

УДК 681.51; 621.391; 519.21

© Г. Ф. Малыхина, А. В. Меркушева

РОБАСТНЫЕ МЕТОДЫ ДЛЯ РАЗДЕЛЕНИЯ СМЕСИ СИГНАЛОВ И АНАЛИЗА НЕЗАВИСИМЫХ КОМПОНЕНТ ПРИ ЗАШУМЛЕННЫХ ДАННЫХ

Рассмотрены робастные методы и соответствующие им алгоритмы на основе нейронных сетей для реализации разделения сигналов смеси. Восстановление формы первичных сигналов, идентификация смешивающей и разделяющей матриц выполняются без существенных ошибок при наличии значительных уровней шума в анализируемых сигналах смеси.

Кл. сл.: смесь сигналов, зашумленность, восстановление компонент, робастные методы, алгоритмы, нейронные сети, матрица смешивания, идентификация

ВВЕДЕНИЕ

Анализ методов разделения смеси сигналов (РСС) (так же как и методов менее общей задачи — анализа независимых компонент) имеет множество приложений. Идентификация первичных сигналов, регистрируемых в виде смеси, требуется в ряде специализированных информационно-измерительных систем, ИИС, (при радарном зондировании, в медико-биологических комплексах), при адаптивном выравнивании усиления и адаптивном снижении шумов в каналах связи, при усилении сигналов многоканальных датчиков, при анализе негауссовых массивов данных, при обработке цветных изображений, в отдельных задачах радиационного контроля [1–6].

Методы РСС достаточно разнообразны. Они определяются видом доступной априорной информации, мерой надежности предположений о типе распределения первичных сигналов (ПС), образующих смесь,¹⁾ или о наличии временной структуры ПС, о разнице их корреляционной функции. Поэтому при создании алгоритма РСС могут использоваться либо методы на основе статистик ПС второго порядка анализируемых ПС [7], либо методы на основе статистик более высокого порядка (на кумулянтах [8]). Однако все анализируемые ранее методы были ориентированы либо на отсутствие зашумленности ПС, либо на наличие пренебрежимо малого уровня шума, который не может существенно влиять на результаты РСС, и в частности не может давать смещение в оценках матрицы смешивания (или обратной ей

матрицы размешивания). Поэтому здесь будет рассмотрено расширение адаптивных методов и алгоритмов, позволяющее существенно снизить смещения оценок, обусловленные шумом в измерительных каналах ПС. Кроме того, анализируются динамические рекуррентные нейронные сети, позволяющие осуществлять одновременную оценку неизвестной смешивающей матрицы, совокупности ПС и снижать уровень шума в восстановленных сигналах. Оптимальный выбор нелинейной функции преобразования нейронов (ФПН) при зашумленных сигналах основан на использовании обобщенного экспоненциально-степенного распределения для сигнала шума.²⁾

В целом задача подавления шума достаточно трудна, а иногда и не поддается полной трактовке, поскольку здесь имеется $(m + n)$ неизвестных сигналов (n — ПС и m — сигналов шума) и в то же время имеется только m доступных или измеряемых сигналов сенсоров. Однако в ряде практических случаев при наличии некоторой информации относительно шума оказываются возможными оценка несмещенной разделяющей матрицы и снижение или полное подавление шума. Например, в некоторых ситуациях шум окружающей среды может быть измерен или моделирован.

¹⁾ В том числе предположений о негауссовом характере распределения первичных сигналов.

²⁾ Плотность обобщенного экспоненциально-степенного распределения при соответствующем выборе ее параметров включает несколько видов распределений: нормальное, Лапласа, подгауссово и надгауссово распределения (т. е. распределения с уменьшенным и с повышенным (относительно гауссова) значением эксцесса) [9].

УДАЛЕНИЕ СМЕЩЕНИЯ В ОЦЕНКАХ НЕЗАВИСИМЫХ КОМПОНЕНТ СМЕСИ И МАТРИЦЫ СМЕШИВАНИЯ

Алгоритмы отбеливания

Алгоритм декорреляции (или отбеливающий³⁾ алгоритм) для задачи анализа независимых компонент (АНК) вектора-сигнала $\mathbf{x}(t)$ имеет вид

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

а его усредненный вариант определяется соотношением $\Delta\mathbf{W}(k) = \eta(k)[\mathbf{I} - \langle \mathbf{y}(k)\mathbf{y}^T(k) \rangle]\mathbf{W}(k)$, где $\mathbf{y}(k) = \mathbf{W}(k)\mathbf{x}(k)$; $\langle \cdot \rangle$ обозначает статистическое усреднение выражения, стоящего в скобках.

Когда сигнал $\mathbf{x}(k)$ зашумлен, т. е. $\mathbf{x}(k) = \hat{\mathbf{x}}(k) + \mathbf{v}(k)$, где $\hat{\mathbf{x}}(k)$ и $\hat{\mathbf{y}}(k) = \mathbf{W}(k)\hat{\mathbf{x}}(k)$ являются оценками без шума для входного и выходного сигналов, то можно показать, что аддитивный шум $\mathbf{v}(k)$, относящийся к $\mathbf{x}(t)$, вводит смещение в оцененной матрице \mathbf{W} . Корреляционная матрица выхода оценивается выражением

$$\hat{\mathbf{R}}_{\mathbf{y}\mathbf{y}} = \langle \mathbf{y}(k)\mathbf{y}^T(k) \rangle = \mathbf{W}\hat{\mathbf{R}}_{\hat{\mathbf{x}}\hat{\mathbf{x}}}\mathbf{W}^T + \mathbf{W}\hat{\mathbf{R}}_{\mathbf{v}\mathbf{v}}\mathbf{W}^T, \quad (2)$$

где

$$\hat{\mathbf{R}}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} = \langle \hat{\mathbf{x}}(k)\hat{\mathbf{x}}^T(k) \rangle = (1/N) \sum_{k=1}^N \hat{\mathbf{x}}(k)\hat{\mathbf{x}}^T(k)$$

$$\text{и } \hat{\mathbf{R}}_{\mathbf{v}\mathbf{v}} = \langle \mathbf{v}(k)\mathbf{v}^T(k) \rangle.$$

Если допустить, что ковариационная матрица шума известна (например, $\hat{\mathbf{R}}_{\mathbf{v}\mathbf{v}} = \hat{\sigma}_v^2\mathbf{I}$) или может быть оценена, то модифицированный алгоритм, обеспечивающий удаление смещения, определяется соотношением

$$\Delta\mathbf{W}(k) = \eta(k)[\mathbf{I} - \mathbf{R}_{\mathbf{y}\mathbf{y}}^{(k)} + \mathbf{W}(k)\hat{\mathbf{R}}_{\mathbf{v}\mathbf{v}}\mathbf{W}^T(k)]\mathbf{W}(k). \quad (3)$$

Стохастический вариант этого алгоритма при условии, что $\hat{\mathbf{R}}_{\mathbf{v}\mathbf{v}} = \hat{\sigma}_v^2\mathbf{I}$, и оценке $\mathbf{R}_{\mathbf{y}\mathbf{y}}^{(k)} = \mathbf{y}\mathbf{y}^T$ имеет вид

$$\mathbf{W}(k) = \eta(k)[\mathbf{I} - \mathbf{y}\mathbf{y}^T + \hat{\sigma}_v^2\mathbf{W}(k)\mathbf{W}^T(k)]\mathbf{W}(k).$$

Альтернативно, чтобы снизить влияние белого аддитивного шума для окрашенных сигналов, может быть использовано модифицированное правило обучения [10]:

³⁾ Термин "отбеливание" используется для краткого обозначения процедуры декорреляции набора сигналов, которая переводит их совместную корреляционную матрицу в диагональную форму, элементами которой являются дисперсии этих сигналов [7], [10].

$$\Delta\mathbf{W}(k) = \eta(k)[\mathbf{I} - (1/2) \cdot (\langle \mathbf{y}(k)\mathbf{y}^T(k-p) \rangle + \langle \mathbf{y}(k-p)\mathbf{y}^T(k) \rangle)]\mathbf{W}(k). \quad (4)$$

где p — малое (целое) время задержки.⁴⁾

Алгоритмы анализа независимых компонент

Логическая схема метода удаления смещения оценок в алгоритмах отбеливания (1)–(4) применима для удаления смещения оценок коэффициентов для группы алгоритмов АНК [11]. Это показано на базовом алгоритме АНК,⁵⁾ форме

$$\Delta\mathbf{W}(k) = \eta(k)[\mathbf{I} - \mathbf{R}_{\mathbf{f}\mathbf{g}}^{(k)}]\mathbf{W}(k). \quad (5)$$

Нелинейные функции векторов подбираются соответствующим образом, как рассмотрено в [12], [13]. Алгоритм работает при условии, что $E\{f_i(y_i)\} = 0$ или $E\{g_i(y_i)\} = 0$. Для того чтобы удовлетворить эти условия при несимметричном распределении источников для каждого индекса i , достаточно использовать только одну нелинейную функцию, а вторая — линейная:

для любого i используется $f_i(y_i)$ и $g_i(y_i) = y_i$ или $f_i(y_i) = y_i$ и $g_i(y_i)$.

Обучающий алгоритм (5) хорошо работает при использовании для РСС без шума, но качество функционирования его значительно ухудшается при измерениях с шумом: появляется нежелательное смещение коэффициентов (элементов размешивающей матрицы \mathbf{W}) и зашумленность восстановленных сигналов. Для оценки смещения коэффициентов используется разложение нелинейных функций $f_i(y_i)$ и $g_i(y_i)$ в ряд Тейлора в точке \hat{y}_i без шума.

Обобщенная ковариационная матрица $\mathbf{R}_{\mathbf{f}\mathbf{g}}$ оценивается по соотношению

⁴⁾ Приведенный алгоритм теоретически нечувствителен к аддитивному белому шуму. Фактически вместо диагонализации ковариационной матрицы с нулевой задержкой $\hat{\mathbf{R}}_{\mathbf{y}\mathbf{y}} = \langle \mathbf{y}(k)\mathbf{y}^T(k) \rangle$ здесь диагонализирована ковариационная матрица с задержкой времени $\hat{\mathbf{R}}_{\mathbf{y}\mathbf{y}}(p) = (1/2) \cdot (\langle \mathbf{y}(k)\mathbf{y}^T(k-p) \rangle + \langle \mathbf{y}(k-p)\mathbf{y}^T(k) \rangle)$, которая нечувствительна к аддитивному белому шуму по той причине, что свободные от шума сигналы сенсоров $\hat{\mathbf{x}}(k)$ являются окрашенными сигналами и число наблюдений достаточно большое (в типовом случае более чем 10^4 отсчетов).

⁵⁾ Более полная и детальная информация о методах и алгоритмах АНК и правиле обучения нейронной сети в условиях низкой зашумленности сигналов содержится в [12] и [13].

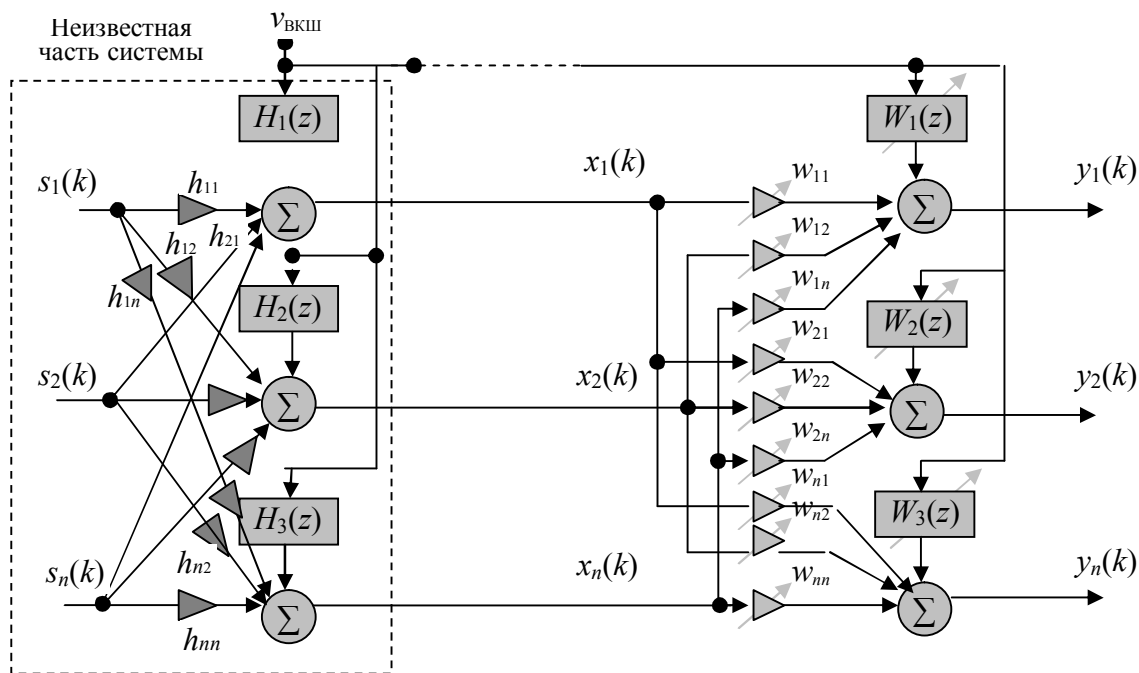


Рис. 1. Система, реализующая смешивание ПС и РСС с практически полным снижением шума (в предположении доступности ВКШ)

$$\mathbf{R}_{fg} = E \{ \mathbf{f}[y(k)] \mathbf{g}[y^T(k)] \} \cong E \{ \mathbf{f}[\hat{y}(k)] \mathbf{g}[\hat{y}^T(k)] \} + \mathbf{k}_f \mathbf{W} \hat{\mathbf{R}}_{vv} \mathbf{W}^T \mathbf{k}_g \quad (6)$$

где \mathbf{k}_f и \mathbf{k}_g — диагональные матрицы с элементами

$$k_{f_i} = E \{ df_i(y_i(k)) / dy_i \};$$

$$k_{g_i} = E \{ dg_i(y_i(k)) / dy_i \}.$$

Полученный так модифицированный адаптивный алгоритм обучения НС при сниженном смещении коэффициентов имеет вид ⁶⁾

$$\Delta \mathbf{W}(k) = \eta(k) [\mathbf{I} - \mathbf{R}_{fg}^{(k)} + \mathbf{k}_f \mathbf{W}(k) \hat{\mathbf{R}}_{vv}] \cdot \mathbf{W}^T(k) \mathbf{k}_g \mathbf{W}(k) \quad (7)$$

РАЗДЕЛЕНИЕ СМЕСИ СИГНАЛОВ С РАЗЛИЧНЫМИ ШУМАМИ

Эффективное РСС при аддитивном окрашенном шуме, которым нельзя пренебречь, зависит от

⁶⁾ В случае, когда все ПС распределены одинаково, т. е. для любого i : $f_i(y_i) = f(y_i)$ и $g_i(y_i) = g(y_i)$ и $\hat{\mathbf{R}}_{vv} = \hat{\sigma}_y^2 \mathbf{I}$, поправка на смещение упрощается и становится равной $\mathbf{B} = \hat{\sigma}_y^2 k_f \cdot k_g \mathbf{W} \mathbf{W}^T$.

того, как и насколько может быть подавлен этот шум. Основная трудность связана с наличием $(m + n)$ неизвестных сигналов (m сенсоров в ИИС и n ПС), т. к. при отсутствии дополнительной информации относительно модели смешивания и (или) шума задача РСС является значительно недоопределенной и ее очень трудно (а иногда невозможно) решить.

Однако во многих практических случаях можно измерить или моделировать шум окружающей среды, который, передаваясь с различными передаточными функциями на выходы смешивающей матрицы, образует зашумленные сигналы x_1, x_2, \dots, x_n , преобразуемые затем с помощью нейронной сети в оценки ПС. В таких случаях говорят о внешнем контролируемом шуме (ВКШ) v_{BKSH} . ⁷⁾ Описанная ситуация показана на рис. 1. Примером ее служит задача выделения определенного речевого сигнала при контроле обсуждения группой людей (часто одновременно говорящих) информационно высокозначимых проблем. При этом наличие микропауз позволяет получать ВКШ, используя отдельный микрофон и контроль on line. В биомедицинских приложениях типа ЭКГ и ЭЭГ таким же образом измеряется ВКШ путем разме-

⁷⁾ Иногда его называют опорным, или референтным шумом, например, в работе Дугласа, Сичоки и Амари (Douglas, Cichocki, Amari) [11].

щения (специально расположенного) дополнительного электрода.

Обусловленный условиями внешней среды ВКШ может влиять (некоторым неизвестным образом) на каждый сенсор ИИС. При этом могут возникать такие эффекты, как задержка, реверберация, эхо и нелинейные искажения. Предполагается, что, прежде чем достигнуть каждого сенсора, $v_{\text{ВКШ}}$ обрабатывается некоторой неизвестной линейной или нелинейной динамической системой (рис. 1). Модель шума может быть различного типа: авторегрессией скользящего среднего (АРСС), нелинейной моделью авторегрессии скользящего среднего (НАРСС) или фильтром (со сверткой) с конечной импульсной характеристикой (КИХ-фильтром) [14, 15]. В случае модели со сверткой ВКШ обрабатывается КИХ-фильтром с функцией

импульсного отклика, параметры которой необходимо оценивать. Следовательно, аддитивный шум на i -м сенсоре моделируется соотношением [16]

$$v_i(k) = \sum_{p=1}^L [h_{ip}(k)z^{-p}]v_{\text{ВКШ}}(k) = \sum_{p=1}^L h_{ip}(k)v_{\text{ВКШ}}(k-p), \quad (8)$$

где $z^{-1} = e^{-sT}$ — оператор, выражающий задержку на единицу времени.

Такая модель отражена на рис. 1 и 2, а. В ней принято, что известный ВКШ поступает на каждый сенсор (добавляясь к смеси ПС) с различной неизвестной временной задержкой и различными коэффициентами $h_{ip}(k)$, определяющими затухание. Процессы неизвестного смешивания и свертки в матричной форме описывает соотношение (9):

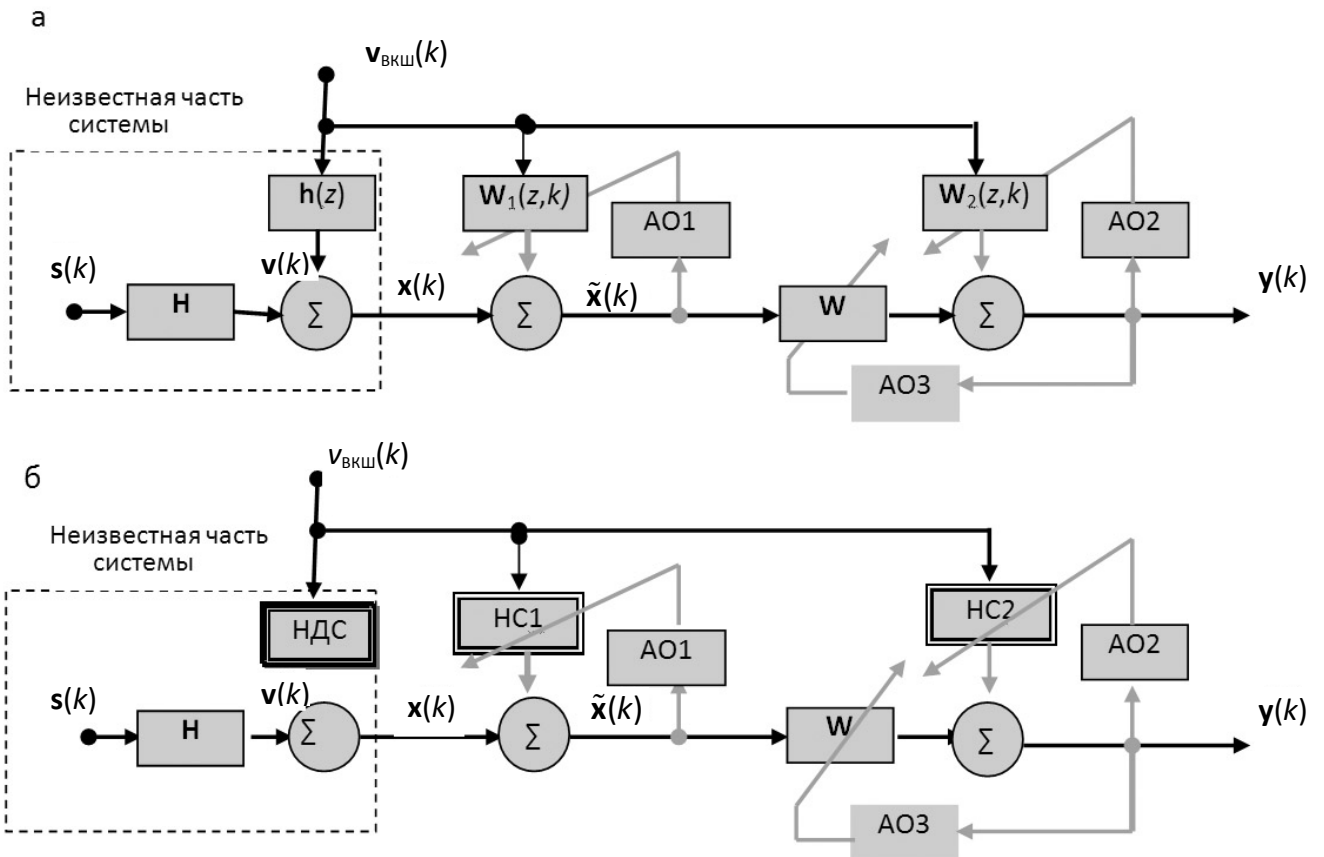


Рис. 2. Схема, отражающая многофазовое устранение шума и РСС.

а — линейная модель шума в форме свертки; б — более общая модель аддитивного шума, определяемого нелинейной динамической системой (НДС) и адаптивными нейронными сетями (НС1 и НС2); АО1 и АО2 — обучающие алгоритмы на основе МНК или обратного распространения ошибки (при супервизорной форме обучения); АО3 — обучающий алгоритм для РСС

$$\mathbf{x}(k) = \mathbf{H}\mathbf{s}(k) + \mathbf{h}(z)v_{\text{ВКШ}}(k), \quad (9)$$

где $\mathbf{h}(z) = [H_1(z), H_2(z), \dots, H_n(z)]^T$ и $H_i(z) = h_{i0}(z) + h_{i1}(z)z^{-1} + \dots + h_{iL}(z)z^{-L}$.

Аналогично РСС и выделение ВКШ (операцией, обратной относительно свертки) определяет выражение

$$\mathbf{y}(k) = \mathbf{W}\mathbf{x}(k) - \tilde{\mathbf{w}}(z)v_{\text{ВКШ}} = \mathbf{W}\mathbf{H}\mathbf{s}(k) + \mathbf{W}\mathbf{h}(z)v_{\text{ВКШ}} - \tilde{\mathbf{w}}(z)v_{\text{ВКШ}}, \quad \text{где}$$

$$\tilde{\mathbf{w}}(z) = [W_1(z), \dots, W_n(z)]^T \text{ и } W_i(z) = \sum_{j=0}^M \tilde{w}_{ij}z^{-j}.$$

Лучшее снижение шума реализуется минимизацией функции стоимости в виде обобщенной энергии всех сигналов на выходе НС (рис. 1) при определенных ограничениях и обеспечении взаимной независимости этих сигналов.

Алгоритм обучения НС для полного снижения шума

Основу алгоритма для снижения шума составляет одновременное использование двух процедур обучения НС: сигналы выделяются из их линейной смеси посредством алгоритма РСС, а аддитивный шум оценивается и вычитается путем минимизации общей энергии выходных сигналов. Функция стоимости (критерий минимизации), за которую принимается энергия, имеет вид

$$J(\tilde{\mathbf{w}}) = (1/2) \sum_{i=1}^n E\{|y_i|^2\}. \quad (10)$$

Минимизация $J(\tilde{\mathbf{w}})$ относительно \tilde{w}_{ip} приводит к процедуре алгоритма наименьших средних квадратов (известного как алгоритм МНК):

$$\begin{aligned} \tilde{w}_{ip}(k+1) &= \tilde{w}_{ip}(k) - \tilde{\eta}(k) \left(\partial J(\tilde{\mathbf{w}}) / \partial \tilde{w}_{ip} \right) = \\ &= \tilde{w}_{ip}(k) + \langle \tilde{\eta}(k) y_i(k) v_{\text{ВКШ}}(k-p) \rangle. \end{aligned} \quad (11)$$

Число единиц временной задержки M в модели разделения (или операции, обратной относительно свертки) должно быть значительно больше, чем L в модели смешивания ($M \gg L$).

Для снижения шума может использоваться многофазная модель (рис. 2, а). В этой модели подавление шума и РСС выполняются последовательно за две или более стадий. Сначала уничтожается шум, содержащийся в смесях (регистрируемых сенсорами), а затем выделяются ПС. Чтобы подавить аддитивный шум и получить адаптивный алгоритм обучения при неизвестных коэффициентах $h_{ip}(k)$, используется минимизация обобщенной энергии выходных сигналов $\tilde{\mathbf{x}}(k) = [\tilde{x}_1(k), \tilde{x}_2(k), \dots, \tilde{x}_n(k)]^T$. Т. е. в качестве критерия — полной функции стоимости как обобщенной энергии — применяется

$J(\tilde{\mathbf{w}}_1) = \sum_{i=1}^n E\{\rho_i(\tilde{x}_i)\}$, где $\rho_i(\tilde{x}_i)$, ($i = 1, 2, \dots, n$) — частные функции потерь в виде $\rho_i(\tilde{x}_i) = (1/\gamma) \log \cosh(\gamma \tilde{x}_i)$ или $\rho_i(\tilde{x}_i) = (1/2) |\tilde{x}_i|^2$ и $\tilde{x}_i(k) = x_i(k) - \sum_{p=1}^M \tilde{w}_{ip} \cdot v_{\text{ВКШ}}(k-p)$ при $\forall i$. Минимизация такой функции стоимости $J(\tilde{\mathbf{w}}_1)$, реализуемой методом градиентного спуска, приводит к алгоритму обучения НС в виде

$$\begin{aligned} \tilde{w}_{ip}(k+1) &= \tilde{w}_{ip}(k) - \tilde{\eta}(k) \cdot \left(\partial J(\tilde{\mathbf{w}}) / \partial \tilde{w}_{ip} \right) \approx \\ &\approx \tilde{w}_{ip}(k) + \tilde{\eta}(k) f_{\text{ВКШ}}[\tilde{x}_i(k)] v_{\text{ВКШ}}(k-p), \end{aligned} \quad (12)$$

где $f_{\text{ВКШ}}[\tilde{x}_i(k)]$ — нелинейная функция $f_{\text{ВКШ}}[\tilde{x}_i(k)] = (\partial \rho_i(\tilde{x}_i) / \partial \tilde{x}_i)$, в качестве которой достаточно использовать $f_{\text{ВКШ}}(\tilde{x}_i) = \tilde{x}_i$ или $f_{\text{ВКШ}}(\tilde{x}_i) = \tanh(\gamma \cdot \tilde{x}_i)$.

В описанной выше модели, которая реализует адаптивное подавление шума и включает линейное КИХ-преобразование КВШ, шум оценивается как взвешенная сумма задержанных отсчетов контролируемого внешнего шума. Однако в отдельных случаях линейная адаптивная система подавления шума не может обеспечить его снижение до приемлемого уровня. Это бывает, когда сигналы сенсоров ИИС подвергаются искажению за счет КВШ, который преобразован сложной динамической нелинейной системой.

В таких приложениях (особенно при обработке биомедицинских сигналов) для оптимального подавления интерференции, связанной с КВШ, и собственно шума, поступающего непосредственно на сенсоры, требуется нелинейная адаптивная обработка регистрируемых и измеряемых сигналов. В таких случаях вместо линейных фильтров целесообразно применять структуры на основе нейронных сетей (рис. 2, б) и обучение их с использованием алгоритма обратного распространения [17, 18].

АДАПТИВНЫЕ АЛГОРИТМЫ АНАЛИЗА НЕЗАВИСИМЫХ КОМПОНЕНТ СМЕСИ, ОСНОВАННЫЕ НА КУМУЛЯНТАХ

Основное достоинство алгоритмов, основанных на кумулянтах, состоит в том, что они не зависят от гауссовых шумов.⁸⁾ Это свойство связано с тем,

⁸⁾ Необходимым условием использования кумулянт в алгоритмах РСС является большое количество отсчетов анализируемых сигналов, позволяющее достаточно корректно оценивать значения кумулянт. Структура кумулянт (кроме их краткого определения, приведенного здесь) более подробно дана в [8].

что статистики, выше второго порядка (в том числе и кумулянты), для гауссова шума равны нулю. Это особенно полезно при обработке и анализе биомедицинских сигналов, которые обычно связаны с высоким уровнем гауссова шума. Однако т. к. кумулянты третьего порядка равны нулю для сигнала с любым симметричным распределением, то для алгоритмов РСС при сигналах с симметричным распределением используются кумулянты четвертого и более высокого порядков.

Функция стоимости, основанная на кумулянтах, и алгоритмы с их применением

В качестве меры независимости сигналов $\{y_i\}_{i=1,2,\dots,n}$ используется критерий Кульбака—Лейблера [19, 20], который приводит к функции стоимости

$$R(\mathbf{y}, \mathbf{W}) = (-1/2) \log(|\det(\mathbf{W}\mathbf{W}^T)|) - \sum_{i=1}^n E\{\log(p_i(y_i))\},$$

или что эквивалентно выражению:

$$R(\mathbf{y}, \mathbf{G}) = (-1/2) \log(|\det(\mathbf{G})|) - \sum_{i=1}^n E\{\log(p_i(y_i))\}, \quad (13)$$

где $\mathbf{G} = \mathbf{W}\mathbf{H}$ (\mathbf{G} — общая матрица смешивания-размешивания размерности $(n \times n)$). Тем не менее во многих приложениях плотность распределения сигналов бывает априори неизвестной, и при выводе практических алгоритмов приходится использовать нелинейные функции, которые в некоторых случаях могут не соответствовать истинному виду распределения сигналов. Поэтому целесообразно использовать другой вид функции стоимости, которая зависит от кумулянт.

Кумулянта $C_q(y_i)$ порядка q сигнала y_i и матрица $\mathbf{C}_{pq}(\mathbf{y}, \mathbf{y})$ кросс-кумулянт с элементами $[\mathbf{C}_{pq}(\mathbf{y}, \mathbf{y})]_{ij} = \text{Кум.}(\underbrace{y_i, y_i, \dots, y_i}_p, \underbrace{y_j, y_j, \dots, y_j}_q)$ выражаются через статистические моменты (высокого порядка) сигналов.⁹⁾

⁹⁾ Оценка кумулянт через статистические моменты определяется соотношением:

$$\text{Кум.}(y_1, y_2, \dots, y_n) = \sum_{(p_1, p_2, \dots, p_m)} (-1)^{m-1} (m-1)! \cdot E\left\{\prod_{i \in p_1} y_i\right\} \cdot E\left\{\prod_{i \in p_2} y_i\right\} \cdot \dots \cdot E\left\{\prod_{i \in p_m} y_i\right\},$$

где сумма распространяется на все возможные разделения (на слагаемые p_1, p_2, \dots, p_m , $m = 1, 2, \dots, n$) множества натуральных чисел $(1, 2, \dots, n)$ [21]. Эти вычисления просты для кумулянт низкого порядка, но их сложность достаточно быстро возрастает для кумулянт высокого порядка.

Случай правильной подгонки нелинейности возникает из замены в (13) слагаемых с псевдоэнтропией $-E\{\log(p_i(y_i))\}$ на функцию кумулянт выходного сигнала $|C_{1+q}(y_i)| / (1+q)$.

Рассмотрение этого случая приводит к функции стоимости (ФС) в виде

$$J(\mathbf{y}, \mathbf{W}) = (-1/2) \log|\det(\mathbf{W}\mathbf{W}^T)| - (1/(1+q)) \sum_{i=1}^n \{|C_{1+q}(y_i)|\}. \quad (14)$$

Минимизация этой ФС приводит к ненулевому решению для $\{y_i\}_{i=1,\dots,n}$ ввиду того, что матрица $\mathbf{W}\mathbf{W}^T \neq 0$ и второе слагаемое заставляют сигналы быть как можно более негауссовыми, т. к. кумулянты высокого порядка являются естественной мерой негауссовости сигнала (и равны нулю для гауссовых сигналов). С учетом определения и свойств кумулянт получается соотношение

$$\begin{aligned} \frac{\partial \text{Кум.}(\overbrace{y_i, \dots, y_i}^{1+q})}{\partial w_{kj}} &= \\ &= (1+q) \text{Кум.}(\overbrace{y_i, \dots, y_i}^q) \cdot \frac{\partial y_i}{\partial w_{kj}} = \\ &= (1+q) \cdot \delta_{ik} \text{Кум.}(\overbrace{y_i, \dots, y_i}^q, x_j). \end{aligned} \quad (15)$$

С использованием этого свойства кумулянт, аналитически выраженное в (15), Круцесом, Кастедо и Сичоки (Cruces, Castedo, Sichoeki) [22] получено соотношение, объединяющее размешивающую матрицу \mathbf{W} (для РСС), сигналы (\mathbf{x}) на сенсорах ИИС, оценку ПС (\mathbf{y}) и кумулянты этих сигналов:

$$\frac{\partial \sum_{i=1}^n |C_{1+n}(y_i)|}{\partial \mathbf{W}} = (1+q) \mathbf{S}_{q+1}(\mathbf{y}) \mathbf{C}_{q,1}(\mathbf{y}, \mathbf{x}), \quad (16)$$

где $\mathbf{S}_{q+1}(\mathbf{y}) = \text{sign}(\text{Диаг.}(C_{1,q}(\mathbf{y}, \mathbf{y})))$. Исходя из этого, поскольку

$$\frac{\partial \log|\det(\mathbf{W}\mathbf{W}^T)|}{\partial \mathbf{W}} = 2(\mathbf{W}\mathbf{W}^T)^{-1} \mathbf{W},$$

получается соотношение для градиента ФС, которое используется при получении ряда обучающих алгоритмов:

$$\frac{\partial J(\mathbf{y}, \mathbf{W})}{\partial \mathbf{W}} = -2(\mathbf{W}\mathbf{W}^T)^{-1} \mathbf{W} + \mathbf{S}_{q+1}(\mathbf{y}) \mathbf{C}_{q,1}(\mathbf{y}, \mathbf{x}). \quad (17)$$

Для улучшения сходимости итеративного процесса обучения НС вместо минимизации ФС обычным методом градиентного спуска может быть использована формула Атика—Редлиха (Atick, Redlich) (18), которая приводит к правилу

обучения (19)¹⁰ [22]:

$$\begin{aligned} \Delta \mathbf{W} &= -\eta \mathbf{W} \left[\frac{\partial J(\mathbf{y}, \mathbf{W})}{\partial \mathbf{W}} \right]^T \mathbf{W} = \\ &= \eta \mathbf{W} \left[(\mathbf{W} \mathbf{W}^T)^{-1} \mathbf{W} - \mathbf{S}_{1+q}(\mathbf{y}) \mathbf{C}_{q,1}(\mathbf{y}, \mathbf{x}) \right]^T \mathbf{W}, \end{aligned} \quad (18)$$

$$\begin{aligned} \Delta \mathbf{W}(l) &= \mathbf{W}(l+1) - \mathbf{W}(l) = \\ &= \eta(l) \left[\mathbf{I} - \mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y}) \mathbf{S}_{1+q}(\mathbf{y}) \right] \mathbf{W}(l). \end{aligned} \quad (19)$$

Умножение (19) справа на смешивающую матрицу \mathbf{H} позволяет получить алгоритм обновления общей матрицы \mathbf{G} системы РСС ($\mathbf{G} = \mathbf{W}\mathbf{H}$):

$$\Delta \mathbf{G}(l) = \eta(l) \left[\mathbf{I} - \mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y}) \mathbf{S}_{1+q}(\mathbf{y}) \right] \mathbf{G}(l). \quad (20)$$

Затем если в (20) применить так называемое неравенство треугольника [23]

$$\begin{aligned} \|\mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y}) \mathbf{S}_{1+q}(\mathbf{y}) - \mathbf{I}\| &\leq 1 + \|\mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y}) \mathbf{S}_{1+q}(\mathbf{y})\| = \\ &= 1 + \|\mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y})\|, \end{aligned}$$

то для численной устойчивости алгоритма (20) достаточно использовать ограничение параметра скорости обучения в виде $\eta(l) < \frac{1}{1 + \|\mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y})\|}$.

При значении $q = 3$ алгоритм (19) позволяет работать с симметрично и несимметрично распределенными сигналами. В этом случае вид алгоритма определяется соотношением

$$\begin{aligned} \mathbf{W}(l+1) &= \\ &= \mathbf{W}(l) + \eta(l) \left[\mathbf{I} - \mathbf{C}_{1,3}(\mathbf{y}, \mathbf{y}) \mathbf{S}_4(\mathbf{y}) \right] \mathbf{W}(l) = \\ &= \mathbf{W}(l) + \eta(l) \left[\mathbf{I} - \langle \mathbf{y} \mathbf{g}^T(\mathbf{y}) \rangle \right] \cdot \mathbf{W}(l), \end{aligned} \quad (21)$$

где

$$\mathbf{C}_{1,3}(\mathbf{y}, \mathbf{y}) = \langle \mathbf{y}(\mathbf{y} \circ \mathbf{y} \circ \mathbf{y})^T \rangle - 3 \langle \mathbf{y} \mathbf{y}^T \rangle \text{Диаг.} \{ \langle \mathbf{y} \circ \mathbf{y} \rangle \}^{11}$$

Диагональный элемент матрицы $\mathbf{S}_4(\mathbf{y})$ определен в

¹⁰ Для отдельных значений q алгоритм (19) переходит в (полученные ранее) частные алгоритмы. Так, например, при $q = 1$ — это алгоритм декорреляции Альмейда и Силва (Almeida, Silva) [24]: $\mathbf{W}(k+1) = \mathbf{W}(k) + \eta(k) \times$

$\times [\mathbf{I} - \langle \mathbf{y} \mathbf{y}^T \rangle] \cdot \mathbf{W}(k)$; при $q = 2$ — это алгоритм Сичоки и Унбехауна (Cichocki, Unbehauen) [25] для разделения сигналов с несимметричным распределением:

$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta(k) [\mathbf{I} - \langle \mathbf{y} \cdot \mathbf{g}(\mathbf{y}) \rangle] \cdot \mathbf{W}(k)$, компоненты вектора $\mathbf{g}(\mathbf{y})$ определяются нелинейными функциями $g(y_i) = \text{sign}(y_i) |y_i|^2$.

¹¹ Символом \circ обозначается кронекерово произведение матриц и векторов (являющихся матрицами с одним столбцом). Матрица $\mathbf{C} = \mathbf{A} \circ \mathbf{B}$ для кронекерова произведения состоит из блоков $[\mathbf{C}_{ik}] = a_{ik} \mathbf{B}$.

в виде $[\mathbf{S}_4(\mathbf{y})]_{ij} = \text{sign}(E\{y_j^4\} - 3(E\{y_j^2\})^2) = \text{sign } k_4(y_j)$.

Алгоритм (21), представленный в реальном времени, имеет вид

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta(k) [\mathbf{I} - \mathbf{R}_{\mathbf{y}\mathbf{g}}^{(k)}] \cdot \mathbf{W}(k), \quad (22)$$

где $\mathbf{R}_{\mathbf{y}\mathbf{g}}^{(k)} = (1 - \eta_0) \mathbf{R}_{\mathbf{y}\mathbf{g}}^{(k-1)} + \eta_0 \mathbf{y}(k) \mathbf{g}^T(\mathbf{y}(k))$ при $g(y_j) = (y_j^2 - 3y_j \hat{\sigma}_{yy}^2) \cdot \text{sign}(k_4(y_j))$ и $\hat{\sigma}_{yy}^2(k) = (1 - \eta_0) \times \hat{\sigma}_{yy}^2(k-1) + \eta_0 y_j^2(k)$.

Один из недостатков ФС (14) и связанного с ней алгоритма (19) состоит в том, что они не могут быть использованы, когда кумулянта $\mathbf{C}_{1+q}(y_i)$ для некоторого сигнала y_i равна нулю (или очень близка к нему). Тогда нужно использовать кумулянты различных порядков и найти набор Ω индексов ($\Omega = \{q_1, \dots, q_{N\Omega}\}$), для которого сумма $\sum_{q \in \Omega} |\mathbf{C}_{1+q}(y_i)|$ (для всех сигналов) не равна нулю. При этом слагаемое $\mathbf{C}_{1+q}(y_i)/(1+q)$ в (14) заменяется взвешенной суммой кумулянт выходных сигналов, индексы которых входят в набор Ω , т. е.

$J(\mathbf{y}, \mathbf{W}) = \sum_{q \in \Omega} \alpha^q \frac{|\mathbf{C}_{1+q}(y_i)|}{1+q}$, а весовые коэффициенты таковы, что $\sum_{q \in \Omega} \alpha^q = 1$.

Так же, как выше, получается обучающий алгоритм [22]

$$\Delta \mathbf{W}(l) = \eta(l) \left[\mathbf{I} - \sum_{q \in \Omega} \alpha_q \mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y}) \mathbf{S}_{1+q}(\mathbf{y}) \right] \mathbf{W}(l), \quad (23)$$

где для параметра скорости обучения η выполняется ограничение

$$\begin{aligned} \eta(k) &= \\ &= \min \left\{ \frac{2\eta_0}{1 + \eta_0 \sum_{q \in \Omega} \alpha_q}, \frac{\eta_0}{1 + \eta_0 \left\| \sum_{q \in \Omega} \alpha_q \mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y}) \mathbf{S}_{1+q}(\mathbf{y}) \right\|} \right\}; \\ \eta_0 &\leq 1 \quad \text{и} \quad \sum_{q \in \Omega} \alpha = 1. \end{aligned}$$

Алгоритм РСС при количестве сенсоров более числа первичных сигналов

Для рассмотренных выше алгоритмов на основе (19), робастных относительно гауссовых шумов, показана устойчивость с использованием общей передаточной матрицы \mathbf{G} . Полученные результаты справедливы при квадратной матрице \mathbf{G} , но это соответствует не только равному числу ПС (n) и сенсоров ИИС (m), но и случаю, когда $m > n$.

Если алгоритм (20), использующий матрицу \mathbf{G}

$$\mathbf{G}(l+1) = [\mathbf{I} + \eta(l)(\mathbf{I} - \mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y})\mathbf{S}_{1+q}(\mathbf{y}))]\mathbf{G}(l),$$

умножить справа на псевдообратную матрицу \mathbf{H}^+ (относительно смешивающей матрицы \mathbf{H}).¹²⁾

$$\mathbf{H}^+ = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T,$$

то будет получено соотношение

$$\begin{aligned} \mathbf{W}(l+1)\mathbf{P}_H &= \\ &= [\mathbf{I} + \eta(l)(\mathbf{I} - \mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y})\mathbf{S}_{1+q}(\mathbf{y}))]\mathbf{W}(l)\mathbf{P}_H, \end{aligned} \quad (24)$$

где \mathbf{P}_H — матрица проектирования на пространство, определяемое столбцами матрицы \mathbf{H} . Этот алгоритм определен с использованием (вместо $\mathbf{W}(l+1)$) разделяющей матрицы $\mathbf{W}(l+1)\mathbf{P}_H$. Если опустить проекцию \mathbf{P}_H ,¹³⁾ то получим ту же самую форму алгоритма для неквадратной разделяющей матрицы \mathbf{W} размера $(n \times m)$:

$$\begin{aligned} \mathbf{W}(l+1) &= \\ &= [\mathbf{I} + \eta(l)(\mathbf{I} - \mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y})\mathbf{S}_{1+q}(\mathbf{y}))]\mathbf{W}(l). \end{aligned} \quad (25)$$

Аналогичным образом получается похожий алгоритм для оценки смешивающей матрицы \mathbf{H} размера $(m \times n)$:

$$\hat{\mathbf{H}}(l+1) = \hat{\mathbf{H}}(l) - \eta(l)[\hat{\mathbf{H}}(l) - \mathbf{C}_{1,q}(\mathbf{x}, \mathbf{y})\mathbf{S}_{1+q}(\mathbf{y})],$$

или эквивалентно

$$\hat{\mathbf{H}}(l+1) = \hat{\mathbf{H}}(l) - \eta(l)\hat{\mathbf{H}}(l)[\mathbf{I} - \mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y})\mathbf{S}_{1+q}(\mathbf{y})]. \quad (26)$$

Приведенные алгоритмы обладают важными свойствами.

1. Сходимость алгоритма (20) не зависит от распределения сигналов (при условии, что кумулянты $(q+1)$ порядка от них не равны нулю).

2. Присутствие в смеси аддитивных гауссовых шумов (при достаточном количестве отсчетов)¹⁴⁾ не изменяет свойства сходимости алгоритма, т. е. оцениваемая разделяющая матрица не имеет смещения, зависящего от шума [22]. Это свойство обусловлено использованием кумулянт высокого порядка (вместо нелинейной ФПН).

¹²⁾ Псевдообратная матрица $\mathbf{H}^+ = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$ является эквивалентом обратной матрицы для неквадратной матрицы \mathbf{H} . Если \mathbf{H} имеет размерность $(n \times m)$, то \mathbf{H}^+ — размерность $(m \times n)$ [26, 27].

¹³⁾ Эта операция (опускание \mathbf{P}_H) не влияет на компоненты выходного сигнала НС-системы РСС, т. к. $\mathbf{W}\mathbf{P}_H\mathbf{H} = \mathbf{I}_n$ соответствует $\mathbf{W}\mathbf{H} = \mathbf{I}_n$ и поэтому РСС все равно выполняется.

¹⁴⁾ При оценке кумулянт достаточным количеством отсчетов можно считать не менее 5000.

РОБАСТНОЕ ИЗВЛЕЧЕНИЕ ГРУППЫ СИГНАЛОВ ИЗ СМЕСИ

Рассмотренные выше алгоритмы могут быть модифицированы для случая, когда требуется извлечь из смеси только определенное количество e ПС, $1 \leq e \leq n$. При этом проанализированы два варианта:

1) алгоритм извлечения малого количества сигналов (составляющих небольшую долю всех ПС и имеющих наибольшее значение эксцесса) с использованием отбеливания и ограничения на ортогональность;

2) алгоритм извлечения произвольной группы сигналов без использования процедуры отбеливания.

Алгоритм 1

Предположение, что сигналы (s_i) упорядочены по убыванию абсолютной величины эксцесса $k_4(s_j) = E\{s_j^4\} - 3E^2\{s_j^2\} \neq 0$, т. е. $|k_4(s_1)| \geq |k_4(s_2)| \geq \dots \geq |k_4(s_e)| \geq \dots \geq |k_4(s_n)|$, позволяет определить задачу извлечения из смеси e сигналов как оптимизацию, т. е. как минимизацию с ограничением (своеобразного критерия) функции стоимости.

Минимизировать

$J(\mathbf{y}) = -\sum_{j=1}^e |k_4(y_j)|$ при ограничении $\mathbf{W}_e \mathbf{W}_e^T = \mathbf{I}_e$, где \mathbf{W} — размешивающая матрица размера $(e \times n)$, $\mathbf{y}_e(k) = \mathbf{W}_e \bar{\mathbf{x}}(k)$, в предположении, что сигналы сенсоров $\bar{\mathbf{x}} = \mathbf{Q}\mathbf{x}$ отбелены или смешивающая матрица устроена таким путем, что матрица $\mathbf{A} = \mathbf{Q}\mathbf{H}$ является ортогональной.

В этом случае минимизация функции стоимости с таким ограничением приведет к извлечению e сигналов с наибольшей величиной эксцесса.

Обучающий алгоритм, построенный на основе использования м-градиента для минимизации ФС, имеет вид $\Delta \mathbf{W}_e(l) = -\eta[\nabla \mathbf{w}_e J - \mathbf{W}_e(\nabla \mathbf{w}_e J)^T \mathbf{W}_e]$ и при выбранной форме ФС сводится к соотношению

$$\begin{aligned} \mathbf{W}_e(l+1) &= \mathbf{W}_e(l) + \\ &+ \eta[\mathbf{S}_4(\mathbf{y})\mathbf{C}_{3,1}(\mathbf{y}, \bar{\mathbf{x}}) - \mathbf{C}_{1,3}(\mathbf{y}, \mathbf{y})\mathbf{S}_4(\mathbf{y})\mathbf{W}_e(l)], \end{aligned} \quad (27)$$

где $\mathbf{S}_4(\mathbf{y})$ — диагональная матрица с элементами $s_{ii} = [\text{sign}(k_4(y_i))]_{ii}$ (знак эксцесса выходных сигналов); $\mathbf{C}_{1,3}(\mathbf{y}, \mathbf{y})$ — матрица кросс-кумулянт четвертого порядка с элементами ij вида

$$[\mathbf{C}_{1,3}(\mathbf{y}, \mathbf{y})]_{ij} = \text{Кум.}(y_i(k), y_j(k), y_j(k), y_j(k));$$

матрица кросс-кумулянт $\mathbf{C}_{3,1}(\mathbf{y}, \bar{\mathbf{x}})$ определена соотношением $\mathbf{C}_{3,1}(\mathbf{y}, \bar{\mathbf{x}}) = \mathbf{C}_{1,3}(\bar{\mathbf{x}}, \mathbf{y})$.

Для $e = n$ (т. е. при извлечении всего набора ПС) алгоритм (27) принимает вид

$$\begin{aligned} \mathbf{W}(l+1) &= \\ &= \mathbf{W}(l) + \eta [\mathbf{S}_4(\mathbf{y})\mathbf{C}_{3,1}(\mathbf{y}, \mathbf{y}) - \mathbf{C}_{1,3}(\mathbf{y}, \mathbf{y})\mathbf{S}_4(\mathbf{y})\mathbf{W}(l)]. \quad (27^*) \end{aligned}$$

Алгоритм (27) может выполняться в режиме реального времени (скользящего среднего):

$$\mathbf{W}_e(k+1) = \mathbf{W}_e(k) + \eta [\mathbf{R}_{\mathbf{g}\bar{\mathbf{x}}}^{(k)} - \mathbf{R}_{\mathbf{y}\mathbf{g}}^{(k)}\mathbf{W}_e(k)], \quad (28)$$

где

$$\mathbf{R}_{\mathbf{g}\bar{\mathbf{x}}}^{(k)} = (1 - \eta_0)\mathbf{R}_{\mathbf{g}\bar{\mathbf{x}}}^{(k-1)} + \eta_0\mathbf{g}(\mathbf{y}(k))\bar{\mathbf{x}}^T(k);$$

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

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

при $g(y_i) = (y_i^3 - 3 y_i E\{y_i^2\})\text{sign}(k_4(y_i))$.

Алгоритм 2

Показано, что смешивающая матрица \mathbf{H} может быть оценена (с использованием кумулянт) на основе обучающего правила (26)

$$\hat{\mathbf{H}}(l+1) = \hat{\mathbf{H}}(l) - \eta(l)\hat{\mathbf{H}}(l)[\mathbf{I} - \mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y})\mathbf{S}_{1+q}(\mathbf{y})].$$

Согласно Круцесу [22], это обучающее правило может быть представлено в более общей форме, описывающей оценку смешивающей матрицы для случая выделения из смеси только e сигналов:

$$\begin{aligned} \hat{\mathbf{H}}_e(l+1) &= \hat{\mathbf{H}}_e(l) - \\ &- \eta(l)\hat{\mathbf{H}}_e(l)[\mathbf{I} - \mathbf{C}_{1,q}(\mathbf{y}, \mathbf{y})\mathbf{S}_{1+q}(\mathbf{y})], \quad (29) \end{aligned}$$

где $\mathbf{y} = \mathbf{W}_e\mathbf{x}$; $\hat{\mathbf{H}}_e = [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \dots, \hat{\mathbf{h}}_e]$ — матрица, составленная из первых e векторов оцениваемой матрицы $\hat{\mathbf{H}}$. Это означает, что может оцениваться произвольное число столбцов матрицы $\hat{\mathbf{H}}$ без знания других ее столбцов (при условии, что вся разделяющая матрица \mathbf{W} также может быть оценена).

Для того чтобы извлечь e сигналов, можно минимизировать среднеквадратичную ошибку между оцениваемым сигналом $\hat{\mathbf{s}}_e$ и выходом НС $\mathbf{y} = \mathbf{W}_e\mathbf{x}$. Но это эквивалентно минимизации общей мощности оцениваемых выходных сигналов при ограничении $\mathbf{W}_e(l)\hat{\mathbf{H}}_e(l) = \mathbf{I}_e$:

$$\min_{\mathbf{W}_e} \text{tr} \left\{ \mathbf{R}_{\mathbf{y}\mathbf{y}} \right\} \quad \text{при} \quad \mathbf{W}_e(l)\hat{\mathbf{H}}_e(l) = \mathbf{I}_e. \quad (30)$$

Метод решения задачи минимизации с ограничением основан на использовании множителей

Лагранжа [29] и, как показано Круцесом (Cruces et al. [30]), приводит к соотношениям:

$$\begin{aligned} \mathbf{y}(k) &= \mathbf{W}(l)\mathbf{x}(k) = \\ &= [\hat{\mathbf{H}}_e^T(l)\mathbf{R}_{\mathbf{v}\mathbf{v}}^+\hat{\mathbf{H}}_e(l)]^{-1}\hat{\mathbf{H}}_e^T(l)\mathbf{R}_{\mathbf{v}\mathbf{v}}^+\mathbf{x}(k), \quad (31) \end{aligned}$$

$$\begin{aligned} \hat{\mathbf{H}}_e(l+1) &= \hat{\mathbf{H}}_e(l) - \\ &- \eta(l) [\hat{\mathbf{H}}_e(l) - \mathbf{C}_{1,q}(\mathbf{x}, \mathbf{y}) + \mathbf{S}_{1+q}(\mathbf{y})], \quad (32) \end{aligned}$$

где $\mathbf{R}_{\mathbf{v}\mathbf{v}}^+ - (m \times n)$ — матрица, псевдообратная относительно оцененной ковариационной матрицы шумов, при этом $m \geq n$ и $1 \leq e \leq n$. Чтобы извлечь из смеси желаемое количество сигналов требуется попеременно выполнять итерации по соотношениям (31) и (32). Эти соотношения представляют суть робастного алгоритма РСС, с помощью которого можно восстановить произвольное количество $e < n$ сигналов без предварительного выполнения операции отбеливания. Преимуществом алгоритма является нечувствительность его к гауссовым шумам (при условии, что их ковариационная матрица известна или может быть надежно оценена). Теневой стороной алгоритма является то, что для него требуется вычисление псевдообратной матрицы и ковариационной матрицы шумов.

СИСТЕМА УНИЧТОЖЕНИЯ ШУМОВ ПРИ РСС НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ (НС)

Установленные выше алгоритмы дают несмещенную оценку разделяющей матрицы \mathbf{W} , что позволяет получать смешивающую матрицу $\hat{\mathbf{H}} = \mathbf{W}^+ = \mathbf{H}\mathbf{P}\mathbf{D}$, где \mathbf{W}^+ — псевдообратная относительно \mathbf{W} ; \mathbf{P} — $(n \times n)$ -матрица перестановок (любая); \mathbf{D} — диагональная масштабирующая $(n \times n)$ -матрица.

Получение системы на НС для уничтожения шума в оценках ПС включает два этапа:

а) вводится вектор ошибки $\mathbf{e}(t) = \mathbf{x}(t) - \hat{\mathbf{H}}\hat{\mathbf{y}}(t)$, где $\mathbf{e}(t) = [e_1(t), e_2(t), \dots, e_m(t)]^T$; $\hat{\mathbf{y}}(t)$ — оценка ПС $\mathbf{s}(t)$;

б) для получения $\hat{\mathbf{y}}(t)$ минимизируется ФС в виде энтропии $E\{J(\mathbf{e}(t))\} = -\sum_{i=1}^m E\{\log[p_i(e_i(t))]\}$, где $p_i(e_i)$ — плотность распределения вероятности шума $v_i(t)$, и предполагается независимый и одинаково распределенный шум в ПС.

Минимизация ФС осуществляется методом градиентного спуска, приводит к независимости компонент ошибки и минимизации их величины. Соответствующая система уравнений имеет вид

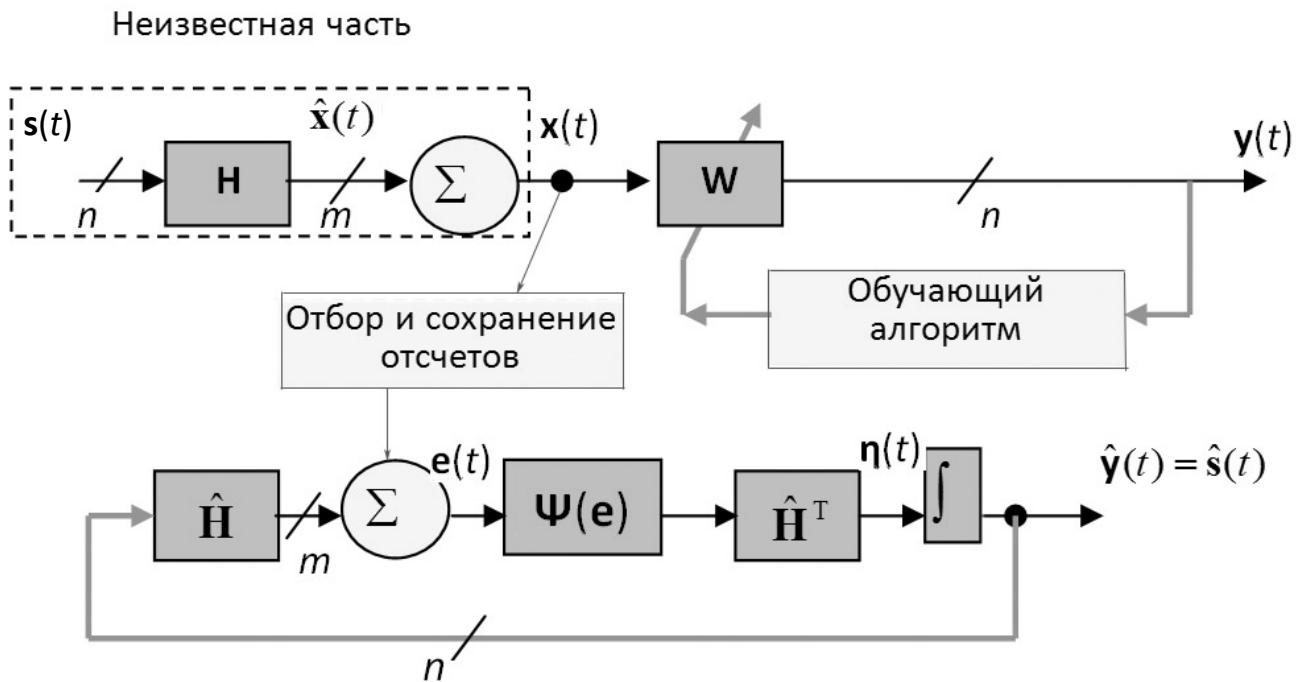


Рис. 3. Блок-схема системы для оценки разделяющей матрицы и ПС при обеспечении снижения шума

$$\frac{\partial \hat{y}(t)}{\partial t} = \eta(t) \hat{H}^T \Psi[e(t)], \quad (33)$$

где $\Psi[e(t)] = [\Psi_1[e_1(t), \dots, \Psi_m[e_m(t)]]^T$ с нелинейностями в виде $\Psi_i(e_i) = -\partial \log p_i(e_i) / \partial e_i$.

Блок-схема системы на основе НС, отражающая выполнение такого алгоритма, приведена на рис. 3, где в блоке "Обучающий алгоритм" использовано правило обучения (7), соответствующее удалению шумов.

В рассматриваемом алгоритме оптимальный выбор нелинейностей $\Psi_i(e_i)$ зависит от распределения шума. Для шума использовано экспоненциально-степенное распределение (ЭСР).¹⁵⁾ Выбор в ЭСР параметра $r_i \geq 1$ определяет нелинейную ФПН в виде:

¹⁵⁾ Плотность ЭСР имеет вид [9]:

$$p_i(e_i) = [r_i / (2\sigma_i \Gamma(1/r_i))] \exp[-|e_i|^{r_i} / (r_i \sigma_i)],$$

где r_i — переменный параметр; $\Gamma(r) = \int_0^\infty u^{r-1} \exp(-u) du$ — гамма-функция; $\sigma^r = E\{|e|^r\}$ — обобщенная мера дисперсии шума. При $r_i = 1$ ЭСР равноценно плотности распределения Лапласа, при $r_i = 2$ — обычному распределению Гаусса, при $r_i > 4-5$ ЭСР равноценно плотности равномерного распределения.

$$\begin{aligned} \Psi_i(e_i) &= -\frac{\partial \log(p_i(e_i))}{\partial e_i} = \\ &= \frac{1}{|\sigma_i|^{r_i}} |e_i|^{r_i-1} \text{sign}(e_i), \quad r_i \geq 1. \end{aligned} \quad (34)$$

Алгоритм (33) для дискретного времени имеет вид $\hat{y}(k+1) = \hat{y}(k) + \eta(k) \hat{H}^T(k) \Psi[e(k)]$ и показан на рис. 4. Он отражает систему РСС рекуррентной НС и нелинейную форму обработки, которая обеспечивает эффективное снижение аддитивных компонент шума в восстанавливаемых ПС.

Несколько более эффективная структура системы на основе НС показана на рис. 5. Система обеспечивает оценку смешивающей матрицы \mathbf{H} и одновременное снижение уровня шума у выделяемых из смеси первичных сигналов. В основе этой системы лежит рассмотренный выше алгоритм и эквивалентность $\hat{\mathbf{H}} = \mathbf{W}^+$. Для условия $m \geq n$ дифференцирование $\mathbf{W} \hat{\mathbf{H}} = \mathbf{W} \mathbf{W}^+ = \mathbf{I}$ ведет к соотношению $\frac{d\mathbf{W}}{dt} \hat{\mathbf{H}} + \mathbf{W} \frac{d\hat{\mathbf{H}}}{dt} = 0$, которое позволяет получить (как это сделано в [12] и [13]) обучающий алгоритм для получения оценки смешивающей матрицы $\hat{\mathbf{H}}$:

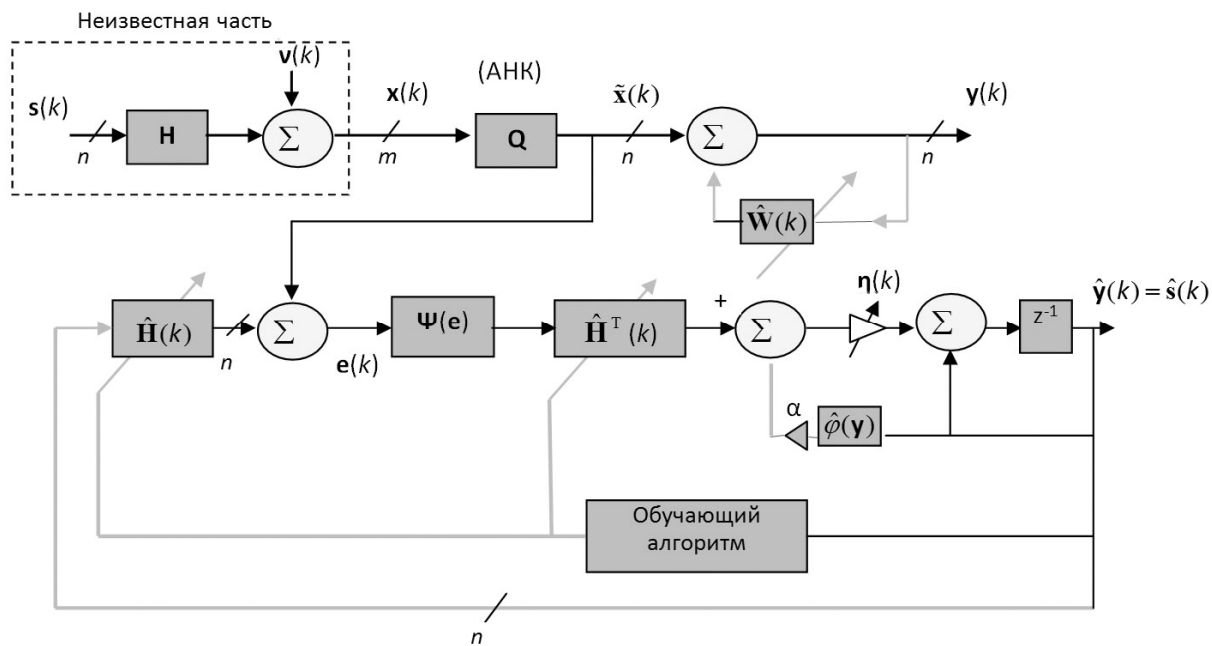


Рис. 4. Блок-схема системы на основе рекуррентной НС для оценки разделяющей матрицы и ПС при обеспечении снижения шума (модель дискретного времени)

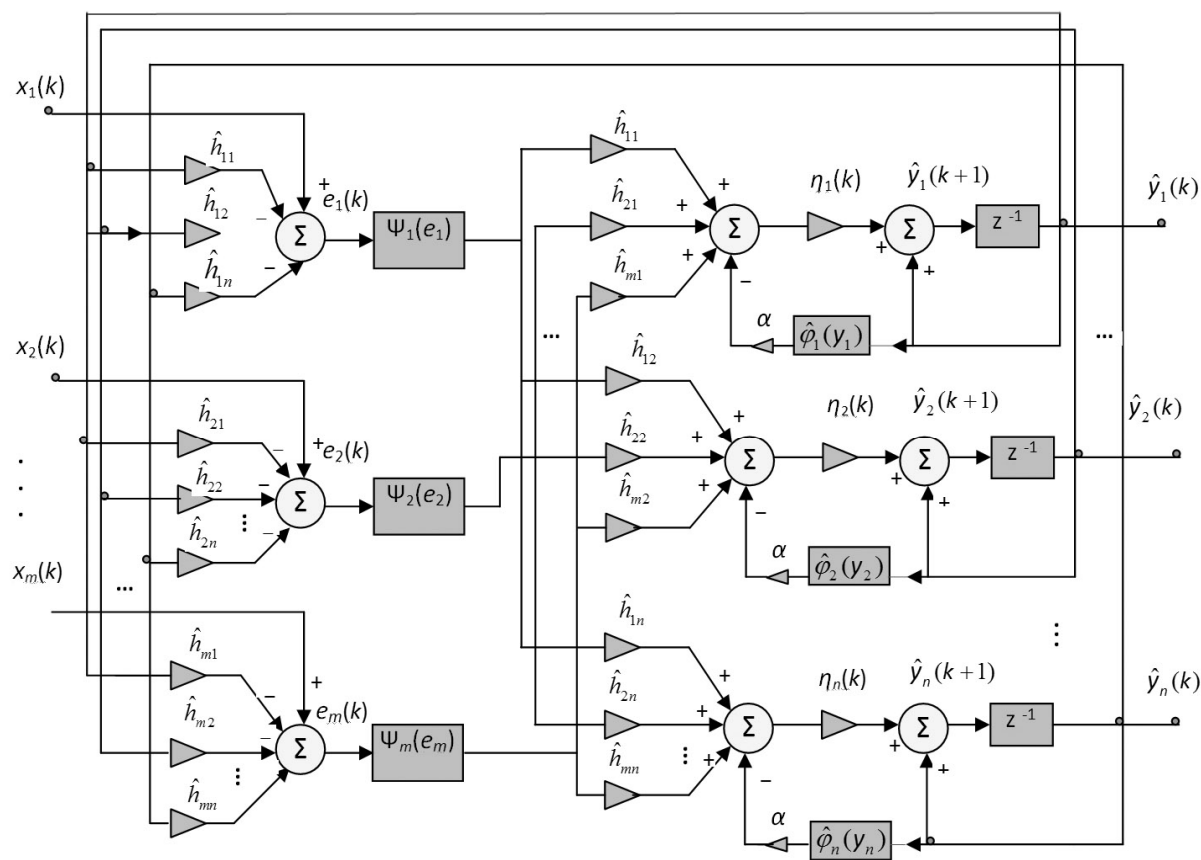


Рис. 5. Блок-схема системы для РСС на рекуррентной НС с робастным алгоритмом и регуляризацией

$$\begin{aligned} \frac{d\hat{\mathbf{H}}(t)}{dt} &= -\hat{\mathbf{H}} \frac{d\mathbf{W}}{dt} \hat{\mathbf{H}} = \\ &= -\mu_1(t) \hat{\mathbf{H}}(t) [\mathbf{\Lambda}(t) = \mathbf{f}[\mathbf{y}(t)] \mathbf{g}^T[\mathbf{y}(t)]], \end{aligned} \quad (35)$$

где $\mathbf{\Lambda} = \text{Диаг.}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ — диагональная матрица (в качестве которой можно использовать $\mathbf{\Lambda} = \mathbf{I}$ или $\mathbf{\Lambda} = \text{Диаг.}\{\mathbf{f}[\mathbf{y}(t)] \mathbf{y}^T(t)\}$); $\mathbf{f}[\mathbf{y}(t)] = [f_1(y_1), \dots, f_n(y_n)]^T$ и $\mathbf{g}[\mathbf{y}(t)] = [g_1(y_1), \dots, g_n(y_n)]^T$ — векторы нелинейных функций. Замена выходного вектора $\mathbf{y}(t)$ улучшенной оценкой $\hat{\mathbf{y}}(t)$ приводит к обучающему алгоритму в виде

$$\begin{aligned} \frac{d\hat{\mathbf{H}}(t)}{dt} &= -\mu_1(t) \hat{\mathbf{H}}(t) [\mathbf{\Lambda}(t) = \mathbf{f}[\hat{\mathbf{y}}(t)] \mathbf{g}^T[\hat{\mathbf{y}}(t)]] \\ \text{и } \frac{d\hat{\mathbf{y}}(t)}{dt} &= \mu(t) \hat{\mathbf{H}}^T(t) \mathbf{\Psi}[\mathbf{e}(t)]. \end{aligned} \quad (36)$$

В дискретной форме алгоритм (36) имеет вид:

$$\begin{aligned} \Delta \hat{\mathbf{H}}(k) &= \hat{\mathbf{H}}(k+1) - \hat{\mathbf{H}}(k) = \\ &= \eta_1(k) \hat{\mathbf{H}}(k) [\mathbf{\Lambda}(k) = \mathbf{f}[\hat{\mathbf{y}}(k)] \mathbf{g}^T[\hat{\mathbf{y}}(k)]] \\ \text{и } \hat{\mathbf{y}}(k+1) &= \hat{\mathbf{y}}(k) + \eta(k) \hat{\mathbf{H}}^T(k) \mathbf{\Psi}[\mathbf{e}(k)], \end{aligned}$$

где $\mathbf{e}(k) = \mathbf{x}(k) - \hat{\mathbf{H}}(k) \hat{\mathbf{s}}(k)$ и $\mathbf{x}(k) = \hat{\mathbf{x}}(k) + \mathbf{v}(k)$.

В отдельных случаях смешивающая матрица \mathbf{H} может быть плохо обусловленной (с очень большим разбросом собственных значений). Тогда, чтобы получить корректную оценку, следует выполнить отбеливание или использовать метод регуляризации (МР) [31, 32]. МР состоит в расширении ФС $J(\mathbf{e})$ — включении дополнительного слагаемого

$$J(\mathbf{e}) = \rho(\mathbf{x} - \hat{\mathbf{H}}\hat{\mathbf{y}}) + (\alpha / 2) \|\mathbf{L}\hat{\mathbf{y}}\|^2,$$

где $\mathbf{e} = \mathbf{x} - \hat{\mathbf{H}}\hat{\mathbf{y}}$; $\alpha > 0$ — константа регуляризации, контролирующая размер решения; а \mathbf{L} — матрица регуляризации, определяющая его норму (при этом обычно принимается $\mathbf{L} = \mathbf{I}$). Минимизация расширения ФС с $\mathbf{L} = \mathbf{I}$ приводит к обучающему правилу

$$\begin{aligned} \hat{\mathbf{y}}(k+1) &= \hat{\mathbf{y}}(k) + \eta(k) [\hat{\mathbf{H}}^T(k) \cdot \mathbf{\Psi}(\mathbf{x}(k) - \\ &\quad - \hat{\mathbf{H}}(k)\hat{\mathbf{y}}(k)) - \alpha \cdot \hat{\phi}(\hat{\mathbf{y}}(k))], \end{aligned} \quad (37)$$

где $\mathbf{\Psi}(\mathbf{e}) = [\Psi_1(e_1), \dots, \Psi_m(e_m)]^T$ с компонентами $\Psi_i(e_i) = \partial \rho / \partial e_i$; $\hat{\phi}(\hat{\mathbf{y}}(k)) = [\hat{\phi}_1(\hat{y}_1), \dots, \hat{\phi}_n(\hat{y}_n)]^T$ с нелинейностями $\hat{\phi}_i(\hat{y}_i) = |\hat{y}_i|^{p-1} \text{sign}(\hat{y}_i)$.

Комбинация (37) с дискретной формой (36) приводит к алгоритму, который обеспечивает одновременную оценку смешивающей матрицы и

ПС для плохообусловленных задач. Система, построенная на основе рекуррентной НС и реализующая этот алгоритм, показана на рис. 5.

Некоторое усовершенствование алгоритма, способствующее дополнительному улучшению обусловленности задачи РСС и снижению шума в восстанавливаемых оценках ПС, может быть достигнуто предварительной обработкой: робастным (без внесения дополнительного шума) отбеливанием или анализом главных компонент для измеренных (на сенсорах) сигналов. Этап предобработки отражен на рис. 5 ($n \times m$)-матрицей \mathbf{Q} . Отбеливание для зашумленных данных может быть выполнено с использованием обучающего алгоритма

$$\begin{aligned} \Delta \mathbf{Q}(k) &= \eta(k) [\mathbf{I} - \hat{\mathbf{x}}(k) \hat{\mathbf{x}}^T(k) + \\ &\quad + \hat{\sigma}_v^2 \mathbf{Q}(k) \mathbf{Q}^T(k)] \mathbf{Q}(k), \end{aligned} \quad (38)$$

где $\hat{\mathbf{x}}(k) = \mathbf{Q}\mathbf{x}(k) = \mathbf{Q} \cdot [\mathbf{H}\mathbf{s}(k) + \mathbf{v}(k)]$. При $m = n$ и несингулярной ковариационной матрице $\mathbf{R}_{xx} = E\{\mathbf{x}(k)\mathbf{x}^T(k)\} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ может использоваться алгоритм

$$\mathbf{Q} = \left[\text{Диаг.} \left\{ \frac{\lambda_1}{\lambda_1 + \hat{\sigma}_v^2} \dots \frac{\lambda_n}{\lambda_n + \hat{\sigma}_v^2} \right\} \right]^{1/2} \mathbf{V}^T,$$

где $\hat{\sigma}_v^2$ — оценка дисперсии шума; λ_i — собственные значения, а $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$ — ортогональная матрица соответствующих собственных векторов ковариационной матрицы \mathbf{R}_{xx} .

ЗАКЛЮЧЕНИЕ

Рассмотрены метод и алгоритмы разделения сигналов смеси (и восстановления формы ее компонент) для условий, когда измеренные сигналы искажены большими уровнями аддитивного шума. Существующие адаптивные алгоритмы разделения сигналов смеси (РСС) видоизменены таким образом, чтобы существенно снизить смещения оценок смешивающей \mathbf{H} и разделяющей \mathbf{W} матриц, обусловленные наличием шума. Этим достигается практическая устойчивость алгоритмов РСС (их робастность) к значительным уровням шумов в анализируемых сигналах смеси.

Система реализации робастных алгоритмов построена на основе нейронных сетей (НС) с рекуррентной структурой. Система обеспечивает одновременную оценку неизвестной смешивающей матрицы, восстановление первичных сигналов (компонент смеси) и снижение шума в них. Этим алгоритмы отличаются от рассмотренных ранее,

которые были ориентированы на отсутствие шумов или их пренебрежимо низкий уровень.

– Показано, каким образом выбирается оптимальная нелинейная функция преобразования нейронов (ФПН) при различных видах вероятностного распределения шумов.

– Для плотности вероятностного распределения шумов использовано семейство экспоненциально-степенных распределений. (Это семейство помимо гауссова включает распределения Лапласа, практически однородное распределение и (надгауссовы) распределения с большим значением эксцесса).

– Удаление смещений в оценках матриц \mathbf{H} и \mathbf{W} проанализировано в рамках методов отбеливания (преобразования анализируемых сигналов, ведущего к диагональной ковариационной матрице) и анализа независимых компонент (АНК).

– Показан метод РСС для случая, когда в первичных сигналах (компонентах смеси) присутствует шум в форме свертки. Дан алгоритм обучения НС (в составе системы РСС), который обеспечивает уничтожение шума. При этом предполагается, что доступен внешне контролируемый шум (ВКШ).

– Дана трактовка адаптивных алгоритмов АНК на основе статистик высокого порядка (кумулянт). Показана робастная форма алгоритма для извлечения из смеси группы сигналов.

– Приведены структуры систем РСС на основе НС для ряда вариантов: для ВКШ в форме свертки (при линейном и нелинейном преобразованиях) с идентификацией матрицы \mathbf{W} и компонент смеси; для идентификации неизвестной смешивающей матрицы (\mathbf{H}) и компонент смеси при подавлении шума. Рассмотрена структура на основе НС, которая обеспечивает оценку смешивающей матрицы \mathbf{H} и одновременное снижение уровня шума у выделенных из смеси первичных сигналов.

– Для плохо обусловленной задачи РСС (при смешивающей матрице с большим разбросом собственных чисел) дан алгоритм с регуляризацией и НС, соответствующая этому алгоритму.

– В представленном анализе методов и робастных алгоритмов РСС использованы работы Дугласа, Сичоки и Амари; Кульбака; Круцеса, Кастедо и Сичоки; Альмейды и Силвы, А.Н. Тихонова и В.Я. Арсенина.

СПИСОК ЛИТЕРАТУРЫ

1. *Al-Kindi M.J., Dunlop J.* Improved adaptive noise cancellation in presence of signal leakage on the noise reference channel // *Signal Processing*. 1989. V. 17. P. 241–250.
2. *Nguyen Thi H.L., Jutten C., Caelen J.* Speech enhancement: Analysis and comparison of methods on various real situations // *Signal Processing VI: Theory and Applications* / Eds. J. Vandewalle, R. Boite, M. Moonen, A. Oosterlinck, 1992. P. 303–306.
3. *Weinstein E., Oppenheim A.F., Feder M., Buck J.R.* Iterative and sequential algorithms for multi-sensor signal enhancement // *IEEE Transactions on Signal Processing*. 1994. V. 42, N 4. P. 846–859.
4. *Shamsunder S., Giannakis G.B.* Modeling of non-Gaussian array data using cumulants: DOA estimation of more sources with less sensors // *Signal Processing*. 1993. V. 30. P. 279–297.
5. *Karhunen J., Joutsensalo J.* Representation and separation of signals using nonlinear PCA type learning // *Neural Networks*. 1994. V. 7, N 1. P. 113–127.
6. *Малыхина Г.Ф., Меркушева А.В.* Сеть с симметричной функцией преобразования нейронов для подавления искажений и восстановления изображения // *Научное приборостроение*. 2008. Т. 18, № 2. С. 73–80.
7. *Меркушева А.В., Малыхина Г.Ф.* Методы и алгоритмы разделения смеси сигналов: Применение декорреляции и статистик второго порядка // *Научное приборостроение*. 2009. Т. 19, № 2. С. 90–103.
8. *Малыхина 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.
9. *Малыхина Г.Ф., Меркушева А.В.* Детектирование речевого сигнала и фильтрация с адаптивным порогом // *Сб. трудов ФТК СПб ГТУ: Микропроцессорные средства измерений*. Вып. 2. С.-Петербург, 2001. С. 26–35.
10. *Douglas S.C., Cichocki A.* Neural networks for decorrelation of signals // *IEEE Transactions on Signal Processing*. 1997. V. 45, N 11. P. 2829–2842.
11. *Douglas S.C., Cichocki A., Amari S.* A bias removal technique for source separation with noisy measurements // *Electronic Letters*. 1998. V. 34, N 14. P. 1379–1380.
12. *Малыхин В.М., Меркушева А.В.* Методы и алгоритмы разделения смеси сигналов: Применение м-градиента к анализу независимых компонент // *Научное приборостроение*. 2009. Т. 19, № 4. С. 83–95.
13. *Меркушева А.В., Малыхина Г.Ф.* Восстановление формы линейно смешанных сигналов на основе адаптивного алгоритма рекуррентной сети // *Научное приборостроение*. 2005. Т. 15, № 3. С. 94–107.
14. *Рабинер Л., Гоулд Б.* Теория и применение цифровой обработки сигналов. Пер. с англ. М.: Мир, 1978. 848 с.
15. *Бендат Дж., Пирсол А.* Прикладной анализ случайных данных. Перевод с англ. М.: Мир, 1989. 540 с.
16. *Уидроу Б., Стернз С.* Адаптивная обработка сигналов. Перевод с англ. М.: Радио и связь, 1989. 439 с.
17. *Gallant A.R., White H.* // *Neural Networks*. 1990. N 5. P. 128–138.
18. *Cichocki A., Kasprzak W.* Nonlinear learning algorithms for separation of natural images // *Neural Network World*. 1996. V. 6, N 4. P. 515–523.
19. *Кульбак С.* Теория информации и статистика. М.: Наука, 1967. 408 с.

20. *Малыхина Г.Ф., Меркушева А.В.* Концепции информации и энтропии в нейросетевых методах анализа сигналов ИИС // Информационные технологии. 2005. № 12. С. 26–36.
21. *Proakis J.G., Nikias C.L.* Blind equalization // Proceedings of SPIE. 1991. V. 1565. P. 76–87.
22. *Cruces S., Castedo L., Cichocki A.* // Proceedings of International Congress on Acoustic "Speech and Signal Processing'2000", Istanbul, Turkey, June, 2000. V. 5. P. 3152–3155.
23. *Голуб Дж., Ван Лоун У.* Матричные вычисления. М.: Мир, 1999. 549 с.
24. *Almeida L.B., Silva F.M.* Adaptive decorrelation // Artificial Neural Networks. Elsevier, 1992. N 2. P. 149–156.
25. *Cichocki A., Unbehauen R.* Robust neural networks with learning for separation and identification of sources // IEEE Transactions on Circuits and Systems. I: Fundamentals Theory and Applications. 1996. V. 43, N 11. P. 894–906.
26. *Воеводин В.В., Воеводин Вл.В.* Энциклопедия линейной алгебры. Электронная система ЛИНЕАЛ. СПб.: БХВ-Петербург, 2006. 541 с.
27. *Гантмахер Ф.Р.* Теория матриц. М.: Наука, 1966. 576 с.
28. *Amari S.* Natural gradient learning for over- and under-complete basis in ICA // Neural Computation. 1999. N 11. P. 1875–1883.
29. *Химмельблау Д.* Прикладное нелинейное программирование. М.: Мир, 1975. 532 с.
30. *Cruces S., Cichocki A., Castedo L.* Source extraction in Gaussian noise // Proceedings of the second International Workshop on ICA and BSS "ICA 2000", Helsinki, Finland, June 19–22, 2000. P. 63–68.
31. *Тихонов А.Н., Арсенин В.Я.* Методы решения некорректных задач. М.: Наука, 1979. 284 с.
32. *Тихонов А.Н., Гончарский А.В., Степанов В.В., Ягола А.Г.* Численные методы решения некорректных задач. М.: Наука, 1990. 232 с.

СПбГТУ, Санкт-Петербург

Контакты: *Малыхина Галина Фёдоровна,*
g_f_malychina@mail.ru

Материал поступил в редакцию 21.04.2010.

ROBUST METHODS FOR SIGNAL MIXTURES SEPARATION AND INDEPENDENT COMPONENTS ANALYSIS IN CONDITIONS OF NOISED DATA

G. F. Malychina, A. V. Merkusheva

SPbSPU, Saint-Petersburg

Robust methods and related algorithms based on neural networks for signal mixture separation are considered. Form reconstruction for original signals, identification of mixture and separation matrixes are implemented practically without of errors for conditions of noise significant levels in the mixtures of original signals.

Keywords: signal mixture, noising, components reconstruction, robust methods, algorithms, neural networks, mixture matrix reconstruction