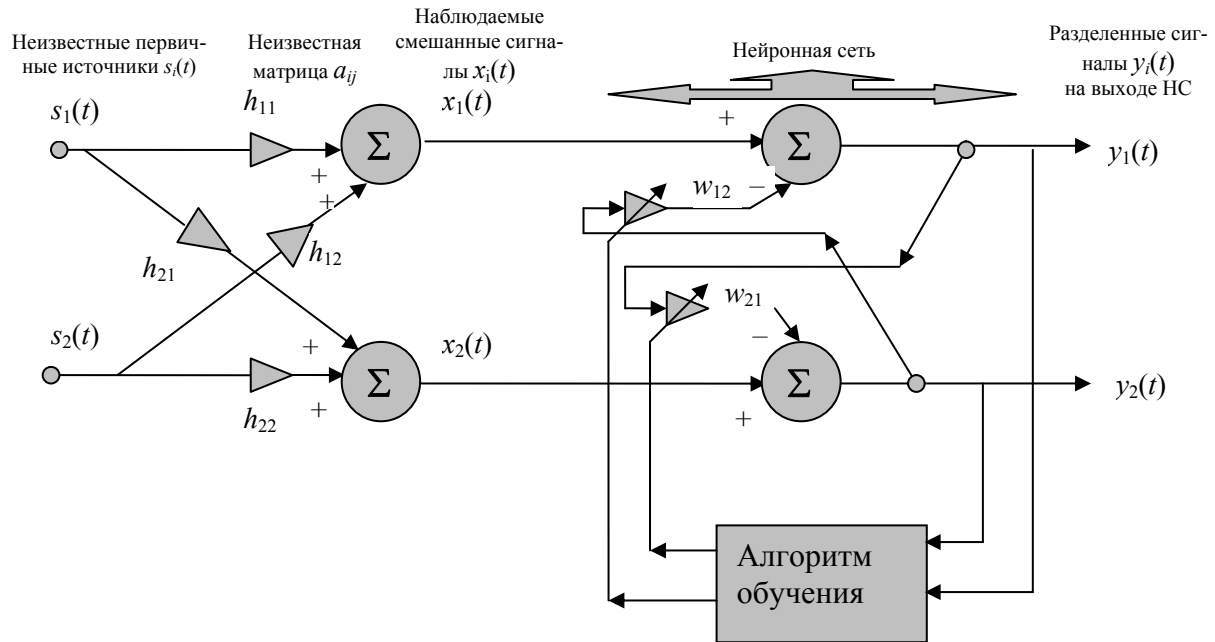


Рис. 2а. Структура операций при восстановлении сигналов на основе НС [8]

где $f(y)$ и $g(y)$ — нечетные ФПН вида $f(y) = y^3$ и $g(y) = \text{tg}(10y)$ для ПС с субгауссовым распределением и $f(y) = \text{tg}(10y)$ и $g(y) = y^3$ — для ПС с надгауссовым распределением ПС.⁴⁾

Алгоритм обучения в векторно-матричной форме имеет вид:

$$\frac{d\hat{\mathbf{W}}(t)}{dt} = \mu(t)\mathbf{f}[\mathbf{y}(t)]\mathbf{g}^T[\mathbf{y}(t)], \quad (5)$$

где

$$\mathbf{f}(\mathbf{y}) = [f(y_1), f(y_2), \dots, f(y_n)]^T, \\ \mathbf{g}(\mathbf{y}) = [g(y_1), g(y_2), \dots, g(y_n)]^T$$

при условии, что $E\{\mathbf{f}(\mathbf{y})\} = 0$ или $E\{\mathbf{g}(\mathbf{y})\} = 0$. Чтобы удовлетворить этим условиям (для несимметричных распределений), используются нелинейные ФПН в виде $f(y_i) = \varphi(y_i)$, $g(y_i) = y_i$, и вид функции $\varphi(y)$ подбирают по виду предполагаемого распределения ПС. Схема РСС на простой структуре НС при двух ПС (s_1 и s_2) и функциональная схема алгоритма для нее показаны на рис. 2а и 2б.

⁴⁾ Субгауссова и надгауссова — это соответственно "сплюснутая" плотность распределения с модой меньше, чем у гауссовой, и плотность распределения с более острой вершиной (т. е. с большей модой), чем у гауссовой плотности распределения.

Соотношение (3) отражает практический критерий независимости выходных сигналов НС. При плохо масштабированных сигналах (когда величины ПС отличаются во много раз) этот критерий оказывается недостаточно эффективным. В этом случае используется дополнительное условие для нормализации дисперсий: $E\{f(y_i)g(y_i)\} = \lambda_i$ (часто принимают все $\lambda_i = 1$).

Модель [10] улучшена использованием рекуррентной структуры НС, где каждый нейрон имеет обратную связь на себя ($w_{ii} \neq 0$) и на все остальные нейроны (рис. 1, а). Эта сеть описывается соотношением $y_i(t) = x_i(t) - \sum_{j=1}^n \hat{w}_{ij}(t)y_j(t)$. Критерий независимости и нормализация дисперсий позволяют получить алгоритм обучения НС в реальном времени

$$\frac{d\hat{w}_{ii}(t)}{dt} = \mu(t)[\lambda_i - f[y_i(t)] \cdot g[y_i(t)]], \\ \frac{d\hat{w}_{ij}(t)}{dt} = \mu(t)f[y_i(t)] \cdot g[y_j(t)] \quad \text{для } i \neq j. \quad (6)$$

и представить его в векторно-матричной форме:

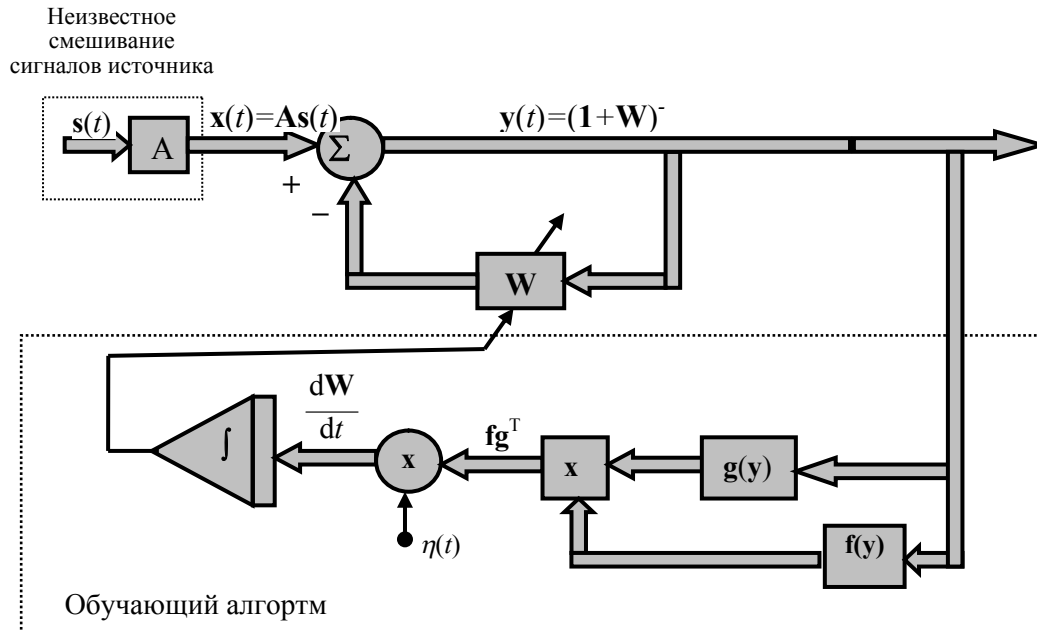


Рис. 26. Функциональная схема алгоритма РСС для НС по модели [10]

$$\frac{d\hat{\mathbf{W}}(t)}{dt} = \mu(t) [\mathbf{\Lambda} - \mathbf{f}[\mathbf{y}(t)]\mathbf{g}^T[\mathbf{y}(t)]], \quad (7)$$

где $\mathbf{\Lambda} = \text{Диаг.}\{\lambda_1, \lambda_2, \dots, \lambda_2\}$.

Условие нормализации дисперсий

$$(E\{f(y_i)g(y_i)\} = 1)$$

существенно улучшает функционирование НС, особенно при плохом масштабировании ПС, когда смешивающая матрица $\mathbf{H}$ имеет очень большой разброс собственных значений, т. е. близка к сингулярной.

Обучающий алгоритм (6), (7) эффективен для хорошо обусловленных задач РСС, однако НС, показанная на рис. 1, а, и алгоритм ее обучения требуют на каждой итерации обращения матрицы $\hat{\mathbf{W}}(t)$ в (2). Кроме того, предусматривается равенство числа ПС и количества сенсоров.

НС С ПРЯМЫМ РАСПРОСТРАНЕНИЕМ СИГНАЛА И АЛГОРИТМЫ ОБУЧЕНИЯ

Рассмотрим восстановление формы ПС, используя НС с прямым распространением сигнала (НС без обратных связей), рис. 1, б. При этом тре-

буется найти веса w_j сети, для которых она определяет оптимальную оценку ПС:

$$\hat{s}_j = y_j(t) = \sum_{i=1}^n w_{ji} x_i(t), \quad (j=1, 2, \dots, n). \quad (8)$$

При РСС реализуются преобразования $\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t)$ и $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) = \mathbf{W} \cdot \mathbf{H}\mathbf{s}(t)$, где $\mathbf{H} = [h]_{n \times n}$ — неизвестная $(n \times n)$ -матрица; $\mathbf{W} = [w]_{n \times n}$ — $(n \times n)$ -матрица адаптивных параметров (синаптических весов) НС; $\mathbf{s}(t) = [s_1(t), \dots, s_n(t)]^T$ — неизвестный вектор независимых ПС; $\mathbf{x}(t) = [x_1(t), \dots, x_n(t)]^T$ — вектор (наблюдаемых) сигналов на сенсорах ИИС; $\mathbf{y}(t) = [y_1(t), \dots, y_n(t)]^T$ — вектор желаемых сигналов на выходе НС. Компоненты $\mathbf{y}(t)$ должны быть независимыми, и это достигается при оптимальной величине параметров НС. РСС реализуется, когда составная матрица $\mathbf{G}(t) = \mathbf{W}(t)\mathbf{H}$ содержит только по одному элементу в каждой строке и столбце, т. е. является матрицей перестановок.

НС с прямым распространением сигнала (НС_ПРС) (рис. 1, б) эквивалентна рекуррентной сети при выполнении условия $\mathbf{W}(t) = [\mathbf{I} + \hat{\mathbf{W}}(t)]^{-1}$, или равноценного условия

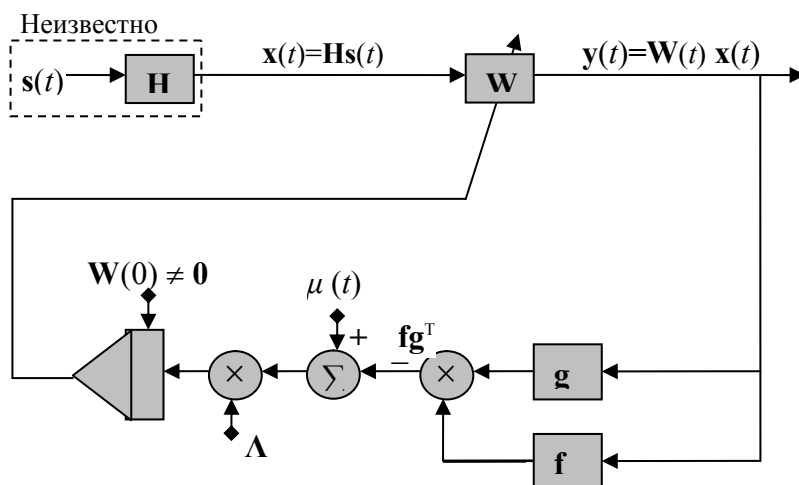


Рис. 3. Функциональная блок-схема локально адаптивного алгоритма (10) для НС_ПРС

$$\hat{\mathbf{W}}(t) = \mathbf{W}(t)^{-1} - \mathbf{I}^{.5)}$$

Дифференцирование $\hat{\mathbf{W}}(t)$ дает соотношение

$$\frac{d\hat{\mathbf{W}}(t)}{dt} = \frac{d}{dt}(\mathbf{W}^{-1} - \mathbf{I}) = -\mathbf{W}^{-1} \frac{d\mathbf{W}}{dt} \mathbf{W}^{-1},$$

или с учетом (7)

$$\mathbf{W}^{-1} \frac{d\mathbf{W}}{dt} \mathbf{W}^{-1} = \mu(t) [\Lambda - \mathbf{f}[\mathbf{y}(t)] \mathbf{g}^T[\mathbf{y}(t)]].$$

Отсюда следует соотношение (9), для которого используется начальное условие $\mathbf{W}(0) = \mathbf{I}$:

$$\frac{d\mathbf{W}}{dt} = \mu(t) \mathbf{W}(t) [\Lambda - \mathbf{f}[\mathbf{y}(t)] \mathbf{g}^T[\mathbf{y}(t)]] \mathbf{W}(t). \quad (9)$$

Соотношение (9) представляет алгоритм обучения НС_ПРС. Свойства его сходимости такие же, как у алгоритма (7), а сложность незначительно выше. Показано, что множители $\mathbf{W}(t)$ не влияют на точку равновесия НС, а только на "траекторию" обучения, ведущую к ней. Поэтому (9) можно использовать в упрощенном виде, где единичная матрица $\mathbf{I}$ заменяет $\mathbf{W}(t)$:

$$\frac{d\mathbf{W}}{dt} = \mu(t) [\Lambda - \mathbf{f}[\mathbf{y}(t)] \mathbf{g}^T[\mathbf{y}(t)]], \quad (10)$$

или в скалярной форме

⁵⁾ Эти условия требуют, чтобы матрицы

$$\hat{\mathbf{W}}(t) + \mathbf{I} \text{ и } \mathbf{W}(t)$$

не были сингулярными (их определители $\neq 0$).

$$\frac{dw_{ij}(t)}{dt} = \mu(t) [\delta_{ij} \lambda_i - f[y_i(t)] g[y_i(t)]]$$

(δ_{ij} — символ Кронекера: $\delta_{ij} = 1$ только при $i = j$, иначе $\delta_{ij} = 0$). Функциональная схема алгоритма (10) показана на рис. 3 и в более детальной форме — на рисунке в Приложении. Этот алгоритм хорошо работает при отношении энергий ПС, не более чем 1 : 10, т. е. при среднем уровне плохой обусловленности данных.

При плохой обусловленности задачи РСС может применяться метод на основе модификации алгоритма (9). При этом вместо скалярной скорости обучения $\mu(t)$ используется матрица обучения $\mu_0(t) \mathbf{W}^{-1}(t)$, где $\mu_0(t)$ — скаляр. Алгоритм обучения определяет правило обновления параметров НС_ПРС в виде

$$\frac{d\mathbf{W}(t)}{dt} = \mu_0(t) [\Lambda - \mathbf{f}[\mathbf{y}(t)] \mathbf{g}^T[\mathbf{y}(t)]] \mathbf{W}(t). \quad (11)$$

Скалярный вид алгоритма (11)

$$\begin{aligned} \frac{dw_{ij}(t)}{dt} &= \\ &= \mu_0(t) \left[\delta_{ij} \lambda_j - f[y_i(t)] \sum_{p=1}^m w_{pj}(t) g[y_p(t)] \right] \end{aligned} \quad (12)$$

соответствует дискретной форме этого алгоритма, представленной соотношением

$$w_{ij}(k+1) = w_{ij}(k) +$$

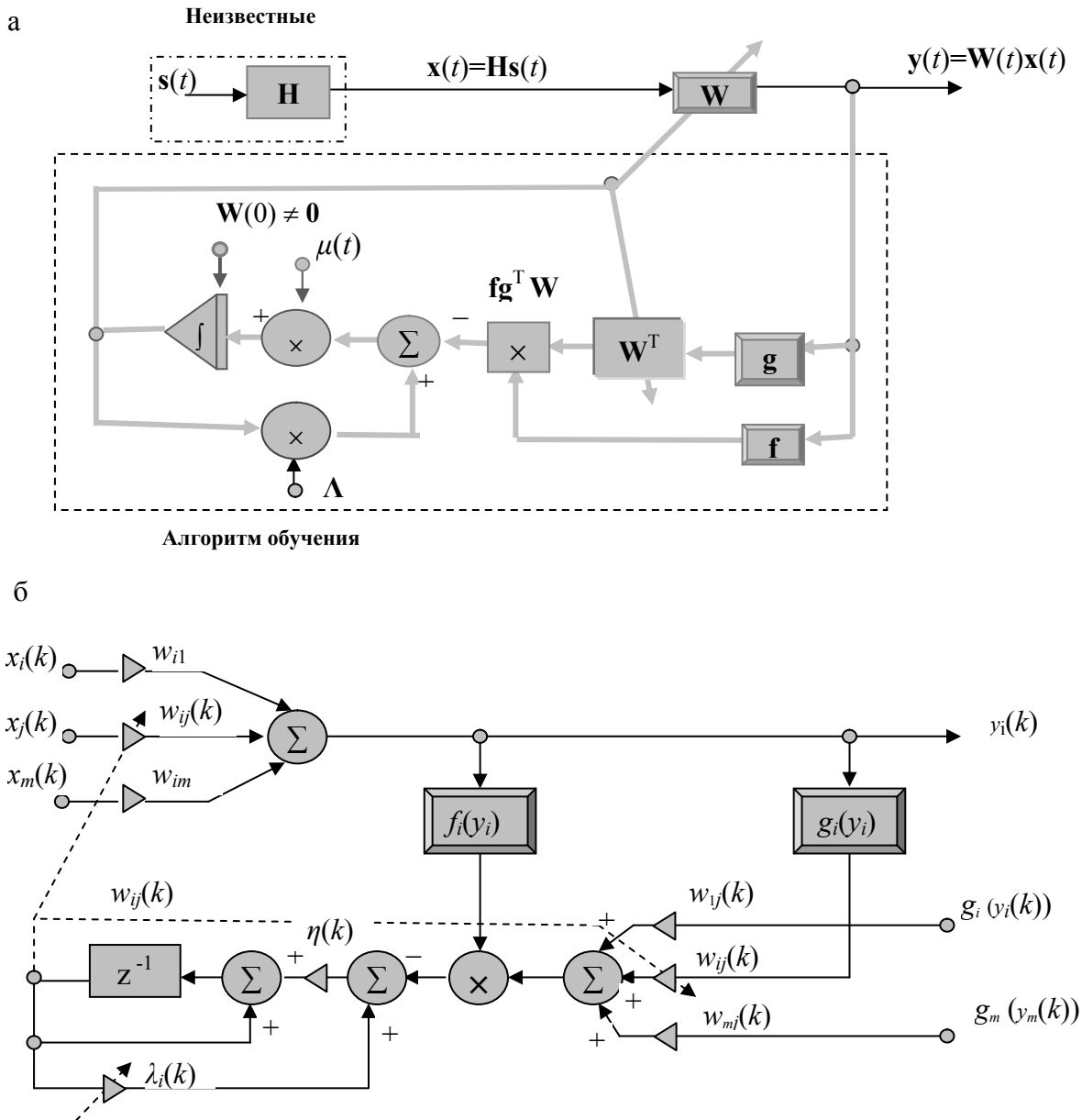


Рис. 4. Функциональная схема алгоритма (11) обучения НС_ПРС.
а — для непрерывного времени, б — для дискретного времени

$$+\eta(k) \left[\delta_{ij} \lambda_j - f[y_i(k)] \sum_{p=1}^m w_{pj}(k) g[y_p(k)] \right].^{6)} \quad (13)$$

⁶⁾ Если дополнительно принять, что предварительно осуществлено преобразование отбеливания сигналов на сенсорах $\mathbf{x}_1 = \mathbf{Q}\mathbf{x}$ (после этого $\mathbf{W}^T \mathbf{W} = \mathbf{I}$), то получается нелинейное РСС в виде $\Delta \mathbf{W} = \eta \mathbf{f}(\mathbf{y}) [\mathbf{x}_1^T - \mathbf{f}^T(\mathbf{y}) \mathbf{W}]$. При допущении, что в процессе обучения НС выполняется ограничение $\mathbf{W}^T \mathbf{W} = \mathbf{I}$, этот алгоритм приближенно соответствует правилу Кардосо и Лахельда (Cardoso, Laheld) [12].

Функциональная схема алгоритма (11) показана на рис. 4, а, а ее дискретный вариант на рис. 4, б.

МОДИФИКАЦИЯ МЕТОДА РСС НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ МНОГОСЛОЙНОЙ НС

Улучшению гибкости и эффективности метода РСС способствует применением многослойной НС (МНС). Рост эффективности обеспечивается, в частности, разнообразием алгоритмов и формой

ФПН в каждом из слоев НС. Кроме того, для осуществления РСС и последовательного извлечения на разных слоях отдельных ПС, которые различаются статистическими свойствами, МНС позволяет комбинировать метод статистик второго порядка с методом статистик высокого порядка [11, 13]. Последовательные слои отчасти независимы: после сходимости параметров первого слоя может происходить настройка параметров второго.

Для структуры МНС, включающей (условно) K слоев, обновление параметров в слое p осуществляется локальным обучающим правилом (14). Его скалярная форма определяется соотношением (15):

$$\begin{aligned} \frac{d\mathbf{W}^{(p)}(t)}{dt} &= \\ &= \mu(t) \left[\Lambda^{(p)} - \left\langle \mathbf{f}_p[\mathbf{y}^{(p)}(t)] \mathbf{g}_p^T[\mathbf{y}^{(p)}(t)] \right\rangle \right] \quad (14) \\ (p=1, 2, \dots, K); \end{aligned}$$

$$\begin{aligned} \frac{dw_{ij}^{(p)}(t)}{dt} &= \\ &= \mu(t) \cdot \left[\delta_{ij} \lambda_i^{(p)} - \left\langle f_p[y_i^{(p)}(t)] \cdot g_p[y_i^{(p)}(t)] \right\rangle \right], \quad (15) \end{aligned}$$

где $\Lambda^{(p)}$ — диагональная матрица (обычно $\Lambda^{(p)} = \mathbf{I}$), а скобки $\langle \cdot \rangle$ означают оператор статистического усреднения.

Локально-адаптивное обучающее правило дискретного времени, связанное с (14) и (15), имеет вид (соответственно в векторно-матричной и скалярной формах):

$$\begin{aligned} \mathbf{W}^{(p)}(k+1) &= \mathbf{W}^{(p)}(k) + \\ &+ \eta^{(p)}(k) \left[\Lambda^{(p)} - \left\langle \mathbf{f}_p[\mathbf{y}^{(p)}(k)] \cdot \mathbf{g}_p^T[\mathbf{y}^{(p)}(k)] \right\rangle \right] \quad (16) \\ (p=1, 2, \dots, K); \end{aligned}$$

$$\begin{aligned} w_{ij}^{(p)}(k+1) &= w_{ij}^{(p)}(k) + \\ &+ \eta^{(p)}(k) \left[\delta_{ij} \lambda_i^{(p)} - \left\langle f_p[y_i^{(p)}(k)] \cdot g_p[y_i^{(p)}(k)] \right\rangle \right]. \quad (17) \end{aligned}$$

Использование МНС обеспечивает лучшие результаты (сравнительно с простой структурой НС) при плохо обусловленной задаче РСС, т. е. при плохом масштабировании ПС и слишком большом разбросе собственных чисел смешивающей матрицы $\mathbf{H}$.

АЛГОРИТМЫ ДЛЯ АНАЛИЗА РСС, УСТОЙЧИВЫЕ К ШУМУ

Проанализированные выше алгоритмы несмотря на простоту и хорошее функционирование основаны на интуитивно-эвристическом подходе. Формализованный подход к получению робастных (устойчивых к шуму) алгоритмов анализа независимых компонент основан на методе итеративного обращения обобщенной ковариационной матрицы: $\mathbf{R}_{fg}(\mathbf{W}) = E\{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})\}$.

Элементы итеративного метода обращения матрицы ковариаций [7, 13, 14]

Допускается, что при оптимальной разделяющей матрице $\mathbf{W}_*$ ковариационная матрица равна единичной, т. е. $\mathbf{R}_{fg}(\mathbf{W}) = \mathbf{I}$. Диагонализация может рассматриваться как неявное обращение матрицы. Так, если определить новую нелинейную функцию выходов системы РСС $\mathbf{h}(\mathbf{y}) = \mathbf{W}^{-1}\mathbf{f}(\mathbf{y})$ и новую обобщенную матрицу ковариаций $\mathbf{R}_{hg}(\mathbf{W}) = E\{\mathbf{h}(\mathbf{y})\mathbf{g}^T(\mathbf{y})\} = \mathbf{W}^{-1}\mathbf{R}_{fg}(\mathbf{W}) = \mathbf{H}\mathbf{R}_{fg}(\mathbf{W})$, то диагонализация $\mathbf{R}_{fg}(\mathbf{W})$ эквивалентна следующему обращению: $\mathbf{R}_{fg}(\mathbf{W}) = \mathbf{I} \Rightarrow \mathbf{W}_* = \mathbf{R}_{hg}^{-1}(\mathbf{W}_*)$. Неявное обращение $\mathbf{R}_{hg}^{-1}(\mathbf{W}_*)$ не может быть выполнено непосредственно из-за незнания $\mathbf{h}(\mathbf{y})$ и $\mathbf{g}(\mathbf{y})$ при оптимальной разделяющей матрице $\mathbf{W}_*$. Вместо этого обращение выполняется итеративно: для разделяющей матрицы $\mathbf{W}(l)$ на l -м шаге итерации принимается, что $\hat{\mathbf{R}}_{hg}^{(l)} = \hat{\mathbf{R}}_{hg}(\mathbf{W}(l))$. Тогда более удобно явно оценить смешивающую матрицу $\hat{\mathbf{H}}(l) = \mathbf{W}^{-1}(l)$, чем разделяющую матрицу $\mathbf{W}(l)$. Это достигается минимизацией (изменяющейся со временем) функции стоимости (ФС), определенной выражением

$$\Phi(\hat{\mathbf{H}}(l+1)) = \|\hat{\mathbf{R}}_{hg}^{(l)} - \hat{\mathbf{H}}(l)\|^2, \quad (18)$$

где $\|\cdot\|$ — символ нормы матрицы.

Таким образом, используется процедура с двумя фазами: "назад" и "вперед". Оценка $\hat{\mathbf{H}}(l+1)$ на основе минимизации (18) используется затем для оценки новой ковариационной матрицы и построения обновленной функции стоимости $\Phi(\hat{\mathbf{H}}(l+2))$, а минимизация $\Phi(\hat{\mathbf{H}}(l+2))$ дает $\hat{\mathbf{H}}(l+2)$ и т. д.

$\hat{\mathbf{H}}(l) = \mathbf{W}^{-1}(l)$ входит в функцию стоимости $\Phi(\hat{\mathbf{H}}(l+1))$ квадратично, поэтому минимум ФС определяется итеративным соотношением

⁷⁾ Обучающее правило (17) может использоваться как для МНС_ПРС, так и для элементов (субмоделей) НС с рекуррентной структурой.

$\hat{\mathbf{H}}(l+1) = \hat{\mathbf{H}}(l) + \eta(\hat{\mathbf{R}}_{hg}^{(l)} - \hat{\mathbf{H}}(l))$, которое может также быть представлено в виде

$$\begin{aligned} \hat{\mathbf{H}}(l+1) &= \\ &= (1-\eta)\hat{\mathbf{H}}(l) + \eta\hat{\mathbf{R}}_{hg}^{(l)} = \hat{\mathbf{H}}(l)\left(\mathbf{I} - \eta(\mathbf{I} - \hat{\mathbf{R}}_{fg}^{(l)})\right). \end{aligned} \quad (19)$$

Из второй части соотношения (19) следует, что $\hat{\mathbf{H}}(l+1) = \mathbf{W}^{-1}(l+1)$ может интерпретироваться как оценка $\hat{\mathbf{R}}_{hg}$. (Это согласуется с отмеченной выше оценкой $\hat{\mathbf{R}}_{hg}$ для смешивающей матрицы $\mathbf{H}$). В алгоритме (19) можно использовать смешивающую матрицу $\mathbf{W}(l)$ (а не $\hat{\mathbf{H}}(l) = \mathbf{W}^{-1}(l)$). Обращение (19) (в предположении, что матрица $(\mathbf{I} - \eta(\mathbf{I} - \hat{\mathbf{R}}_{fg}^{(l)}))$ не сингулярная) приводит к соотношению

$$\mathbf{W}(l+1) = \left[\mathbf{I} - \eta(l)(\mathbf{I} - \hat{\mathbf{R}}_{fg}^{(l)})\right]^{-1} \mathbf{W}(l), \quad (20)$$

в котором

$$\begin{aligned} \hat{\mathbf{R}}_{fg} &= \left(\frac{1}{N}\right) \sum_{k=1}^N \{\mathbf{f}(\mathbf{y}^{(l)}(k)) \cdot \mathbf{g}^T(\mathbf{y}^{(l)}(k))\} \quad \text{и} \\ \mathbf{y}^{(l)}(k) &= \mathbf{W}(l)\mathbf{x}(k). \end{aligned}$$

Устойчивость алгоритма (20) обеспечивается выбором малого значения параметра η скорости обучения НС [15]: $\eta(l) < \frac{1}{\|\hat{\mathbf{R}}_{fg}^{(l)} - \mathbf{I}\|}$.⁸⁾ При этом условия в соотношении (20) возможно обращение матрицы и, следовательно, представление этого обращения в виде ряда [16]:

$$\begin{aligned} \left[\mathbf{I} - \eta(l)(\mathbf{I} - \hat{\mathbf{R}}_{fg}^{(l)})\right]^{-1} &= \\ &= \sum_{i=0}^{\infty} \left(\eta(l)(\mathbf{I} - \hat{\mathbf{R}}_{fg}^{(l)})\right)^i \approx \mathbf{I} + \eta(l)(\mathbf{I} - \hat{\mathbf{R}}_{fg}^{(l)}). \end{aligned} \quad (21)$$

Использование (21) в (20) позволяет получить достаточно простой устойчивый алгоритм обучения:

$$\mathbf{W}(l+1) = \mathbf{W}(l) + \eta(\mathbf{I} - \hat{\mathbf{R}}_{hg}^{(l)})\mathbf{W}(l). \quad (22)$$

⁸⁾ Это условие влечет отсутствие сингулярности матрицы $\mathbf{I} - \eta(l)(\mathbf{I} - \hat{\mathbf{R}}_{fg}^{(l)})$ и возможность ее обращения.

⁹⁾ Применение итеративного варианта обращения $\hat{\mathbf{R}}_{hg}^{(l)}$ позволяет получить алгоритм эквивалентный (22) в виде $\mathbf{W}(l+1) = \mathbf{W}(l) + \eta(l)[\mathbf{I} - \mathbf{W}(l)\hat{\mathbf{R}}_{hg}^{(l)}]$. Аналогично с помощью итеративного обращения матриц получены дуальные формы адаптивных алгоритмов для оценки

Получение устойчивого алгоритма анализа независимых компонент на основе обобщенного м-градиента

Семейство алгоритмов для анализа независимых компонент (АНК) в несколько менее общей форме может быть получено при использовании обобщенной формы м-градиента (ОФМГ):

$$\Delta \mathbf{W} = -\eta \frac{\partial J(\mathbf{y}, \mathbf{W})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{D}_1(\mathbf{y}) \mathbf{W}, \quad (23)$$

где $J(\mathbf{y}, \mathbf{W})$ — выбранная функция стоимости, $\mathbf{D}_1(\mathbf{y})$ — масштабирующая положительно определенная диагональная матрица.

ОФМГ (так же как м-градиент) обеспечивает устойчивый градиентный спуск для поиска локального минимума функции стоимости (ФС). Так, если ФС имеет вид:

$$J(\mathbf{y}, \mathbf{W}) = -\log |\text{Det}(\mathbf{W})| - \sum_{i=1}^n E\{\log(q_i(y_i))\}, \quad (24)$$

то градиент этой ФС определяется соотношением

$$\frac{\partial J(\mathbf{y}, \mathbf{W})}{\partial \mathbf{W}} = -\mathbf{W}^T + E\{\mathbf{f}(\mathbf{y})\mathbf{x}^T\},$$

где

$$\begin{aligned} \mathbf{f}(\mathbf{y}) &= [f_1(y_1), f_2(y_2), \dots, f_n(y_n)]^T \\ \text{при } f_i(y_i) &= -d \log(q_i(y_i)) / dy_i. \end{aligned}$$

Следовательно, применение ОФМГ приводит к дискретной форме анализируемого выше обучающего алгоритма (см. выражение (11)):

$$\Delta \mathbf{W}(l) = \eta(l) \left[\Lambda(l) - \langle \mathbf{f}[\mathbf{y}(l)] \mathbf{g}^T[\mathbf{y}(l)] \rangle \right] \mathbf{W}(l),$$

где $\Lambda = \mathbf{D}_1(\mathbf{y})$ и $\mathbf{g}(\mathbf{y}) = \mathbf{y}^T \mathbf{D}_1(\mathbf{y}) = [g_1(y_1), g_2(y_2), \dots, g_n(y_n)]^T$.

В случае симметричной функции распределения ПС, нечетности ФПН ($f_i(y_i)$) и при $\mathbf{D}_1 = \text{Диаг.}\{\langle |y_1|^p \rangle, \langle |y_2|^p \rangle, \dots, \langle |y_n|^p \rangle\}$ получается медианное обучающее правило при $p = -1$:

$$\Delta \mathbf{W}(l) = \eta(l) \left[\Lambda(l) - \langle \mathbf{f}[\mathbf{y}(l)] \cdot [\text{sign}(\mathbf{y})^T] \rangle \right] \mathbf{W}(l), \quad (25)$$

смешивающей матрицы

$$\hat{\mathbf{H}}(k+1) = \hat{\mathbf{H}}(k) \left[\mathbf{I} - \eta(k)(\mathbf{I} - \hat{\mathbf{R}}_{fg}^{(k)}) \right] \quad \text{и}$$

$$\hat{\mathbf{H}}(k+1) = \hat{\mathbf{H}}(k) \left[\mathbf{I} + \eta(k)(\mathbf{I} - \hat{\mathbf{R}}_{fg}^{(k)}) \right]^{-1}$$

и для оценки разделяющей матрицы $\mathbf{W}$:

$$\mathbf{W}(k+1) = \left[\mathbf{I} + \eta(k)(\mathbf{I} - \hat{\mathbf{R}}_{fg}^{(k)}) \right] \mathbf{W}(k) \quad \text{и}$$

$$\mathbf{W}(k+1) = \left[\mathbf{I} - \eta(k)(\mathbf{I} - \hat{\mathbf{R}}_{fg}^{(k)}) \right]^{-1} \mathbf{W}(k),$$

где $\hat{\mathbf{R}}_{fg}^{(k)} = (1 - \eta_0) \hat{\mathbf{R}}_{fg}^{(k-1)} + \eta_0 \mathbf{f}(\mathbf{y}(k)) \cdot \mathbf{g}^T(\mathbf{y}(k))$ [17, 18].

где $\text{sign}(\mathbf{y}) = [\text{sign}(y_1), \text{sign}(y_2), \dots, \text{sign}(y_n)]^T$. Такой обучающий алгоритм с ФПН в виде знаковой функции вполне устойчив к аддитивному шуму.

Практическая форма реализации алгоритма

Обобщенная ковариационная матрица $\hat{\mathbf{R}}_{fg}^{(l)}$, полученная в робастном (т. е. устойчивом к шуму) обучающем алгоритме (19) и (22), может быть оценена как статистическое среднее на основе доступных (поступающих) выходных данных НС. Имеются две возможности.

– Адаптация в реальном времени, при которой матрица $\hat{\mathbf{R}}_{fg}^{(l)}$ заменяется своей выборочной оценкой (в РВ) для каждого дискретного момента времени, т. е. $\hat{\mathbf{R}}_{fg}^{(l)} \approx \mathbf{f}(\mathbf{y}^{(l)}(k)) \cdot \mathbf{g}^T(\mathbf{y}^{(l)}(k))$.

– Адаптация по группе данных, когда статистическое среднее заменяется скользящим средним $\hat{\mathbf{R}}_{fg}^{(k)} = (1 - \eta)\hat{\mathbf{R}}_{fg}^{(k-1)} + \eta \cdot \mathbf{f}(\mathbf{y}^{(l)}(k))\mathbf{g}^T(\mathbf{y}^{(l)}(k))$. При этом предполагается стационарность на блоке наблюдений из L отсчетов.

Значение параметра η выбирается так, чтобы обеспечить устойчивость алгоритма и одновременно достаточную скорость его сходимости. Для обучающего алгоритма в РВ (11) с оценкой ковариационной матрицы по текущему отсчету НС:

$$\mathbf{W}(k+1) = [\mathbf{I} - \eta(\mathbf{I} - \mathbf{f}(\mathbf{y}(k))\mathbf{g}^T(\mathbf{y}(k)))]^{-1} \mathbf{W}(k) \quad (26)$$

при условии, что $\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y}) < 0$, параметр скорости обучения $\eta < \frac{1}{1 + |\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})|}$.

При использовании $\eta_0 = \eta/(1 - \eta)$ соотношение (26) приобретает вид

$$\mathbf{W}(k+1) = (1 + \eta_0)[\mathbf{I} - \eta_0\mathbf{f}(\mathbf{y}(k))\mathbf{g}^T(\mathbf{y}(k))]^{-1} \mathbf{W}(k),$$

а применение матричной формы для обращения бинома приводит к выражению

$$\mathbf{W}(k+1) = (1 + \eta_0) \left(\mathbf{I} - \eta_0 \frac{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})}{1 + \eta_0\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})} \right) \mathbf{W}(k).$$

Поскольку при малых значениях η_0 можно пренебречь ее степенями, выше первой, то нормализованная форма алгоритма (19) выражается соотношением:

$$\mathbf{W}(k+1) = \left(\mathbf{I} - \eta_0 \frac{\mathbf{I} - \mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})}{1 + \eta_0\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})} \right) \mathbf{W}(k). \quad (27)$$

Выражение (27) может интерпретироваться как алгоритм (22) с самоадаптацией параметра скоро-

сти обучения $\eta(k) = \frac{\eta_0}{1 + \eta_0\mathbf{g}^T(\mathbf{y}(k))\mathbf{f}(\mathbf{y}(k))}$. Для параметра η используется и более осторожная оценка:

$$\eta(k) = \frac{\eta_0}{1 + \eta_0|\mathbf{g}^T(\mathbf{y}(k))\mathbf{f}(\mathbf{y}(k))|}. \quad (28)$$

■ **Робастный алгоритм обучения НС.** Соотношение (22), отражающее обучающий алгоритм НС, служит основой для ряда прикладных схем РСС, позволяет интерпретировать их как способ итеративного обращения обобщенной корреляционной матрицы и получать адаптивные алгоритмы РВ, а также применять обучение по группе отсчетов с выхода НС.

1. При выборе $\mathbf{f}(\mathbf{y}) = \mathbf{g}(\mathbf{y}) = \mathbf{y}$, $\mathbf{\Lambda} = \mathbf{I}$ и нормализованного параметра скорости обучения η по (28) адаптивный алгоритм (22) переходит в алгоритм декорреляции сигналов [7, 18]:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta_0 \frac{\mathbf{I} - \mathbf{y}\mathbf{y}^T}{1 + \eta_0\mathbf{y}^T\mathbf{y}} \mathbf{W}(k). \quad (29)$$

2. Если использовать динамическую оценку (адаптивного значения) η по (28) и положить $\mathbf{g}(\mathbf{y}) = \mathbf{y}$, то алгоритм (22) упрощается к выражению (30), которое является вариантом алгоритма на основе м-градиента:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta_0 \frac{\mathbf{I} - \mathbf{f}(\mathbf{y})\mathbf{y}^T}{1 + \eta_0\mathbf{y}^T\mathbf{f}(\mathbf{y})} \mathbf{W}(k). \quad (30)$$

Более полная характеристика его содержится в [8, 15] и [19].

3. Адаптивное правило обучения, представленное соотношением (22), порождает группу прикладных алгоритмов расширенного анализа смеси и идентификации ПС.

В процедуре разделения смеси, реализуемой матрицей $\mathbf{W}$, выделяются две стадии: $\mathbf{W} = \mathbf{W}_b \mathbf{W}_a$. Первая матрица реализует декорреляцию входных (для НС) сигналов, т. е. диагонализацию симметричной корреляционной матрицы $\mathbf{R}_{yy}$. Это достигается с помощью (29). Вторая матрица $\mathbf{W}_b$ выбирается так, чтобы диагонализировать матрицу $\mathbf{R}_{fg}$. Это выполняется с использованием схемы РВ, соответствующей (22):

$$\mathbf{W}_b(k+1) = \mathbf{W}_b(k) + \eta_0 \frac{\mathbf{I} - \mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})}{1 + \eta_0|\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})|} \mathbf{W}_b(k). \quad (31)$$

Для объединения (29) и (31) в рекурсивное соотношение для общей размешивающей матрицы $\mathbf{W}$ необходимо, чтобы составляющие ее компоненты ($\mathbf{W}_b$ и $\mathbf{W}_a$) были ортогональны с точностью

до η . Это реализуется заменой $\mathbf{I} - \mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})$ ее проекцией на пространство кососимметрических матриц [12, 15]. Полученный таким методом алгоритм имеет вид

$$\mathbf{W}_b(k+1) = \left(\mathbf{I} - \frac{\eta_0}{2} \frac{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y}) - \mathbf{g}(\mathbf{y})\mathbf{f}^T(\mathbf{y})}{1 + \eta_0 |\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})|} \right) \mathbf{W}_b(k).$$

Выполнение комбинации соотношения (29) и преобразованной формы (31) в общую размещающую матрицу $\mathbf{W}(k+1) = \mathbf{W}_b(k+1) \mathbf{W}_a(k+1)$ приводит к соотношению:¹⁰⁾

$$\mathbf{W}(k+1) = \left(\mathbf{I} + \eta_0 \frac{\mathbf{I} - \mathbf{y}\mathbf{y}^T}{1 + \eta_0 |\mathbf{y}^T\mathbf{y}|} - \frac{\eta_0}{2} \frac{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y}) - \mathbf{g}(\mathbf{y})\mathbf{f}^T(\mathbf{y})}{1 + \eta_0 |\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})|} \right) \mathbf{W}(k). \quad (32)$$

Алгоритм (32) обеспечивает расширенный анализ смеси и идентификацию ПС на основе НС с адаптивным изменением параметра скорости обучения; разработка его выполнена Кардосо и Лахельдом (Cardoso J.F., Laheld D.L.) [12].

▪ **Нелинейный алгоритм анализа главных компонент смеси.** Если ограничиться только ортогональными размещающими матрицами $\mathbf{W}$, то алгоритм (22) становится аналогом нелинейного алгоритма анализа главных компонент (АГК), полученного ранее Каруненом, Вигарио и Ойа (Karhunen, Vigarío, Oja) [21]. Сначала переопределяется обобщенная (нелинейная) ковариационная матрица: $\bar{\mathbf{R}}_{fg} = \langle \mathbf{f}(\mathbf{y}) \cdot \mathbf{g}^T(\mathbf{y}) \rangle + \mathbf{I}$, где $\mathbf{f}(\mathbf{y})$ — нечетная функция, а $\mathbf{g}(\mathbf{y}) = \mathbf{f}(\mathbf{y}) - \mathbf{y}$. После этого реализация алгоритма (22) в реальном времени сводится к итеративному выполнению соотношения

$$\begin{aligned} \mathbf{W}(k+1) &= \mathbf{W}(k) - \eta \mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})\mathbf{W}(k) = \\ &= \mathbf{W}(k) - \eta \mathbf{f}(\mathbf{y})(\mathbf{f}^T(\mathbf{y}) - \mathbf{y}^T\mathbf{W}(k)). \end{aligned} \quad (33)$$

С учетом ортогональности размещающей матрицы (т. е. $\mathbf{W}^T(k)\mathbf{W}(k) = \mathbf{I}$) соотношение (33) преобразуется к виду

$$\mathbf{W}(k+1) = \mathbf{W}(k) - \eta \mathbf{f}(\mathbf{y})(\mathbf{y}^T - \mathbf{f}^T(\mathbf{y})\mathbf{W}(k)) \quad (34)$$

и представляет нелинейный алгоритм АГК.

▪ **Адаптация алгоритма АНК при неизвестном числе ПС и их статистической особенности.** На практике число ПС часто бывает неизвестным и может изменяться со временем.

¹⁰⁾ При его получении выполняется преобразование с точностью до первой степени (малого значения) параметра η скорости обучения НС.

Статистические особенности ПС также обычно неизвестны. В условиях, когда число сенсоров ИИС (m) превышает количество ПС (n), возможны три различных подхода к решению этой сложной задачи.

– Применение АНК или робастной ортогонализации. Ортогонализация использует разложение по сингулярным числам для матрицы, которая образована набором векторов с компонентами из сигналов, регистрируемых сенсорами ИИС [7]. Ортогонализация служит для определения числа ПС и одновременно для сведения смешивающей матрицы к квадратной матрице размера $n \times n$.

– Извлечение ПС последовательно (один за другим) до тех пор, пока не будет выделено нужное количество сигналов [9, 22].

– Непосредственное использование разделяющей нейронной сети с количеством выходов, равным числу сенсоров. При этом разделяющая матрица — квадратная и обратимая.

Для рассмотренных подходов к РСС при условии $m > n$ получение разделяющей матрицы $\mathbf{W}$ основано на использовании алгоритма

$$\Delta \mathbf{W}(k) = \eta(k) (\mathbf{\Lambda}^{(k)} - \mathbf{R}_{fg}^{(k)}) \mathbf{W}(k), \quad (35)$$

где

$$\mathbf{\Lambda}^{(k)} = (1 - \eta_0) \mathbf{\Lambda}^{(k-1)} + \eta_0 \text{Диаг.}\{\mathbf{f}(\mathbf{y}(k))\mathbf{g}^T(\mathbf{y}(k))\},$$

$$\mathbf{R}_{fg}^{(k)} = (1 - \eta_0) \mathbf{R}_{fg}^{(k-1)} + \eta_0 \mathbf{f}(\mathbf{y}(k))\mathbf{g}^T(\mathbf{y}(k)),$$

а $\text{Диаг.}\{\mathbf{R}\}$ — диагональная матрица с элементами, стоящими на диагонали матрицы $\mathbf{R}$.

В случае отсутствия шума и когда число сенсоров больше количества ПС, некоторые выходы НС автоматически устанавливаются в нуль, если они ниже некоторого порогового значения. Так что при отсутствии шума (избыточные) $(m - n)$ выходных сигналов y_i в процессе адаптивного обучения НС приближаются к нулю и только n сигналов на выходе НС будут соответствовать восстановленным ПС.¹¹⁾

Для малой величины аддитивного шума и $m > n$ алгоритм (35) может оказаться нестабильным: при большом количестве итераций норма матрицы $\mathbf{W}$

¹¹⁾ Обычно принимается, что матрица $\mathbf{W}$ квадратная, т. е. число выходов разделяющей системы равно числу сенсоров ИИС, но на практике число ПС может быть и меньше, чем число сенсоров ($m \geq n$). Такая модель оправдана тем, что количество ПС может изменяться с течением времени. Кроме того, практически аддитивный шум может рассматриваться как дополнительный неизвестный ПС, так что разумно этот сигнал шума тоже извлекать из смеси. Заметим также, что в (36) символ $\text{tr}(\mathbf{A})$ означает след, т. е. сумму диагональных элементов матрицы $\mathbf{A}$ (от англ. trace — след).

начинает возрастать. Чтобы избежать этого достаточно использовать модификацию алгоритма "с фактором забывания γ " в виде

$$\Delta \mathbf{W}(k) = \eta(k) (\Lambda^{(k)} - \mathbf{R}_{fg}^{(k)} - \gamma(k) \mathbf{I}_m) \mathbf{W}(k), \quad (36)$$

где $\gamma(k) = \text{tr} [\mathbf{W}^T(k) (\Lambda^{(k)} - \mathbf{R}_{fg}^{(k)}) \mathbf{W}(k)]$.

Этот алгоритм стабилизирует норму матрицы $\mathbf{W}(k)$, т. е. $\|\mathbf{W}(k)\|=1$, если $\|\mathbf{W}^T(0)\|=1$.

Как отмечено выше, векторная форма функций $\mathbf{f}[\mathbf{y}(k)]$ и $\mathbf{g}[\mathbf{y}(k)]$ использована для обозначения

$$\mathbf{f}(\mathbf{y}) = [f_1(y_1), f_2(y_2), \dots, f_n(y_n)]^T \quad \text{и} \\ \mathbf{g}(\mathbf{y}) = [g_1(y_1), g_2(y_2), \dots, g_n(y_n)]^T.$$

Если измеренные сигналы $x_i(k)$ содержат смеси как с подгауссовым, так и с надгуссовым распределениями ПС, то сохранение стабильности алгоритма обеспечивает ПФН с переключением

$$f_i(y_i) = \begin{cases} \varphi_i(y_i) & \text{для } k_4(y_i) > \delta, \\ y_i & \text{в другом случае;} \end{cases} \quad (37)$$

$$g_i(y_i) = \begin{cases} y_i & \text{для } k_4(y_i) > -\delta, \\ \varphi_i(y_i) & \text{в другом случае,} \end{cases}$$

где $k_4(y_i) = (E\{y_i^4\}/E^2\{y_i^2\} - 3)$ — нормализованная величина эксцесса; $\delta \geq 0$ — маленькое значение порога; $\varphi_i(y_i)$ — нелинейная функция, зависящая от распределения ПС [8]. В качестве ПФН возможен выбор таких функций, как $\varphi_i(y_i) = \tanh(\theta_i y_i)$ или $\varphi_i(y_i) = \text{sign}(y_i) \exp(-\theta_i |y_i|)$ при $\theta_i \geq 0$.

Алгоритмы обучения (35)–(37) отслеживают и оценивают статистические характеристики каждого сигнала на выходе НС и в зависимости от знака и величины нормализованного эксцесса автоматически выбирают (или переключают) подходящую нелинейную ПФН такую, которая обеспечивает успешное выделение всех негауссовых ПС. Это становится возможным, т. к. нормализованный эксцесс служит мерой отличия распределения сигнала от гауссовой формы. В описанном методе ПФН "активационная функция" НС выбрана в виде нелинейности, которая адаптивно изменяется со временем.

ЗАКЛЮЧЕНИЕ

Проанализированы методы и алгоритмы, которые служат для разделения сигналов, образующих смесь, и ориентированы на нейронные сети (НС). Дана трактовка нескольких подходов к разделению сигналов смеси (РСС) и рассмотрена группа адаптивных алгоритмов обучения НС, основанных на использовании декорреляции, статистик второго порядка и модифицированного градиента. Выделены алгоритмы для выполнения РСС при наличии зашумленных сигналов, сигналов с большим разбросом величин или при условии, когда количество первичных сигналов (ПС), образующих смесь, и их статистические свойства неизвестны.

— Для НС с рекуррентной структурой и с прямым распространением сигнала приведено адаптивное правило Коммона, Джуттен и Хераулт для обучения НС разделению линейно смешанных сигналов и способ выбора нелинейной функции преобразования нейронов (ПФН). Приведено обобщение структуры НС для разделения смеси произвольного числа ПС и модификация метода РСС на основе использования многослойной НС.

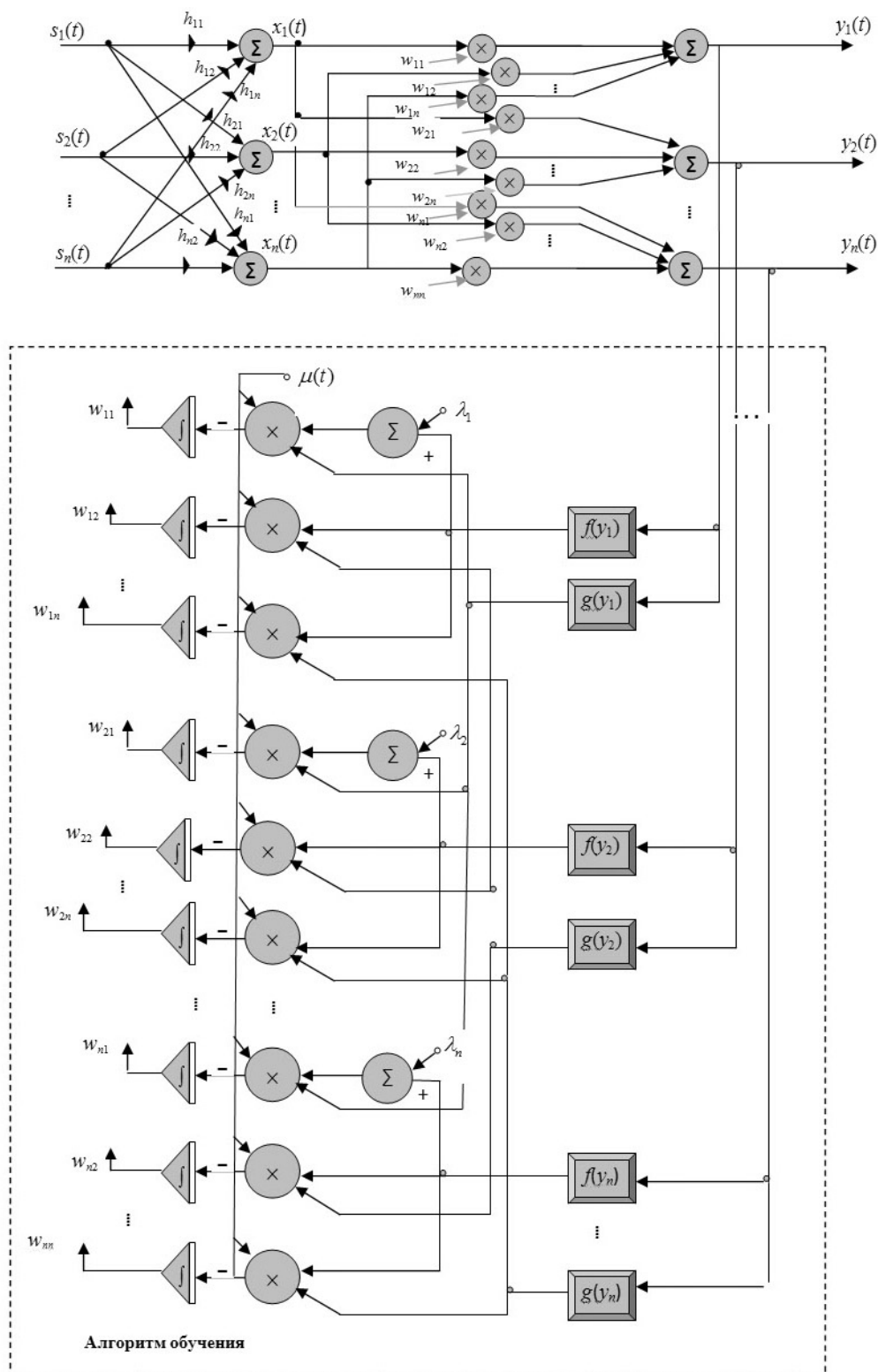
— Показано, что формализованный подход к получению алгоритмов, устойчивых к шуму, основан на методе итеративного обращения обобщенной ковариационной матрицы $\mathbf{R}_{fg}(\mathbf{W}) = E\{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})\}$ и что возможность такого обращения и стабильность процесса обучения НС гарантируются выбором достаточно малого параметра скорости обучения η : $\eta(l) < \|\hat{\mathbf{R}}_{fg}^{(l)} - \mathbf{I}\|^{-1}$, который меняется на каждом шаге итеративного процесса обучения.

— Дано определение и трактовка понятий м-градиента и обобщенного м-градиента. Рассмотрен метод формирования робастного алгоритма анализа независимых компонент (АНК) на основе обобщенного м-градиента.

— Сформированы основные положения практической реализации адаптивных алгоритмов для РСС на основе НС: робастного алгоритма обучения НС (устойчивого к зашумленности измеряемых сигналов), алгоритма АНК при неизвестном числе ПС и их статистической особенности.

— Методические основы представления адаптивных алгоритмов РСС (в том числе алгоритмов реального времени) базируются на разработках Ху, Принсайра, Ву, Сичоки, Тавонмаса, Кардосо, Лахельда, Крусеза, Кастедо и Дугласа.

Приложение

РАСШИРЕННАЯ ФУНКЦИОНАЛЬНАЯ СХЕМА ВЫПОЛНЕНИЯ
ЛОКАЛЬНО АДАПТИВНОГО АЛГОРИТМА ОБУЧЕНИЯ НС_ПРС

СПИСОК ЛИТЕРАТУРЫ

1. Pope K.J., Bogner R.E. Signal Separation: Linear Instantaneous Combinations // *Digital Signal Processing*. 1996. V. 6. P. 5–16.
2. Weinstein E., Oppenheim A.F., Feder M., Buck J.R. Iterative and Sequential Algorithms for Multisensor Signal Enhancement // *IEEE Transactions on Signal Processing*. 1994. V. 42, N 4. P. 846–859.
3. Меркушева А.В. Нейросетевые методы обработки сигналов в информационных системах: Элементы структуры, принципы обучения и мера многообразия отображений информации // *Информационные технологии*. 2005. № 3. С. 9–20.
4. Karhunen J., Joutsensalo J. Representation and Separation of Signals Using Nonlinear PCA Type Learning // *Neural Networks*. 1994. V. 7, N 1. P. 113–127.
5. Малихина Г.Ф., Меркушева А.В. Концепции информации и энтропии в нейросетевых методах анализа сигналов в ИИС // *Информационные технологии*. 2005. № 12. С. 26–36.
6. Douglas S.C., Cichocki A. Neural Networks for Decorrelation of Signals // *IEEE Transaction on Signal Processing*. 1997. V. 45, N 11. P. 2829–2842.
7. Меркушева А.В., Малихина Г.Ф. Методы и алгоритмы разделения смеси сигналов: Применение декорреляции и статистик второго порядка / *Научное приборостроение*. 2009. Т. 19, № 2. С. 90–102.
8. Малихин В.М., Меркушева А.В. Методы и алгоритмы разделения смеси сигналов II. Применение м-градиента к анализу независимых компонент // *Научное приборостроение*. 2009. Т. 19, № 4. С. 83–95.
9. Малихин В.М. Выделение сигналов из их смеси каскадной нейронной сетью // *Сб. материалов XVI Международной конференции по вычислительной механике и современным прикладным программным системам (ВМСППС'2009)*. М.: Вузовская книга, 2009. С. 512–515.
10. Common P., Jutten C., Herault J. Blind Separation of Sources: Problem Statement // *Signal Processing*. 1991. V. 24. P. 11–20.
11. Малихина Г.Ф., Меркушева А.В. Сеть с симметричной функцией преобразования нейронов на основе статистик высокого порядка (СВП) для снижения шума нестационарного сигнала // *Научное приборостроение*. 2008. Т. 18, № 4. С. 129–136.
12. Cardoso J.F., Laheld D.L. Equivariant Adaptive Source Separation // *IEEE Transaction on Signal Processing*. 1996. V. 44, N 12. P. 3017–3030.
13. Cichocki A., Thavonmas R. On-Line Algorithm for Signal Extraction of Arbitrary Distributed, but Temporally Correlated Sources Using Second Order Statistics // *Neural Processing Letters*. 2000. V. 12, N 1, August. P. 91–98.
14. Xu D., Principe J.C., Wu H.-C. Generalized Eigendecomposition with an On-Line Local Algorithm // *IEEE Signal Processing Letters*. 1998. V. 5, N 11. P. 298–301.
15. Cruces S., Cichocki A., Castedo L. An Iterative Inversion Approach to Source Separation // *IEEE Transactions on Neural Networks*. 2000. V. 11, N 6. P. 1423–1427.
16. Голуб Дж., Ван Лоун У. Матричные вычисления. М.: Мир, 1999. 548 с.
17. Cichocki A., Unbehauen R. Robust Neural Networks with On-Line Learning for Identification and Separation of Sources // *IEEE Transactions on Circuits and Systems. I: Fundamentals of Theory and Applications*. V. 43, N 11. P. 894–906.
18. Cichocki A., Unbehauen R., Rummert E. Robust Learning Algorithm for Separation of Signals // *Electronic Letters*. 1994. V. 30, N 17. P. 1386–1387.
19. Amari S. Natural Gradient Works Efficiently in Learning // *Neural Computation*. 1998. N 10. P. 271–276.
20. Amari S., Cichocki A. Adaptive Signal Processing — Neural Network Approaches // *Proceedings of IEEE*. 1998. V. 86. P. 1186–1187.
21. Karhunen J., Oja E., Wang L., Vigario L., Joutsensalo J. A Class of Neural Networks for Independent Component Analysis // *IEEE Transaction on Neural Networks*. 1997. V. 8, N 3. P. 486–503.
22. Картамышев А.В., Малихин В.М., Малихина Г.Ф., Меркушева А.В. Выделение сигналов из их смеси (элементы алгоритмов и нейросетевых структур) // *Научное приборостроение*. 2010. Т. 20, № 1. С. 96–112.

СПбГТУ, Санкт-Петербург

Контакты: Малихина Галина Фёдоровна,
g_f_malychina@mail.ru

Материал поступил в редакцию 5.04.2010.

ADAPTIVE METHODS AND ALGORITHMS FOR SEPARATION OF SIGNAL MIXTURE WITH INDEPENDENT COMPONENTS

G. F. Malykhina, A. V. Merkusheva

SPbSTU, Saint-Petersburg

Methods and algorithms that are intended for separation of signal mixture of independent components (initial signals generating the mixture) are discussed. Methods realization is oriented on neural networks. The models of neural networks and the rules for their adaptive learning are discussed.

Keywords: signal mixtures, generating signals, initial components, independence, adaptive methods, algorithms