

УДК 681.51; 621.391; 519.21

© А. В. Меркушева, Г. Ф. Малыхина

МЕТОД ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ПРОЦЕДУР АНАЛИЗА НЕЗАВИСИМЫХ КОМПОНЕНТ И ОБРАЩЕНИЯ СВЕРТКИ ПРИ ВОССТАНОВЛЕНИИ ФОРМЫ СИГНАЛОВ ПО ИЗМЕРЕНИЮ ИХ СМЕСИ

Рассмотрена концепция, порождающая единую точку зрения на различные алгоритмы разделения смеси сигналов (РСС), включая операцию обращения многоканальной свертки первичных сигналов. Анализируется метод оценивающих функций, позволяющий объяснить структуру различных адаптивных алгоритмов, реализующих оценку размешивающей матрицы и восстановление формы сигналов, исходя из измерений их смеси. Различие большинства алгоритмов связано с выбором оценивающих функций. Задача РСС сформулирована на основе полупараметрической статистической модели и семейства оценивающих функций. На основе понятий теории группы Ли анализируется геометрическая структура множества фильтров.

Кл. сл.: смеси сигналов, многоканальная свертка, разделение сигналов, алгоритмы, критерии эффективности, размешивающая матрица, группа Ли, КИХ-фильтры, геометрические структуры

ВВЕДЕНИЕ

Ранее рассмотрены методы разделения смеси сигналов, анализа независимых компонент, а также процедуры, обратной относительно смешивания сигналов, выполненного на основе совместной свертки [1–5]. Проанализированы как алгоритмы реального времени (РВ), так и алгоритмы обучения с формированием обучающей выборки. Большая часть алгоритмов РВ методом простого усреднения может быть преобразована в алгоритмы с обучающей выборкой. Некоторые алгоритмы позволяют извлекать из смеси все компоненты одновременно (параллельно), другие — выделяют первичные сигналы (ПС) смеси один за другим (последовательно).

Эта статья отражает поиск единой концепции, которая позволит объяснить большинство алгоритмов со статистической точки зрения. В ней введен метод оценивающих функций, который объясняет общие структуры в большинстве существующих алгоритмов разделения смеси сигналов (РСС). Для этой цели использована геометрия информации и определены оценивающие функции (ОФ) в полупараметрических статистических моделях, которые включают неизвестные функции как параметры (Цханг, Амари, Сичоки (Zhang, Amari, Sichocki) [6]). Различие в большей части существующих алгоритмов РСС состоит только в выборе оценивающих функций. Ниже в терминах ОФ приведен анализ точности алгоритмов и анализ их устойчивости. Это позволяет конструировать различные адаптивные методы для выбора

неизвестных параметров, включенных в ОФ и контролирующие точность и устойчивость. При этом метод Ньютона выводится автоматически с помощью стандартизованных оценивающих функций.

В рамках полупараметрической модели и семейства ОФ будет сформулирована стандартная задача РСС (или задача анализа независимых компонент (АНК)). После этого для задач РСС при линейной смеси и при смеси в форме свертки обсуждается и получает дальнейшее расширение вопрос о сходимости и эффективности оценивания при обучении нейронной сети (НС) методом обучающей выборки и методом м-градиента. На основе структуры типа группы Ли представлены геометрические свойства многообразия КИХ-фильтров и сформулирована задача восстановления сигналов при их взаимном смешивании в форме свертки. Затем анализируется эффективность оценивания при обучении с использованием выборки и использовании м-градиента.

ОЦЕНИВАЮЩИЕ ФУНКЦИИ ДЛЯ МЕТОДА АНАЛИЗА НЕЗАВИСИМЫХ КОМПОНЕНТ

Понятие оценивающей функции удобно ввести на простой модели смешивания сигналов:

$$\mathbf{x}(k) = \mathbf{H}\mathbf{s}(k) + \mathbf{v}(k),$$

где \mathbf{H} — неизвестная смешивающая матрица; s_i и s_j — независимые компоненты сигнала $\mathbf{s}(k)$ (каждая из которых может быть коррелированной по

времени); число первичных сигналов (т. е. размерность $\mathbf{s}(k)$) равно n ; число детекторов информационно-измерительной системы (ИИС), регистрирующих сигналы смеси, в общем случае равно m , так же как размерность сигнала смеси \mathbf{x} . В анализируемой здесь системе РСС принимается, что $m = n$, а вектор шума \mathbf{v} пренебрежимо мал. Если матрица \mathbf{W} (размерности $n \times n$) является предполагаемой размешивающей матрицей — $\mathbf{y}(k) = \mathbf{W}\mathbf{x}(k)$ (т. е. $\mathbf{y}(k)$ — вектор-сигнал такой, что его компоненты $y_i(k)$, $i = 1, \dots, n$ — независимые случайные ПС), — то при обучении \mathbf{W} обновляется по правилу

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta \mathbf{F}(\mathbf{x}(k), \mathbf{W}(k)), \quad (1)$$

где η — скорость обучения (которая может зависеть от k); $\mathbf{F}(\mathbf{x}, \mathbf{W}) \in \mathbf{R}^{n \times n}$; $\mathbf{F}(\mathbf{x}, \mathbf{W})$ — матрица-функция такая, что $\mathbf{W}(k)$ сходится к истинному решению. Обычно \mathbf{F} зависит от \mathbf{x} через $\mathbf{y}(k) = \mathbf{W}\mathbf{x}(k)$ и при этом имеет форму $\mathbf{F}(\mathbf{y})\mathbf{W}$, как в случае с m -градиентом [7].

Предложены различные виды \mathbf{F} , которые во многих (но не во всех) случаях получены как градиент функции стоимости — градиент критерия, который должен минимизироваться. Функциями стоимости могут быть, например, кумулянты высокого порядка, энтропия, отрицательный логарифм функции правдоподобия. Во многих случаях алгоритмы включают свободные параметры, а иногда свободные функции, которые должны быть адекватно выбраны или определены адаптивно. Поскольку функция плотности вероятности (ФПВ) распределения первичных сигналов, образующих смесь, обычно неизвестна, то нет возможности обойтись без использования таких свободных параметров.

Существуют условия, которым должна удовлетворять функция \mathbf{F} для того, чтобы алгоритм сошелся к истинному решению. Истинное \mathbf{W} должно быть точкой равновесия динамического соотношения (1). Но (1) является стохастическим дифференциальным уравнением, поэтому для математического анализа более удобно использовать его форму с непрерывным временем:

$$\frac{d}{dt} \mathbf{W}(t) = \mu \cdot \mathbf{F}[\mathbf{x}(t), \mathbf{W}(t)]. \quad (2)$$

И поскольку $\mathbf{x}(t)$ — стохастический процесс, его ожидаемая величина определена соотношением

$$\frac{d}{dt} \mathbf{W}(t) = \mu \cdot \mathbf{E} \{ \mathbf{F}[\mathbf{x}(t), \mathbf{W}(t)] \}. \quad (3)$$

Условием, что истинное решение \mathbf{W} определяет равновесие в (3), является соотношение (4)

$$\mathbf{E} \{ \mathbf{F}(\mathbf{x}, \mathbf{W}) \} = 0, \quad (4)$$

где математическое ожидание \mathbf{E} берется по $\mathbf{x} = \mathbf{H}\mathbf{s}$.

Функция $\mathbf{F}(\mathbf{x}, \mathbf{W})$, удовлетворяющая (4) для истинной (желаемой) матрицы \mathbf{W} и условию $\mathbf{E} \{ \mathbf{F}(\mathbf{x}, \mathbf{W}') \} \neq 0$ для ложной матрицы \mathbf{W}' , называется оценивающей функцией. Это определение введено по отношению к полупараметрической статистической модели. Теория оценивающих функций приводит к выводу, что обучение с выборкой (при достаточной ее величине) дает сходимость к истинному решению. Кроме того, преимущество использования полупараметрической модели состоит в том, что для задачи РСС она позволяет не оценивать мешающие параметры, т. е. ФПВ первичных сигналов, образующих смесь.

Полупараметрическая статистическая модель

Задача РСС может быть сформулирована в рамках ее статистической интерпретации. Если фактическая ФПВ сигнала s_i равна $r_i(s_i)$ и компоненты s_i линейно независимы, то совместная функция плотности вероятности (ФПВ) вектора-сигнала \mathbf{s} определяется соотношением

$$r(\mathbf{s}) = \prod_{i=1}^n r_i(s_i). \quad (5)$$

Вектор наблюдений \mathbf{x} является линейной функцией \mathbf{s} (т. е. $\mathbf{x} = \mathbf{H}\mathbf{s}$ и $\mathbf{s} = \mathbf{H}^{-1}\mathbf{x} = \mathbf{W}\mathbf{x}$), поэтому выражение ФПВ \mathbf{x} через $\mathbf{W} = \mathbf{H}^{-1}$ имеет вид:

$$p(\mathbf{x}, \mathbf{W}, \mathbf{r}) = \det |\mathbf{W}| \cdot \mathbf{r}(\mathbf{W}\mathbf{x}). \quad (6)$$

Поскольку ФПВ \mathbf{r} неизвестна (кроме того, что она удовлетворяет (5)), то модель вероятности (6) для \mathbf{x} включает два параметра: \mathbf{W} , который требуется оценить, и неизвестный параметр-функцию $\mathbf{r} = r_1 \dots r_n$ ("мешающий параметр"), о котором можно пока не заботиться. Такую статистическую модель, включающую (бесконечную) степень свободы в виде функции, можно назвать полупараметрической. В общем случае оценка \mathbf{W} — достаточно трудная задача именно из-за неизвестной функции \mathbf{r} . Однако преимущество использования полупараметрической модели состоит в том, что для задачи РСС (включая смеси со сверткой) она позволяет не оценивать мешающие параметры, т. е. ФПВ первичных сигналов, образующих смесь.

Метод получения оценок с помощью анализа полупараметрической статистической модели использует основные понятия концепции информационной геометрии.¹⁾ Последующее изложение построено на выводах и следствиях полупарамет-

¹⁾ Модель, концепция и методы информационной геометрии разработаны Амари, Дугласом и Сичоки (Amari, Douglas, Cichoci [8]), Каванобе (Kawanabe [9]) и Нагаокой (Nagaoka [10]).

рической статистической модели, концепции информационной геометрии и прямых результатов работ, указанных в приведенной выше ссылке.

Оценивающей функцией является функция (с матричным значением) $F(\mathbf{x}, \mathbf{W}) = |F_{ab}(\mathbf{x}, \mathbf{W})|$ от аргументов \mathbf{x} и \mathbf{W} , не включающая мешающего параметра r . Эта функция удовлетворяет соотношениям:

$$1) E_{\mathbf{w},r}\{F(\mathbf{x}, \mathbf{W}')\} = 0 \text{ при } \mathbf{W}' = \mathbf{W}, \quad (7)$$

$$2) E_{\mathbf{w},r}\{F(\mathbf{x}, \mathbf{W}')\} \neq 0 \text{ при } \mathbf{W}' \neq \mathbf{W}, \quad (8)$$

где $E_{\mathbf{w},r}$ — символ математического ожидания, определяемого по плотности вероятности (6). Кроме того, накладывается требование, чтобы (7) выполнялось для всех r вида (5). Индексы a, b, c и другие (здесь и далее) представляют компоненты первичного и восстановленного векторов-сигналов, т. е. \mathbf{s} и \mathbf{u} .

В отдельных случаях вместо 2) необходимо выполнение более мягкого условия, которое состоит в том, чтобы матрица \mathbf{K} в (9) не была вырожденной:

$$2) \mathbf{K} = E_{\mathbf{w},r} \left\{ \frac{\partial}{\partial \mathbf{W}} F(\mathbf{x}, \mathbf{W}) \right\}. \quad (9)$$

Другими словами, условие 2) выполняется только локально. Следует отметить, что \mathbf{K} — линейный оператор, отображающий матрицу на матрицу. Компоненты \mathbf{K} имеют вид

$$K_{ab,ij} = E_{\mathbf{w},r} \left\{ \frac{\partial}{\partial W_{ij}} F_{ab}(\mathbf{x}, \mathbf{W}) \right\}, \quad (10)$$

где W_{ij} — элементы \mathbf{W} ; индексы i, j, a, b и др. соответствуют представлению компонент наблюдаемого сигнала \mathbf{x} . Удобно использовать индексы A, B, \dots для представления пар индексов $(a, b), (i, j)$ и др. Тогда для $A = (a, b), B = (c, i)$ \mathbf{K} имеет матричное представление $\mathbf{K} = [K_{A,B}]$, которое действует на $(W_B) = (W_{ij})$ по правилу: $\mathbf{K}\mathbf{W} = \sum_B K_{AB} W_B = \sum_{i,j} K_{ab,ij} W_{ij}$. Обратный относительно \mathbf{K} оператор определяется обратной матрицей $\mathbf{K} = [K_{AB}]$.

Если имеется оценивающая функция (ОФ) $F(\mathbf{x}, \mathbf{W})$, то для наблюдаемых данных $\mathbf{x}(1), \dots, \mathbf{x}(N)$ оценка величины $\tilde{\mathbf{W}}$ с использованием этой обучающей выборки определяется уравнением

$$\sum_{k=1}^N F\{\mathbf{x}(k), \mathbf{W}\} = 0.$$

Оно получено заменой математического ожидания в (7) суммой наблюдаемых величин. Так же

как обучающий алгоритм в РВ определен выражением (1), уравнение для оценки по выборке выполняет свою роль без использования неизвестной ПРВ r ("мешающего параметра"). Так что правы две задачи:

1. Существует ли оценивающая функция, которая работает без знания r .

2. Каким образом найти "хорошую" ОФ F (в то время как их достаточно много).

Допустимый класс оценивающих функций

Алгоритмы, предложенные Джутен и Хераулт (Jutten, Herault [11]), Беллом и Сейновским (Bell, Sejnowski [12]), Амари (Amari [13]), Кардосо и Лакхельдом (Cardoso, Laheld [14]), Ойа и Каруненом (Oja, Kahrunen [15]), использовали различные ОФ, найденные на основе эвристик. Они получались и хорошие, и плохие. ОФ F лучше, чем F' , когда ожидаемая ошибка оценки разделяющей матрицы $\tilde{\mathbf{W}}$, полученная по F , меньше, чем полученная по F' . Но может случиться, что F лучше, чем F' , при условии, что истинное (неизвестное) распределение — $r(\mathbf{s})$, но F' лучше, когда это распределение — $r'(\mathbf{s})$. Следовательно, вообще говоря они не сравнимы. Семейство (или класс) ОФ считается допустимым, если в этом семействе содержится (и ее можно найти) эквивалентная или лучшая ОФ сравнительно с любой ранее выбранной ОФ. Целесообразно ограничиться рассмотрением допустимого класса ОФ.

Амари и Кардосо (Amari, Cardoso [16]) с использованием геометрической теории информации [17], [18] показано, что множество ОФ в виде

$$(\mathbf{x}, \mathbf{W}) = \mathbf{I} - \boldsymbol{\varphi}(\mathbf{y})\mathbf{y}^T$$

(или в покомпонентной форме) (11)

$$F_{ij}(\mathbf{x}, \mathbf{W}) = \delta_{ij} - \varphi_i(y_i)y_j$$

дает допустимый класс (множество) ОФ, где $\boldsymbol{\varphi}(\mathbf{y}) = [\varphi_1(y_1), \varphi_2(y_2), \dots, \varphi_n(y_n)]^T$ включает произвольные нетривиальные функции $\varphi_i(y_i)$. Можно показать, что это действительно ОФ. Когда \mathbf{W} — истинное решение, то y_i и y_j являются независимыми. Поэтому для любого r выполняется соотношение

$$E_{r,\mathbf{w}}\{\varphi_i(y_i)y_j\} = E\{\varphi_i(y_i)\} \cdot E\{y_j\} = 0 \text{ при } i \neq j.$$

Но, когда \mathbf{W} не является истинным решением, это соотношение в общем случае не выполняется. Для диагональных элементов (т. е. при $i = j$) $E\{\varphi_i(y_i)y_i\} = 1$, и это определяет величину восстановленного сигнала y_i . Поскольку амплитуда сигнала может быть произвольной, то можно положить диагональные элементы F_{ii} тоже произвольными (включая элементы типа $F_{ii} = 0$).

Рассмотрим типичные примеры ОФ. Положим,

что $q(\mathbf{s}) = \prod_{i=1}^n q_i(s_i)$ — неверно определенная ФПВ \mathbf{s} , отличающаяся от реальной ФПВ $r(\mathbf{s}) = \prod_{i=1}^n r_i(s_i)$. Отрицательный логарифм функции правдоподобия для \mathbf{x} (при ФПВ $q(\mathbf{s})$) имеет вид

$$\rho(\mathbf{x}, \mathbf{W}) = -\det|\mathbf{W}| - \sum_{i=1}^n \log q_i(y_i),$$

где y_i — i -я компонента от $\mathbf{W}\mathbf{x}$ (зависящая от \mathbf{x} и от \mathbf{W}). Критерий минимизации ρ интерпретируется как минимизация энтропии, или максимизация правдоподобия. Положим, что $\varphi_i(y_i) = -\frac{d}{dy_i} \log q_i(y_i)$. Градиент ρ дает ОФ в виде

$$\tilde{\mathbf{F}}(\mathbf{x}, \mathbf{W}) = -\frac{\partial \rho(\mathbf{x}, \mathbf{W})}{\partial \mathbf{W}} = \mathbf{W}^{-T} - \varphi(\mathbf{y})\mathbf{x}^T.$$

Можно показать, что $\tilde{\mathbf{F}}$ — ОФ. Однако, когда $\tilde{\mathbf{F}}$ есть ОФ, то $\mathbf{F}(\mathbf{y}) = \tilde{\mathbf{F}}(\mathbf{x}, \mathbf{W})\mathbf{W}^T\mathbf{W} = [\mathbf{I} - \varphi(\mathbf{y})\mathbf{y}^T]\mathbf{W}$ также является ОФ. При этом выполняются соотношения: $E\{\mathbf{F}(\mathbf{y})\} = 0$ и $\{E\{\tilde{\mathbf{F}}(\mathbf{x}, \mathbf{W})\}\} = 0$.

Для истинного распределения компонент сигнала \mathbf{s} в виде ПРВ r_i лучший выбор φ_i ($i = 1, 2, \dots, n$) обеспечивает полученное Пхамом (Pham [19]) выражение

$$\varphi_i = -\frac{d}{ds} \log r_i(s).$$

Показано также, что $\tilde{\mathbf{F}}(\mathbf{x}, \mathbf{W})$ и $\mathbf{F}(\mathbf{y})$ связаны линейно и оценивающие их уравнения дают одно и то же решение, поэтому их можно считать эквивалентными ОФ.

В качестве общего случая целесообразно рассмотреть произвольный (обратимый) линейный оператор $\mathbf{R}(\mathbf{W})$, действующий на матрицах. Когда $\mathbf{F}(\mathbf{x}, \mathbf{W})$ является ОФ-матрицей,

$$\mathbf{R}(\mathbf{W}) \cdot \mathbf{F}(\mathbf{x}, \mathbf{W})$$

также будет ОФ-матрицей, поскольку

$$\begin{aligned} E_{\mathbf{w},r}\{\mathbf{R}(\mathbf{W}) \cdot \mathbf{F}(\mathbf{x}, \mathbf{W})\} &= \\ &= \mathbf{R}(\mathbf{W})E_{\mathbf{w},r}\{\mathbf{F}(\mathbf{x}, \mathbf{W})\} = 0. \end{aligned} \tag{11}$$

Кроме того, $\mathbf{F}(\mathbf{x}, \mathbf{W})$ и $\mathbf{R}(\mathbf{W}) \cdot \mathbf{F}(\mathbf{x}, \mathbf{W})$ являются эквивалентными в том смысле, что выведенные оценки в точности одинаковые, т. к. оба уравнения (12) и (13) дают одно и то же решение $\hat{\mathbf{W}}_*$ (пренебрегая произвольным масштабированием и изменением порядка компонент сигнала, восстановленного из смеси):

$$\sum_{k=1}^N \mathbf{F}[\mathbf{x}(k), \mathbf{W}] = 0, \tag{12}$$

$$\sum_{k=1}^N \mathbf{R}(\mathbf{W}) \cdot \mathbf{F}[\mathbf{x}(k), \mathbf{W}] = 0. \tag{13}$$

Это определяет эквивалентный класс ОФ, т. к. при осуществлении оценивания по выборке эти ОФ по существу равноценны.

Однако две эквивалентных ОФ $\mathbf{F}(\mathbf{x}, \mathbf{W})$ и $\mathbf{R}(\mathbf{W}) \cdot \mathbf{F}(\mathbf{x}, \mathbf{W})$ дают различные динамические свойства при обучении нейронной сети в РВ, т. е. динамические свойства алгоритмов (14) и (15) для оценки \mathbf{W} в РВ полностью различны:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta \mathbf{F}(\mathbf{x}(k), \mathbf{W}(k)), \tag{14}$$

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta \cdot \mathbf{R}(\mathbf{W}(k)) \cdot \mathbf{F}(\mathbf{x}(k), \mathbf{W}(k)). \tag{15}$$

При этом, чтобы получить хороший алгоритм РВ для получения размешивающей матрицы, необходимо вместо (1) рассматривать расширенную форму ОФ $\mathbf{R}(\mathbf{W}) \cdot \mathbf{F}(\mathbf{x}, \mathbf{W})$.

Стандартизованная ОФ и адаптивный алгоритм Ньютона

Динамика обучения в виде

$$\Delta \mathbf{W}(k) = \mathbf{W}(k+1) - \mathbf{W}(k) = \eta \cdot \mathbf{F}[\mathbf{x}(k), \mathbf{W}(k)] \cdot \mathbf{W}(k) \tag{16}$$

может быть ускорена на основе использования метода Ньютона, который определяется соотношением

$$\Delta \mathbf{X}(k) = \eta \cdot \mathbf{K}^{-1}[\mathbf{W}(k)] \cdot \mathbf{F}[\mathbf{x}(k), \mathbf{W}(k)].$$

Поскольку $\mathbf{K}^{-1}\mathbf{F}$ является эквивалентной ОФ относительно \mathbf{F} , то видно, что метод Ньютона определяется с помощью ОФ $\mathbf{F}^*(\mathbf{x}, \mathbf{W}) = \mathbf{K}^{-1}(\mathbf{W}) \cdot \mathbf{F}(\mathbf{x}, \mathbf{W})$. Метод Ньютона сходится линейно. Решение, получаемое этим методом, всегда устойчиво, потому что гессиан \mathbf{F}^* является единичной матрицей. Это следует из соотношения

$$\mathbf{K} \bullet = E \left\{ \frac{\partial \mathbf{F}^*}{\partial \mathbf{X}} \right\} = \frac{\partial \mathbf{K}^{-1}}{\partial \mathbf{X}} E\{\mathbf{F}\} + \mathbf{K}^{-1} \circ \mathbf{K} = \mathbf{I}.$$

Стандартизованная ОФ-матрица \mathbf{F}^* описывается выражением

$$F_{ab}^* = c_{ab} \{k_b \cdot \sigma_a^2 \cdot \varphi_a(y_a) \cdot y_b - \varphi_b(y_b) \cdot y_a\}, \quad a \neq b; \tag{17}$$

включает параметры σ_a^2 и k_a , которые обычно известны [14], [16], [20]. Они зависят от статистических свойств первичного сигнала s_a , т. е. от компонент первичного сигнала \mathbf{s} , образующего смесь. С параметрами σ_a и k_b связан параметр $c_{ab} =$

$$= \frac{1}{k_a k_b \cdot \sigma_a^2 \sigma_b^2 - 1}. \quad (\text{Напомним, что } a \text{ и } b \text{ — буквен-})$$

ные обозначения индексов, об использовании которых было сказано выше).

Для выполнения метода Ньютона необходимо использовать адаптивный алгоритм, который оценивает параметры. Это не только ускоряет сходимость, но и автоматически обеспечивает устойчивость решения. Если $\sigma_a^2(k)$ и $k_a(k)$ — оценка параметров в дискретный момент времени k , то для обновления параметров может использоваться адаптивное правило вида

$$k_a(k+1) = (1-\eta_0)k_a(k) + \eta_0\phi'_a(y_a(k)),$$

$$\sigma_a^2(k+1) = (1-\eta_0)\sigma_a^2(k) + \eta_0y_a^2(k),$$

где η_0 — параметр скорости обучения. При этом в качестве диагонального элемента матрицы \mathbf{F} возможно использовать $F_{aa} = 1 - y_a^2$. Тогда восстановленный сигнал будет нормализован: $\sigma_a^2 = 1$, так что \mathbf{F}^* принимает упрощенный вид:

$$F_{ab}^* = -\frac{1}{k_a k_b} \{k_b \cdot \phi_a(y_a) \cdot y_b - \phi_b(y_b) \cdot y_a\}, \quad a \neq b. \quad (18)$$

Метод адаптивного выбора функции ϕ

Величина ошибки оценки \mathbf{W} зависит от вида $\mathbf{F}(\mathbf{x}, \mathbf{W})$ или $\mathbf{F}^*(\mathbf{x}, \mathbf{W})$, т. е. от функций ϕ . Хотя стандартизованная ОФ \mathbf{F}^* улучшает устойчивость и сходимость решения, ошибка алгоритма при обучении с использованием выборки и в режиме реального времени зависит от ПРВ ϕ . Снижение величины ошибки достигается адаптивным методом выбора ϕ .

Если для получения ϕ использованы данные фактической ПРВ первичных сигналов, то оцененное значение $\hat{\mathbf{W}}$ является оценкой максимального правдоподобия и оно эффективно в том смысле, что (асимптотическая) ошибка минимальна. Но установить ПРВ сигналов источника трудно, поэтому используется параметрическое семейство ПРВ ϕ : $\phi_a = \phi_a(y; \Theta_n)$ для каждой компоненты s_a сигнала источника (ПС) \mathbf{s} , а обновление параметра Θ_n осуществляется по соотношению

$$\Delta\Theta_a = \eta_0 \frac{\partial \rho}{\partial \Theta_a}.$$

Существует несколько моделей выбора ϕ_a . Смесь гауссианов — один из методов для аппроксимации распределений сигнала. Это — параметрическое семейство

$$q(y; \Theta) = \sum_{i=1}^u v_i \exp \left\{ -\frac{(x - \mu_i)^2}{2\sigma_i^2} \right\},$$

где Θ состоит из численных значений v_i, μ_i и σ_i^2 . Отсюда выводится соответствующая параметри-

ческая плотность $\phi(y; \Theta)$. Такая модель ПРВ включает "надгауссову" (обостренную) и "подгауссову" (сплюсненную) плотности распределения. Более простой метод состоит в использовании семейства обобщенных распределений Гаусса: $q(y; \Theta) = c \cdot \exp\{-|y|^\Theta\}$, где Θ — единственный настраиваемый параметр. Это семейство также покрывает надгауссову и подгауссову плотности. Адаптивные нелинейные активационные функции, используемые в алгоритмах анализа независимых компонент, в этом случае имеют вид $q(y; \Theta) = \tilde{c} \cdot \text{sign}(y)|y|^{\Theta-1}$, где \tilde{c} — положительная константа масштаба.

Другая форма экспоненциального семейства ПРВ предложена Цхангом (Zhang et al. [20]). Она тоже объединяет три вида типовых распределений вероятности (гауссово, надгауссово и подгауссово распределения) и имеет вид $q_a(s, \theta_a) = \exp\{\theta_a^T \mathbf{g}(s) - \psi(\theta_a)\}$, где θ_a — канонические параметры, $\mathbf{g}(s)$ — адекватная вектор-функция, а $\psi(\theta_a)$ определяет нормализацию распределения. Плотность вероятности ϕ_a , соответствующую этому распределению, дает выражение

$$\phi_a(y) = -\frac{d}{dy} \log q_a(y, \theta_a) = \theta_a^T \cdot \mathbf{g}'(y).$$

В [20] предложено использовать также трехмерную модель (вектор распределений): $\mathbf{g}(y) = [\log \text{sech}(y), -y^4, -y^2]^T$ или $\mathbf{g}'(y) = [\tanh(y), y^3, y]^T$. Такой векторной модели соответствуют типичные виды показанных выше плотностей, т. е. модель Цханга также определяет надгауссову, подгауссову и гауссову плотности распределения. При этом $\phi_a(y)$ является их линейной комбинацией, объединяющей все случаи. Параметр θ_a адаптивно определяется по соотношению

$$\theta_a(k+1) = \theta_a(k) - \eta(k) \cdot [\mathbf{g}(y_a(k)) + E\{\mathbf{g}(y_a)\}],$$

где $E\{\mathbf{g}(y_a)\}$ может также быть оценено адаптивно.

ОЦЕНИВАЮЩИЕ ФУНКЦИИ В СЛУЧАЕ ШУМА

При анализе случая с наличием шума $\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{v}$, где \mathbf{v} — вектор шума при измерении сигналов смеси, полагается, что шум гауссов с некоррелированными компонентами и что его ковариационная матрица $\mathbf{R}_{\mathbf{v}\mathbf{v}}$ — диагональная:

$$\mathbf{R}_{\mathbf{v}\mathbf{v}} = E\{\mathbf{v}\mathbf{v}^T\} = \text{Диаг.}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2\}.$$

Кроме того, чтобы фиксировать масштаб первичных сигналов, образующих смесь, принимает- ся, что $E\{s_i^2\} = 1$.

Размешивающая матрица $\mathbf{W} = \mathbf{H}^{-1}$ определяет оценку \mathbf{y} первичного вектора сигнала \mathbf{s} : $\mathbf{y} = \mathbf{W}\mathbf{x}$.

С учетом принятых условий (и обозначений) справедливо соотношение: $\mathbf{y} = \mathbf{s} + \mathbf{W}\mathbf{v} = \mathbf{s} + \tilde{\mathbf{v}}$, где $\tilde{\mathbf{v}} = \mathbf{W}\mathbf{v}$ — вектор шума, у которого компоненты коррелированные. В случае наличия шума функции типа $\mathbf{F} = \mathbf{I} - \boldsymbol{\Phi}(\mathbf{y})\mathbf{y}^T$ в общем случае не являются ОФ. Действительно, даже если \mathbf{y} выведено из реальной размешивающей матрицы \mathbf{W} , то все же $E\{\mathbf{I} - \boldsymbol{\Phi}(\mathbf{y})\mathbf{y}^T\} \neq 0$, т. к. y_i и y_j не являются больше независимыми, даже когда $\mathbf{W} = \mathbf{H}^{-1}$. Однако даже для случая наличия шума существуют ОФ.

Для истинного значения размешивающей матрицы $\mathbf{W} = \mathbf{H}^{-1}$ слагаемое с шумом $\tilde{\mathbf{v}} = \mathbf{W}\mathbf{v}$ имеет (как и \mathbf{v}) гауссово распределение. Для ковариационной матрицы \mathbf{V} преобразованного шума $\tilde{\mathbf{v}}$ справедливо соотношение $\mathbf{V} = E\{\tilde{\mathbf{v}}\tilde{\mathbf{v}}^T\} = E\{\mathbf{W}\mathbf{v}\mathbf{v}^T\mathbf{W}^T\} = \mathbf{W}\mathbf{R}_v\mathbf{W}^T$.

При анализе возможных видов ОФ для условий наличия шума Каванабе и Муратой (Kawanabe, Murata [21]) в качестве наиболее простого варианта найдена функция-матрица $\mathbf{F}(\mathbf{y}, \mathbf{W})$, элементы (F_{ab}) которой определяются выражением

$$F_{ab}(\mathbf{y}, \mathbf{W}) = y_a^3 y_b - 3 \cdot v_{aa} \cdot y_a \cdot y_b - 3 \cdot v_{ab} \cdot y_a^2 + 3 \cdot v_{aa} \cdot v_{ab},$$

где v_{ab} — элементы матрицы \mathbf{V} . Причем показано, что при $\mathbf{W} = \mathbf{H}^{-1}$ выполняется $E\{\mathbf{F}(\mathbf{y}, \mathbf{W})\} = 0$, т. е. найденная функция-матрица $\mathbf{F}(\mathbf{y}, \mathbf{W})$ является ОФ. Следовательно, адаптивный обучающий алгоритм (19) эффективен даже при значительном гауссовом шуме:

$$\Delta \mathbf{W}(k) = \mathbf{W}(k+1) - \mathbf{W}(k) = \boldsymbol{\eta}(k) \cdot \mathbf{F}[y(k), \mathbf{W}(k)] \cdot \mathbf{W}(k). \quad (19)$$

Когда ковариационная матрица \mathbf{R}_v шума измерения неизвестна, необходимо произвести ее оценку. С использованием метода факторного анализа Икедой и Тойамой (Ikeda, Toyama [22]) получено адаптивное правило для определения недиагональных элементов матрицы \mathbf{V} :

$$v_{ab}(k+1) = (1 - \eta_0) v_{ab}(k) + \eta_0 y_a(k) y_b(k),$$

где η_0 — параметр скорости обучения.

К сожалению, обучающий алгоритм (19) не всегда обладает устойчивостью. Устойчивость обеспечивается только при использовании стандартизованной ОФ \mathbf{F}^* , которая реализуется адаптивным методом Ньютона. При этом ОФ \mathbf{F}^* может

быть получена методом, подобным рассмотренному выше случаю без шума.

ОЦЕНИВАЮЩИЕ ФУНКЦИИ ДЛЯ КОРРЕЛИРОВАННЫХ ПО ВРЕМЕНИ СИГНАЛОВ ИСТОЧНИКА

Независимые сигналы источника смеси (или первичные сигналы (ПС)) во многих случаях бывают коррелированными во времени. Если известен этот факт, процедура РСС становится проще (даже если не известны точные значения временных коэффициентов корреляции). Для алгоритма разделения смеси в таком случае достаточно корреляций второго порядка. Полное представление о методе РСС для описанных условий дает анализ моделей первичных сигналов.

Модель первичных сигналов

Для взаимно независимых ПС источника s_i ($i = 1, \dots, n$) принимается линейная стохастическая модель:

$$s_i(k) = \sum_{p=1}^{L_i} a_{ip} s_i(k-p) + \varepsilon_i(k), \quad (20)$$

где L_i — параметр, определяющий конечный интервал временной корреляции i -го ПС;²⁾ ε_i — независимые, одинаково распределенные (со средним, равным нулю) элементы временного ряда "обновления"³⁾. Элементы этого ряда могут быть гауссовыми или негауссовыми. Здесь достаточно принять, что для них выполняются условия:

$$E\{\varepsilon_i(k)\} = 0; \quad E\{\varepsilon_i(k)\varepsilon_j(k')\} = 0$$

при $i \neq j$ или $k \neq k'$.

Введение оператора сдвига z^{-1} (такого, что $z^{-1}s_i(k) = s_i(k-1)$) преобразует соотношение (20) к виду:

$$A_i(z^{-1})s_i(k) = \varepsilon_i(k), \quad \text{где} \quad A_i(z^{-1}) = 1 - \sum_{p=1}^{L_i} a_{ip} z^{-p}.$$

При использовании обращения полинома A_i ПС представляется в виде

$$s_i(k) = [A_i^{-1}(z^{-1})] \cdot \varepsilon_i(k), \quad \text{где} \quad A_i^{-1}(z^{-1}) = \sum_{p=0}^{\infty} \bar{a}_{ip} \cdot z^{-p}$$

Функция $A_i^{-1}(z^{-1})$ представляет импульсный отклик, с помощью которого i -й источник ПС $s_i(k)$

²⁾ Конечное значение L_i соответствует авторегрессионной модели порядка L_i .

³⁾ Такой ряд часто (например, при описании алгоритмов теории фильтров Калмана и других рекуррентных конструкций) называют рядом инновации, а элементы этого ряда просто инновациями.

определяется по $\varepsilon_i(k)$ и соответственно набор $\{s_i(k)\}$ определяется по сигналам $\{\varepsilon_i(k)\}$. При этом если $r_i(\varepsilon_i)$ — ПРВ для $\varepsilon_i(k)$, то условная плотность ПРВ $s_i(k)$ (условная относительно прошлых значений сигнала) представляется в виде

$$p_i\{s_i(k)|s_i(k-1), s_i(k-2), \dots\} = r_i \left\{ s_i(k) - \sum_p a_{ip} s_i(k-p) \right\} = r_i \{A_i(z^{-1})s_i(k)\}. \quad (21)$$

Поэтому для вектора сигналов источников $\mathbf{s}(k)=[s_1(k), \dots, s_n(k)]^T$ в момент (k) условная плотность вероятности определяется соотношением

$$p\{\mathbf{s}(k)|\mathbf{s}(k-1), \mathbf{s}(k-2), \dots\} = \prod_{i=1}^n r_i \{[A_i(z^{-1})]s_i(k)\}. \quad (22)$$

Удобно (для краткости) ввести обозначения (23)

$$\begin{aligned} \boldsymbol{\varepsilon} &= [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n]^T, \\ \mathbf{A}(z^{-1}) &= \text{Диаг.} \{A_1(z^{-1}), \dots, A_n(z^{-1})\}, \\ r(\boldsymbol{\varepsilon}) &= \prod_{i=1}^n r_i(\varepsilon_i), \\ \mathbf{s}(k, \text{nped.}) &= \{\mathbf{s}(k-1), \mathbf{s}(k-2), \dots\}, \end{aligned} \quad (23)$$

а также сокращения в форме (24) там, где это не будет затруднять их понимание:

$$\mathbf{s}_k = \mathbf{s}(k), \quad \mathbf{x}_k = \mathbf{x}(k), \quad \mathbf{y}_k = \mathbf{y}(k) = \mathbf{W}(k). \quad (24)$$

При этом соотношение (22) принимает вид

$$p\{\mathbf{s}(k)|\mathbf{s}(k, \text{nped.})\} = r\{\mathbf{A}(z^{-1})\mathbf{s}(k)\}.$$

Совместная плотность вероятности набора $\{\mathbf{s}(1), \mathbf{s}(2), \dots, \mathbf{s}(N)\}$ представляется в форме:

$$p(\mathbf{s}(1), \mathbf{s}(2), \dots, \mathbf{s}(N)) = \prod_{k=1}^N p\{\mathbf{s}(k)|\mathbf{s}(k, \text{nped.})\} = \prod_{k=1}^N r\{\mathbf{A}(z^{-1})\mathbf{s}(k)\}, \quad (25)$$

где при $k \leq 0$ полагаем $\mathbf{s}(k)$ равным нулю. (Практически при $k < 0$ $\mathbf{s}(k)$ не равно нулю, так что (25) является приближением, которое выполняется асимптотически, т. е. для больших значений N).

Модели источника (ПС) определяются n функциями $r_i(\varepsilon_i)$ и n функциями, обратными относительно функций импульсного отклика $A_i(z^{-1})$. Процедура РСС извлекает независимые сигналы из их мгновенной линейной смеси $\mathbf{x}(k)$ без знания точной формы $r_i(\varepsilon_i)$ и $A_i(z^{-1})$. Можно сказать, что $r_i(\varepsilon_i)$ и $A_i(z^{-1})$ следует трактовать как неизвестные и мешающие параметры.

Для данных N наблюдений смеси $\{\mathbf{x}(1), \mathbf{x}(2), \dots$

$\dots, \mathbf{x}(N)\}$ их совместная плотность распределения получается из (25) и соотношений $\mathbf{s}_k = \mathbf{W}\mathbf{x}_k$, где $\mathbf{W} = \mathbf{H}^{-1}$. Эта совместная плотность имеет вид (26) и определяется размешивающим матричным параметром $\mathbf{W} = \mathbf{H}^{-1}$ и мешающими параметрами \mathbf{A} и ПРВ $\{r_i\}_{i=1, \dots, N}$ модели источника сигналов:

$$p\{\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{W}; \mathbf{A}, r\} = \text{Det} |\mathbf{W}|^N \prod_{k=1}^N r_k \{\mathbf{A}(z^{-1})\mathbf{W}\mathbf{x}_k\}. \quad (26)$$

Оценки максимального правдоподобия

Если представить (на момент), что r и \mathbf{A} известны, то для того, чтобы оценить \mathbf{W} , будет возможно использовать метод максимального правдоподобия (МП). Логарифм МП, полученный на основе (26), имеет вид:

$$\begin{aligned} \rho^{(N)}(\mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{W}, \mathbf{A}, r) &= \\ &= -\log p\{\mathbf{x}_1, \dots, \mathbf{x}_N; \mathbf{W}, \mathbf{A}, r\} = \\ &= -N \cdot \log |\mathbf{W}| - \sum_{k=1}^N \log r\{\mathbf{A}(z^{-1})\mathbf{W}\mathbf{x}_k\} = \\ &= -N \cdot \log |\mathbf{W}| - \sum_{k=1}^N \log r\{\mathbf{A}(z^{-1})\mathbf{y}_k\}, \end{aligned} \quad (27)$$

где $\mathbf{y}_k = \mathbf{W}\mathbf{x}_k$. Оценка МП \mathbf{W} максимизирует приведенную выше функцию правдоподобия при данных N наблюдений $\mathbf{x}_1, \dots, \mathbf{x}_N$.

Полагая $\rho(\mathbf{y}_k, \mathbf{W}) = -\log r\{\mathbf{A}(z^{-1})\mathbf{y}_k\}$, можно видеть, что ρ зависит не только от \mathbf{y}_k , но (из-за оператора $\mathbf{A}(z^{-1})$) зависит также от прошлых значений $\mathbf{y}_{k-1}, \mathbf{y}_{k-2}, \dots$. Кроме того, ρ функционально зависит от \mathbf{W} только посредством набора \mathbf{y}_k . Это позволяет представить $\rho^{(N)}$ в виде

$$\rho^{(N)} = -N \cdot \log |\mathbf{W}| - \sum_{k=1}^N \rho(\mathbf{y}_k, \mathbf{W}).$$

Малое изменение $d\rho$, связанное с малым изменением \mathbf{W} (от \mathbf{W} до $\mathbf{W} + d\mathbf{W}$), определяется выражением

$$d\rho = -\varphi_r(\mathbf{A}\mathbf{y}_k)^T d(\mathbf{A}\mathbf{y}_k), \quad (28)$$

где $\varphi_r(\mathbf{y}) = -\frac{\partial}{\partial \mathbf{y}} \log r(\mathbf{y})$ — вектор. Теперь, поскольку $d(\mathbf{A}\mathbf{y}_k) = \mathbf{A} d\mathbf{y}_k$ и $d\mathbf{y}_k = d\mathbf{W}\mathbf{x}_k = d\mathbf{X}\mathbf{y}_k$, то справедливо соотношение $d\rho = \varphi_r(\mathbf{A}\mathbf{y}_k)^T \mathbf{A} d\mathbf{X}\mathbf{y}_k$. А это приводит к скорости функции МП на единицу изменения $d\mathbf{X}$:

$$\frac{\partial \rho^{(N)}}{\partial \mathbf{X}} = \frac{\partial \rho^{(N)}}{\partial \mathbf{W}} \mathbf{W} = \sum_{k=1}^N [\mathbf{I} - \{\varphi_r(\mathbf{A}\mathbf{y}_k)\mathbf{A}\} \mathbf{y}_k^T], \quad (29)$$

где $\varphi_r \mathbf{A}$ — вектор-столбец с компонентами $\varphi_j A_j(z^{-1})$.

Далее, поскольку $d\rho$ в компонентной форме — $d\rho = \sum c_{ij} dX_{ij}$, то $d\rho/d\mathbf{X}$ — матрица с элементами c_{ij} . Следовательно, $\varphi_r(\mathbf{A}\mathbf{y})\mathbf{A}\mathbf{y}^T$ в компонентах представляется в виде $\varphi_i(A_i(z^{-1})y_i)A_i(z^{-1})y_j$.

Если положить $\frac{\partial \rho}{\partial \mathbf{X}} = \mathbf{F}(\mathbf{y}, \mathbf{W}; r, \mathbf{A}) = \mathbf{I} - \{\varphi_r(\mathbf{A}\mathbf{y}) \cdot \mathbf{A}\} \mathbf{y}^T$, то уравнение правдоподобия (с учетом сокращений (24)) приобретает форму

$$\sum_{k=1}^N \mathbf{F}(\mathbf{y}_k, \mathbf{W}; r, \mathbf{A}) = \mathbf{0}, \tag{30}$$

а решение $\hat{\mathbf{W}}$ уравнения (30) дает оценку максимального правдоподобия размещивающей матрицы.

Оценивающие функции

Поскольку ПРВ сигналов источника $\{r_i\}$ и фильтры $\{A_i(z^{-1})\}$ неизвестны, нет практической возможности использовать рассмотренные выше оценивающие функции \mathbf{F} , которые зависят от \mathbf{r} и \mathbf{A} . Поэтому целесообразно выполнить поиск ОФ в классе функций вида

$$\mathbf{F}(\mathbf{y}, \mathbf{W}, q, \mathbf{B}) = \frac{d\rho(\mathbf{y}, \mathbf{W}, q, \mathbf{B})}{d\mathbf{X}} = \mathbf{I} - \varphi_q \{(\mathbf{B}(z^{-1})\mathbf{y}) \cdot \mathbf{B}(z^{-1})\} \cdot \mathbf{y}^T, \tag{31}$$

где q — любое (фиксированное) независимое распределение; матрица $\mathbf{B}(z^{-1}) = \text{Диаг.}\{B_1(z^{-1}), \dots, B_n(z^{-1})\}$

с фиксированными фильтрами $B_i(z^{-1}) = \sum_{p=0}^{L_i} b_{ip} z^{-p}$.

Это — оценивающая функция (при любых q и \mathbf{B}), т. к. она удовлетворяет соотношению $\mathbf{E}_{\mathbf{W}, r, \mathbf{A}}[\mathbf{F}(\mathbf{y}, \mathbf{W}, q, \mathbf{B})] = \mathbf{0}$ для любых сигналов источника, имеющих независимые распределения компонент, и фильтров $\mathbf{A}(z^{-1})$. Когда истинный мешающий параметр $r = q$ и $\mathbf{A} = \mathbf{B}(z^{-1})$, то $\mathbf{F}(\mathbf{y}, \mathbf{W}, q, \mathbf{B})$ является функцией МП. Однако даже если q и \mathbf{B} определены неверно, это соотношение может служить в качестве ОФ.

Условия, определяющие возможность идентификации

Более последовательно и корректно условия возможности идентификации \mathbf{W} установлены Тонгом (Tong et al. [23]) и Кономом (Common [24]). Должно выполняться (по крайней мере) одно из двух условий:

1. Все независимые сигналы — источники сме-

си имеют различные спектры, т. е. передаточные функции $A_i(z^{-1})$ различны.

2. Когда некоторые сигналы имеют одинаковые спектры, распределения r_i этих сигналов — негауссовы, кроме одного сигнала.

В целом эти условия порождают общую формулировку: когда удовлетворяются условия, определяющие возможность идентификации, минимальный допустимый класс оценивающих функций является линейной комбинацией недиагональных элементов матрицы $\mathbf{F}(\mathbf{y}, \mathbf{W}, q, \mathbf{B})$, в которой q и \mathbf{B} произвольные. Оценивающее уравнение имеет вид $\sum_{k=1}^N \mathbf{F}(\mathbf{y}_k, \mathbf{W}, q, \mathbf{B}) = \mathbf{0}$.

Адаптивный обучающий алгоритм на основе такой ОФ принимает вид (32), а в более общем случае при использовании стандартизованной ОФ принимает вид (33):

$$\Delta \mathbf{W}(k) = \eta(k) \mathbf{F}[\mathbf{y}(k), \mathbf{W}(k)] \cdot \mathbf{W}(k), \tag{32}$$

$$\Delta \mathbf{W}(k) = \eta(k) \mathbf{F}^*[\mathbf{y}(k), \mathbf{W}(k)]. \tag{33}$$

Стандартизованная оценивающая функция и метод Ньютона

Если взять обратимый (т. е. не сингулярный) матричный оператор $\mathbf{R}(\mathbf{W}) = (\mathbf{R}_{AB})$, который может зависеть от \mathbf{W} , тогда \mathbf{F} и $\tilde{\mathbf{F}} = \mathbf{R} \cdot \mathbf{F}$ — эквивалентные оценивающие функции (ОФ). Одна ОФ \mathbf{F}^* из класса эквивалентных функций, которая удовлетворяет условию

$$\mathbf{K}^* = \mathbf{E} \left\{ \frac{\partial \mathbf{F}^*}{\partial \mathbf{X}} \right\} = \text{единичный оператор (оператор тождественности)},$$

называется стандартизованной ОФ [25]. При этом если задана ОФ \mathbf{F} , то ее стандартизованная форма определяется выражением

$$\mathbf{F}^* = \mathbf{K}^{-1} \mathbf{F}, \text{ где } \mathbf{K} = \mathbf{E} \left\{ \frac{\partial \mathbf{F}}{\partial \mathbf{X}} \right\}.$$

Теперь можно вычислить

$$\mathbf{K} = \mathbf{E} \left\{ \frac{\partial \mathbf{F}(\mathbf{y}, \mathbf{W}, q, \mathbf{B})}{\partial \mathbf{X}} \right\}$$

при истинном решении $\mathbf{W} = \mathbf{H}^{-1}$. Переписывая $\mathbf{F} = \partial \rho / \partial \mathbf{X}$ (или $d\rho = -\text{tr}(d\mathbf{X}) + [\varphi(\tilde{\mathbf{y}})^T \mathbf{B}(z^{-1})] d\mathbf{X}\mathbf{y}$ в компонентной форме, где положено $\tilde{\mathbf{y}} = [\mathbf{B}(z^{-1})]\mathbf{y}$), можно вычислить дифференциал второго порядка, который представляется в виде выражения:

$$d^2 \rho = d \left(\sum_{i,j,p} \varphi_i(\tilde{y}_i) b_{ip} y_j (k-p) \right) dX_{ij} =$$

$$= \sum_{i,j,m,q,k} \varphi'_i(\tilde{y}_i) b_{iq} y_m (k-p) \cdot b_{iq} y_j (k-p) dX_{im} dX_{ij} + \sum_{i,j,m,k} \varphi'_i(\tilde{y}_i) b_{iq} y_m (k-p) dX_{jm} dX_{ij}, \quad (34)$$

где $\varphi'_i(y) = d\varphi_i(y)/dy$.

Для истинного решения получается соотношение

$$E\{\varphi'(\tilde{y}_i) y_j y_m\} = E\{\varphi(\tilde{y}_i) y_j\} = 0,$$

если не выполняется $i = j = m$, и

$$d^2\rho = \sum_i E\{\varphi'_i(\tilde{y}_i) \tilde{y}_i^2\} (dX_{ii})^2 + \sum_{i,j} E\{\varphi(\tilde{y}_i) \tilde{y}_i\} dX_{ij} dX_{ji} + \sum_{i \neq j} E\left\{\varphi'_i(\tilde{y}_i) \left[\sum_{p=0} b_{ip} y_j (k-p)\right]^2\right\} (dX_{ij})^2.$$

Следовательно, квадратичная (относительно набора dX_{ij}) форма $d\rho^2$ распадается на диагональные члены $\sum_i (\tilde{m}_i + 1)(dX_{ii})^2$ и на матрицы более низкого порядка (2×2), состоящие из dX_{ij} и dX_{ji}

($i \neq j$): $\sum_{i \neq j} \left\{ \tilde{k}_i \tilde{\sigma}_{ij}^2 (dX_{ij})^2 + dX_{ij} dX_{ji} \right\}$. В этих матрицах использованы обозначения:

$$\tilde{m}_i \Rightarrow \left\{ \varphi'_i(\tilde{y}_i) \tilde{y}_i^2 \right\}; \quad \tilde{k}_i = E\left\{ \varphi'_i(\tilde{y}_i) \right\};$$

$$\tilde{\sigma}_{ij}^2 = E\left\{ \left[\sum_{p=0} b_{ip} y_j (k-p) \right]^2 \right\}.$$

Если положить $F_{ii}(\mathbf{y}) = 1 - y_i^2$, то восстановленные сигналы удовлетворяют условию $E\{y_i^2\} = 1$. В этом случае диагональные члены равны $2\sum_i dX_{ii}^2$, а диагональные члены (2×2)-матриц равны

$$\tilde{k}_i \tilde{\sigma}_{ij}^2 (dX_{ij})^2 + \tilde{h}_i dX_{ij} dX_{ji},$$

где $\tilde{h}_i = E\{\varphi(\tilde{y}_i) \tilde{y}_i\}$.

Из проведенного анализа получается условие устойчивости алгоритма: $\Delta\mathbf{W} = \eta\mathbf{F}(\mathbf{y}, \mathbf{W})\mathbf{W}$.

Таким образом, решение для разделяющей матрицы \mathbf{W} асимптотически (т. е. для большой выборки сигналов смеси) устойчиво только при условии: $\tilde{m}_i + 1 > 0$, $\tilde{k}_i > 0$, $\tilde{k}_i \tilde{k}_j \tilde{\sigma}_{ij}^2 \tilde{\sigma}_{ji}^2 > 1$.

Обратная матрица \mathbf{K}^{-1} имеет ту же структуру, что и \mathbf{K} . Ее диагональная часть \mathbf{K}_{AA} при $A = (i, i)$

равна $k_{ii,ii} = \frac{1}{1 + \tilde{m}_i}$, а ее (2×2)-диагональная часть \mathbf{K}_{AA} для $A = (i, j)$ и $A' = (j, i)$ равна:

$$\mathbf{K}_{AA} = c_{ij} \begin{bmatrix} \tilde{k}_j \tilde{\sigma}_{ji}^2 & -1 \\ -1 & \tilde{k}_i \tilde{\sigma}_{ij}^2 \end{bmatrix},$$

где $c_{ij} = \frac{1}{\tilde{k}_i \tilde{\sigma}_{ij}^2 \tilde{k}_j \tilde{\sigma}_{ji}^2 - 1}$.

Кроме того, элементы стандартизированной матрицы-функции $\mathbf{F}^*(\mathbf{y}, \mathbf{W})$ определяются в виде

$$F_{ij}^* = c_{ij} [-\tilde{k}_j \tilde{\sigma}_{ji}^2 \varphi_i(\tilde{y}_i) \tilde{y}_j^{(i)} + \varphi_j(\tilde{y}_j) \tilde{y}_i^{(j)}];$$

$$F_{ij}^* = \frac{1}{\tilde{m}_i + 1} \{1 - \varphi_i(\tilde{y}_i) \tilde{y}_i\}, \quad \text{где } \tilde{y}_j^{(i)} = B_i(z^{-1}) y_j.$$

Соответствующий алгоритм обучения, использующий метод Ньютона, определяется соотношением

$$\Delta\mathbf{W} = \eta \cdot \mathbf{F}^*(\mathbf{y}, \mathbf{W}). \quad (35)$$

ПОЛУПАРАМЕТРИЧЕСКИЕ МОДЕЛИ МНОГОКАНАЛЬНОГО РСС ДЛЯ СВЕРТКИ ПС

В работах [4], [5] развит анализ моделей, алгоритмов и нейросетевых структур, решающих задачу РСС для линейного смешивания ПС. Задача РСС для смесей ПС со сверткой значительно сложнее, и попытки ее решения наметились только после использования нового подхода на основе модернизированной формы градиента [7]. Развитие более совершенного метода РСС (на основе обращения операции многоканальной свертки (ООМС)) связано с появлением элементов теории геометрических структур на многообразиях фильтров [Цханг, Сичоки, Амари [26]].

Геометрические свойства многообразия КИХ-фильтров, основанные на структуре группы Ли, позволяют формулировать задачу РСС при многоканальной свертке ПС в рамках полупараметрической модели, что позволяет вывести семейство ОФ для разделения сигналов из их многоканальной свертки. При этом полезен анализ эффективности приближенной оценки $\mathbf{W}(z)$ по выборке на основе использования ОФ. Кроме того, в [20] и [26] показано, что при определенных условиях (отсутствия так называемой сингулярности \mathbf{W}_0) обучение алгоритма как по выборке, так и на основе м-градиента обладает высокой эффективностью.

Элементы формализации задачи и обозначений

В качестве модели смешивания ПС в форме многоканальной свертки рассматривается линейная инвариантная по времени система (ЛИВС):

$$\mathbf{x}(k) = \sum_{p=0}^{\infty} \mathbf{H}_p \mathbf{s}(k-p), \quad (36)$$

где \mathbf{H}_p — $(n \times n)$ -матрица смешивающих коэффициентов при временной задержке p (называемая импульсным откликом на время p); $\mathbf{s}(k) = [s_1(k), s_2(k), \dots, s_n(k)]^T$ — n -мерный вектор сигналов источника (с нулевым средним и идентично распределенными компонентами); $\mathbf{x}(k) = [x_1(k), x_2(k), \dots, x_n(k)]^T$ — n -мерный вектор сигналов сенсоров ИИС. Для простоты будет использоваться обозначение $\mathbf{H}(z) = \sum_{p=0}^{\infty} \mathbf{H}_p z^{-p}$, где z — переменная z -преобразования. $\mathbf{H}(z)$ может быть названо смешивающим фильтром. В задаче РСС для ПС в форме многоканальной свертки смешивающий фильтр неизвестен, а целью задачи является восстановление сигналов источника (ПС) с использованием только сигналов сенсоров $\mathbf{x}(k)$ и некоторой информации относительно распределений ПС.

ООМС для реализации РСС осуществляется с помощью другой (отличной от (36)) ЛИВС общего вида, т. е. "неказуальной" системы:

$$\mathbf{y}(k) = \sum_{p=-\infty}^{\infty} \mathbf{W}_p \mathbf{x}(k-p),$$

где $\mathbf{y}(k) = [y_1(k), y_2(k), \dots, y_n(k)]^T$ — n -мерный вектор выхода; \mathbf{W}_p — $(n \times n)$ -мерная матрица коэффициентов (при временной задержке p), компоненты которой являются параметрами, определяемыми в процессе обучения соответствующего алгоритма.

Матричная передаточная функция размешивающих фильтров представляется в виде

$$\mathbf{W}(z) = \sum_{p=-\infty}^{\infty} \mathbf{W}_p \cdot z^{-p}.$$

Цель ООМС состоит в получении выходных сигналов $\mathbf{y}(k)$ размешивающей модели максимально взаимно ("пространственно") независимыми и с независимыми, но одинаковыми распределениями по времени. Для этого используется полупараметрическая модель, с помощью нее создается семейство ОФ, после чего строятся эффективные обучающие алгоритмы для определения параметров разделяющего фильтра $\mathbf{W}(z)$.

На практике ООМС выполняется с использованием КИХ-фильтра, т. е. фильтра с конечной импульсной передаточной функцией: $\mathbf{W}(z) = \sum_{p=0}^L \mathbf{W}_p \cdot z^{-p}$, где L — максимальный порядок (длина) разделяющего фильтра в ООМС. Альтернативно, можно использовать неказуальной фильтр симметричной формы $\mathbf{W}(z) = \sum_{p=-L/2}^{L/2} \mathbf{W}_p \cdot z^{-p}$.

Геометрические структуры многообразия КИХ-фильтров

При создании обучающего алгоритма использование оптимизации функции стоимости на осно-

ве m -градиента эффективно лишь для задачи итеративной оценки параметров. Для случая оптимизации функции стоимости, включающей КИХ-фильтр как целое (т. е. весь набор его параметров), метод m -градиента не оптимален. Поэтому при создании эффективных обучающих алгоритмов для получения оценки параметров размешивающего фильтра полезно рассмотреть подход на основе геометрических свойств многообразия КИХ-фильтров.

Множество $\mathcal{M}(L)$ всех КИХ-фильтров $\mathbf{W}(z)$ длины L (с ограничением, что матрица \mathbf{W}_0 обратима)

$$\mathcal{M}(L) = \left\{ \mathbf{W}(z) \mid \mathbf{W}(z) = \sum_{p=0}^L \mathbf{W}_p z^{-p}, \det(\mathbf{W}_0) \neq 0 \right\}$$

имеет размерность $n^2(L+1)$. В общем случае умножение двух фильтров в $\mathcal{M}(L)$ приводит к увеличению длины у результирующего фильтра. Поэтому, чтобы использовать возможные геометрические структуры в $\mathcal{M}(L)$, которые приведут к эффективным обучающим алгоритмам для $\mathbf{W}(z)$, следует определить алгебраические операции фильтров по концепции операций в группе Ли.⁴⁾

Использование m -градиента для РСС с операцией, обратной многоканальной свертке (ООМС)

Группа Ли имеет важное свойство — она допускает инвариантную метрику. С использованием структуры группы Ли можно получить m -градиент

⁴⁾ Для многообразия $\mathcal{M}(L)$ операции в группе Ли имеют своеобразную форму. Умножение \otimes и операция обращения \lrcorner определены следующим образом. Для $\mathbf{W}(z), \mathbf{H}(z) \in \mathcal{M}(L)$: $\mathbf{W}^\lrcorner(z) = \sum_{p=0}^L \mathbf{W}_p^\lrcorner \cdot z^{-p}$; $\mathbf{W}(z) \otimes \mathbf{H}(z) = \sum_{p=0}^L \sum_{q=0}^p \mathbf{W}_p \mathbf{H}_{(p-q)} z^{-p}$, где \mathbf{W}_p^\lrcorner рекуррентно определено выражениями: $\mathbf{W}_p^\lrcorner = \mathbf{W}_0^{-1}$; $\mathbf{W}_p^\lrcorner = -\sum_{q=1}^p \mathbf{W}_{p-q}^\lrcorner \mathbf{W}_q \mathbf{W}_0^{-1}$, $p=1, 2, \dots, L$. С этими операциями как $\mathbf{W}(z) \otimes \mathbf{H}(z)$, так и \mathbf{W}_p^\lrcorner остаются в многообразии $\mathcal{M}(L)$. При этом $\mathcal{M}(L)$ (с введенными выше операциями) образует группу Ли. Единичным элементом в группе $\mathbf{E}(z)$ является единичная матрица \mathbf{I} . Фактически умножение двух в группе Ли $\mathbf{W}(z)$ и $\mathbf{H}(z) \in \mathcal{M}(L)$ — это усеченная (до порядка L) форма обычного умножения, т. е. $\mathbf{W}(z) \otimes \mathbf{H}(z) = [\mathbf{W}(z) \cdot \mathbf{H}(z)]_L$, где $[\mathbf{W}(z)]_L$ — оператор усечения такой, что любые члены порядка выше L в матричном полиноме $\mathbf{W}(z)$ опускаются. Если для $\mathbf{W}(z)$ величина L берется достаточно большой, то флуктуации будут незначительными. Однако умножение в смысле группы Ли дает $\mathbf{G}(z) = \mathbf{W}(z) \otimes \mathbf{H}(z) = \mathbf{I}$. Далее будет рассматриваться объединенная передаточная функция (в смысле группы Ли) $\mathbf{G}(z) = \mathbf{W}(z) \otimes \mathbf{H}(z)$.

функции стоимости $\rho(\mathbf{W}(z))$, определенной на многообразии $\mathcal{M}(L)$:

$$\tilde{\nabla}\rho(\mathbf{W}(z)) = \frac{\partial\rho(\mathbf{W}(z))}{\partial\mathbf{X}(z)} \otimes \mathbf{W}(z) = \nabla\rho(\mathbf{W}(z)) \otimes \mathbf{W}(z),$$

где $d\mathbf{X}(z)$ — переменная, определенная соотношением:⁵⁾

$$d\mathbf{X}(z) = d\mathbf{W}(z) \otimes \mathbf{W}^{-1}(z) = [d\mathbf{W}(z)\mathbf{W}^{-1}(z)]_L. \quad (37)$$

Альтернативно м-градиент может быть выражен в виде

$$\tilde{\nabla}\rho(\mathbf{W}(z)) = \nabla\rho(\mathbf{W}(z)) \otimes \mathbf{W}^T(z^{-1}) \otimes \mathbf{W}(z).$$

Однако оценивать м-градиент значительно легче при введении (неголономной) дифференциальной переменной $d\mathbf{X}(z)$, определенной приведенным выше соотношением. Вычисление $\frac{\partial\rho(\mathbf{W}(z))}{\partial\mathbf{X}(z)}$ может быть выполнено двумя способами:

1. Оценить его по соотношению

$$\frac{\partial\rho(\mathbf{W}(z))}{\partial\mathbf{X}(z)} = \frac{\partial\rho(\mathbf{W}(z))}{\partial\mathbf{W}(z)} \otimes \mathbf{W}^T(z^{-1}).$$

2. Непосредственно вычислить его с использованием свойства $dy(k) = d\mathbf{W}(z)\mathbf{x}(k) = d\mathbf{X}(z)\mathbf{y}(k)$. Из этого выражения видно, что дифференциал $d\mathbf{X}(z)$ определяет изменения, которые переносятся каналом РСС на сигнал выхода $\mathbf{y}(k)$. Это свойство является основным фактором, влияющим на вывод обучающихся алгоритмов для модели разделения сигналов при ООМС.

Принимая, что $d\mathbf{X}(z) = \sum_{p=0}^L d\mathbf{X}_p z^{-p} \in \mathcal{M}(L)$ и что $d\mathbf{X}(z)$ — функция стоимости, определенная на

$$\mathcal{M}(L), \text{ можно выразить } \frac{\partial\rho(\mathbf{X}(z))}{\partial\mathbf{X}_p} = \left(\frac{\partial\rho(\mathbf{X}(z))}{\partial\mathbf{X}_{p,ij}} \right)_{n \times n}.$$

Из этого следует, что

$$\frac{\partial\rho(\mathbf{X}(z))}{\partial\mathbf{X}(z)} = \sum_{p=0}^L \frac{\partial\rho(\mathbf{X}(z))}{\partial\mathbf{X}_p} z^{-p}.$$

Оценивающая функция для ООМС обозначена, как $\mathbf{F}(\mathbf{y}, \mathbf{X}(z)) = \sum_{p=0}^L \mathbf{F}_p(\mathbf{y}, \mathbf{X}(z)) z^{-p}$, где $\mathbf{F}_p \in \mathbf{R}^{n \times n}$, $p = 0, 1, 2, \dots, L$ — матричные функции на $\mathcal{M}(L)$. При задании p и q производная $\frac{\partial\mathbf{F}_p}{\partial\mathbf{X}_q}$ является четырехмерным тензором, который опреде-

⁵⁾ Эту переменную называют неголономной — обеспечивающей дополнительную связь.

$$\text{лен соотношением } \frac{\partial\mathbf{F}_p}{\partial\mathbf{X}_q} = \left(\frac{\partial\mathbf{F}_{p,ij}}{\partial\mathbf{X}_{q,ij}} \right)_{n \times n \times n \times n}.$$

Для любой матрицы $\mathbf{F}_p \in \mathbf{R}^{n \times n}$ операция $\frac{\partial\mathbf{F}_p}{\partial\mathbf{X}_q} \mathbf{P}$ имеет вид $\frac{\partial\mathbf{F}_p}{\partial\mathbf{X}_q} \mathbf{P} = \sum_{l,k} \frac{\partial\mathbf{F}_p}{\partial\mathbf{X}_{q,ij}} P_{lk}$. Поэтому производная $\frac{\partial\mathbf{F}(\mathbf{y}, \mathbf{X}(z))}{\partial\mathbf{X}(z)}$ является оператором, отображающим $\mathcal{M}(L)$ на $\mathcal{M}(L)$, который представляется в виде соотношения

$$\frac{\partial\mathbf{F}(\mathbf{y}, \mathbf{X}(z))}{\partial\mathbf{X}(z)} \mathbf{P}(z) = \sum_{p=0}^L \sum_{q=0}^L \frac{\partial\mathbf{F}_p}{\partial\mathbf{X}_q} \mathbf{P}_q z^{-p},$$

для любого фильтра $\mathbf{P}(z) \in \mathcal{M}(L)$.

На основе элементов проведенного выше анализа осуществлено построение обучающихся по м-градиенту алгоритмов для ООМС. Для выполнения адаптивного обучения в РВ использована функция стоимости в виде

$$J(\mathbf{y}, \mathbf{W}(z)) = E\{\rho(\mathbf{y}, \mathbf{W}(z))\} = -\log |\det(\mathbf{W}_0)| - \sum_{i=1}^n E\{\log q(y_i)\},$$

где $q(y_i)$ — оценка для реальной функции ПРВ сигналов источника.

Оценка полного дифференциала $d\rho(\mathbf{y}, \mathbf{W}(z))$ имеет вид

$$d\rho(\mathbf{y}, \mathbf{W}(z)) = d\left(-\log |\det(\mathbf{W}_0)| - \sum_{i=1}^n \log q(y_i)\right) = -\text{tr}(d\mathbf{W}_0 \mathbf{W}_0^{-1}) + \boldsymbol{\Phi}^T(\mathbf{y})(\mathbf{y})^T d\mathbf{y}, \quad (39)$$

где tr — след матрицы, $\boldsymbol{\Phi}(\mathbf{y})$ — вектор нелинейных активационных функций с компонентами $\varphi_i(y_i)$

$$\varphi_i(y_i) = -\frac{d \log q_i(y_i)}{d y_i} = -\frac{q'_i(y_i)}{q_i(y_i)}.$$

Введение неголономной дифференциальной связи по типу (37) позволяет преобразовать (39) к выражению, имеющему форму $d\rho(\mathbf{y}, \mathbf{W}(z)) = -\text{tr}(d\mathbf{X}_0) + \boldsymbol{\Phi}^T(\mathbf{y}) d\mathbf{W}(z) \mathbf{W}^{-1}(z) \mathbf{y}$. При этом как следствие получается соотношение, определяющее компоненты м-градиента:

$$\frac{\partial\rho(\mathbf{y}, \mathbf{W}(z))}{\partial\mathbf{X}_p} = -\delta_{0p} \mathbf{I} + \boldsymbol{\Phi}^T(\mathbf{y}) \mathbf{y}^T (k-p), \quad p = 1, 2, \dots, L. \quad (40)$$

После этого с использованием метода (м-) градиентного спуска получается эффективный обучающийся алгоритм в РВ, имеющий вид

$$\begin{aligned} \Delta \mathbf{W}_p(k) &= -\eta(k) \sum_{q=0}^p \frac{\partial \rho(\mathbf{W}(z))}{\partial \mathbf{X}_p} \mathbf{W}_{p-q}(k) = \\ &= \eta(k) \sum_{q=0}^p [\delta_{0q} \mathbf{I} - \boldsymbol{\Phi}(\mathbf{y}(k)) \mathbf{y}^T(k-q)] \mathbf{W}_{p-q}(k) \end{aligned} \quad (41)$$

для $p = 1, 2, \dots, L$,

где η — параметр скорости обучения. В частности, обучающийся алгоритм для \mathbf{W}_0 имеет выражение

$$\Delta \mathbf{W}_0(k) = \eta(k) [\mathbf{I} - \boldsymbol{\Phi}(\mathbf{y}(k)) \mathbf{y}^T(k)] \mathbf{W}_0(k). \quad (42)$$

Альтернативно может быть использован адаптивный алгоритм с обучением его на выборке измерительных сигналов смеси, т. е. сигналов с сенсоров ИИС:

$$\Delta \mathbf{W}_p(k) = \eta(k) \sum_{q=1}^p [\delta_{0q} \mathbf{I} - \mathbf{R}_{\varphi \mathbf{y}}^{(k)}(q)] \mathbf{W}_{p-q}(k), \quad (43)$$

где $\mathbf{R}_{\varphi \mathbf{y}}^{(k)}(q) = (1 - \eta_0) \mathbf{R}_{\varphi \mathbf{y}}^{(k-1)}(q) + \eta_0 \boldsymbol{\Phi}(\mathbf{y}(k)) \mathbf{y}^T(k-q)$.

Алгоритмы на основе м-градиента (41) и (43) обладают свойством получать решение для размешивающей матрицы \mathbf{W} с примерно равной скоростью для ее компонент. Упомянутым свойством обладают алгоритмы, динамическое поведение которых зависит от общей передаточной функции (ПФ) $\mathbf{G}(z) = \mathbf{W}(z) \otimes \mathbf{H}(z)$, а не от ПФ смешивающего фильтра $\mathbf{H}(z)$. Фактически, обучающийся алгоритм (41) обладает этим свойством в смысле группы Ли. Так, если написать (первую часть) (41) по правилам группы Ли и умножить обе части этого соотношения на смешивающий фильтр $\mathbf{H}(z)$ (по правилам группы Ли), то будет получено выражение

$$\Delta \mathbf{G}(z) = -\eta \frac{\partial \rho(\mathbf{W}(z))}{\partial \mathbf{X}(z)} \otimes \mathbf{G}(z), \quad (44)$$

где $\mathbf{G}(z) = \mathbf{W}(z) \otimes \mathbf{H}(z)$.

Поскольку из (40) видна формальная независимость $\frac{\partial \rho(\mathbf{y}, \mathbf{W}(z))}{\partial \mathbf{X}_p}$ от смешивающего канала $\mathbf{H}(z)$,

то динамические свойства алгоритма (44) зависят от $\mathbf{G}(z)$, а сам алгоритм обладает тем же свойством, как у алгоритмов (41) и (43).

Другое важное свойство алгоритма (42) состоит в том, что он поддерживает обратимость матрицы \mathbf{W}_0 , если начальное значение \mathbf{W}_0 обратимо [27].

ОЦЕНИВАЮЩИЕ ФУНКЦИИ И СТАНДАРТИЗОВАННЫЕ ОЦЕНИВАЮЩИЕ ФУНКЦИИ ДЛЯ РСС НА ОСНОВЕ ООМС

Наиболее выигрышной является ОФ, которая представлена соотношением

$$\mathbf{F}(\mathbf{x}(k), \mathbf{W}(z)) = \sum_{p=0}^L \boldsymbol{\Phi}(\mathbf{y}(k)) \mathbf{y}(k-p)^T z^{-p} - \mathbf{I}, \quad (45)$$

где $\mathbf{y}(k) = \sum_{p=0}^L \mathbf{W}_p \mathbf{x}(k-p)$; $\boldsymbol{\Phi}$ — вектор данной активационной функции; соблюдается условие, что оператор производной $\mathbf{K}(z) =$

$= \mathbf{E} \left\{ \frac{\partial \mathbf{F}(\mathbf{x}, \mathbf{W}(z))}{\partial \mathbf{X}(z)} \right\}$ обратим. ОФ является эффективной, когда дополнительно к этому выполняется соотношение $F_{ii}(y_i) = \varphi_i(y_i) y_{i-1}$ [26], [27].

Оператор производной $\mathbf{K}(z) = \mathbf{E} \left\{ \frac{\partial \mathbf{F}(\mathbf{x}, \mathbf{W}(z))}{\partial \mathbf{X}(z)} \right\}$ — это фильтр тензорного типа. Он может быть представлен в форме $\mathbf{K}(z) = \sum_{p=0}^L \mathbf{K}_p z^{-p}$, где

$$\mathbf{K}_{p,ij,lm} = \mathbf{E} \{ \varphi'_i(y_i(k)) (s_j^2(k-p)) \} \delta_{il} \delta_{jm} + \delta_{im} \delta_{jl} \delta_{0p}.$$

Кроме того, при выполнении условия $k_i \neq 0$, $k_i k_j \sigma_i^2 \sigma_j^2 - 1 \neq 0$, $m_i + 1 \neq 0$ производная оператора $\mathbf{K}(z)$ обратима. В последнем условии использованы обозначения, введенные Цхангом:

$$m_i = \mathbf{E} \{ y_i^2 \varphi'_i(y_i) \}, \quad k_i = \mathbf{E} \{ \varphi'_i(y_i) \}, \quad \sigma_i^2 = \mathbf{E} \{ y_i^2 \}. \quad (46)$$

$$c_{ij} = [k_i k_j \sigma_i^2 \sigma_j^2 - 1]^{-1}, \quad l_i = \mathbf{E} \{ \varphi(y_i) \}. \quad (47)$$

Метод, основанный на полупараметрической модели, для получения оценки параметров (z -преобразования) размешивающей матрицы, использует соотношение $\sum_{k=1}^N \mathbf{F}(\mathbf{x}(k), \mathbf{W}(z)) = 0$. При достаточно большом значении N оценка $\mathbf{W}(z, k)$ сходится к фактическому размешивающему фильтру, причем для этого не требуется знания ПРВ сигналов источника $\mathbf{r}(s)$. ОФ не является единственной, т. к. для любого обратимого оператора $\mathbf{R}(z)$ (отображающего $\mathbf{M}(L)$ на $\mathbf{M}(L)$) $\mathbf{R}(z) \mathbf{F}(\mathbf{x}, \mathbf{W}(z))$ — также ОФ. При этом, как было установлено, две ОФ эквивалентны в том смысле, что выведенные из них оценки размешивающей матрицы по алгоритму с обучающей выборкой совершенно одинаковы. Однако анализ алгоритма РВ показывает, что динамика обучения различается и, следовательно, целесообразно вводить ОФ, которая обеспечивает более эффективный и устойчивый обучаемый алгоритм. Для этой цели вводится концепция стандартизованной ОФ. ОФ называется стандартизованной, если оператор производной

$$\mathbf{K}(z) = \mathbf{E} \left\{ \frac{\partial \mathbf{F}(\mathbf{x}, \mathbf{W}(z))}{\partial \mathbf{X}(z)} \right\}$$
 является единичным оператором.

Далее, если оператор $\mathbf{K}(z)$ обратим, то для любой ОФ $\mathbf{F}(\mathbf{x}, \mathbf{W}(z))$ преобразование оператором $\mathbf{K}^{-1}(z)$, т. е. $\mathbf{K}^{-1}(z) \mathbf{F}(\mathbf{x}, \mathbf{W}(z))$, является стандартизованной ОФ.

Кроме того, при эффективной ОФ в форме (45) соответствующее выражение стандартизованной ОФ имеет вид

$$\mathbf{F}^*(\mathbf{x}, \mathbf{W}(z)) = \sum_{p=0}^L \mathbf{F}_p^*(\mathbf{x}, \mathbf{W}(z)) z^{-p}, \quad (48)$$

где

$$F_{0,ii}^* = \frac{1}{m_i + 1} \{ \varphi_i(y_i) y_i - 1 \} \quad \text{для } i = 1, 2, \dots, n;$$

$$F_{0,ij}^* = c_{ij} \{ k_j \sigma_i^2 \varphi_i(y_i) y_j - \varphi_j(y_j) y_i \} \quad \text{для } i \neq j;$$

$$F_{p,ij}^* = \varphi_i(y_i) y_j (k - p) / (k_i \sigma_j^2) \quad \text{для } p \geq 1.$$

Использование стандартизованной формы ОФ имеет определенные преимущества при получении алгоритма с обучением в РВ для оценки размещающей матрицы. Алгоритм обучения на основе использования м-градиента определяется соотношением

$$\Delta \mathbf{W}(z) = -\eta \cdot \mathbf{F}^*(\mathbf{x}, \mathbf{W}(z)) \otimes \mathbf{W}(z). \quad (49)$$

Этот алгоритм обучения при выполнении условия (47) приводит к устойчивому равновесию, которое обеспечивает получение правильного решения для размещающей матрицы $\mathbf{W}(z) = \mathbf{H}^{-1}(z)$. Для выполнения обучающего алгоритма (49) требуется в РВ оценивать статистики (46) и (47). В частности, если ПС бинарные (принимающие значения 1, -1), то вычисление таких статистик для стандартизованной ОФ достаточно просто. Если принять в качестве активационной функции кубическую функцию $\varphi_i(y_i) = y_i^3$, то статистики оцениваются по соотношениям: $m_i = 3$, $k_i = 3$, $\sigma_i^2 = 1$, $\tilde{\gamma}_{ij} = c_{ij}^{-1} = 8$. Поэтому стандартизованные ОФ могут быть получены в явном виде.

Повышение эффективности при использовании оценок с обучающей выборкой

Показано, что мгновенная замена $V_{ij}(N)$ ковариационной матрицы $E\{y_i y_j\}$ ($i \neq j$) при больших значениях выборки ПС убывает со скоростью $1/N^2$. Это свойство называется повышением эффективности при соответствующем оценивании. Цхангом [6], [26] показано, что это свойство сохраняет справедливость при РСС с операцией, обратной относительно многоканальной свертки ПС.

Положим, что $\mathbf{F}^*(\mathbf{x}, \mathbf{W}(z))$ — это стандартизованная оценивающая функция

$$E\{\Delta \mathbf{X}_N(z, k) \otimes \Delta \mathbf{X}_N^T(z, k)\} = \frac{1}{N} \mathbf{G}^*(z) + O\left(\frac{1}{N^2}\right),$$

где $\mathbf{G}^*(z) = \mathbf{K}^{-1}(z) \mathbf{G}(z) \mathbf{K}^{-T}(z) =$

$$= E\{\mathbf{F}^*(\mathbf{x}, \mathbf{W}(z)) \otimes \mathbf{F}^{*T}(\mathbf{x}, \mathbf{W}(z))\}.$$

Тогда коэффициенты $\mathbf{G}^*(z)$ выражаются соотношениями:

$$G_{0,il,jl}^* = c_{il} c_{jl} \sigma_i^2 \sigma_j^2 \sigma_l^2 k_i^2 l_i l_j \quad \text{для } i \neq j, j \neq l, l \neq i;$$

$$G_{0,ii,ji}^* = \frac{1}{m_i + 1} c_{ji} k_i \sigma_j^2 l_j E\{s_i^2 \varphi_i(s_i)\} \quad \text{для } i \neq j;$$

$$G_{p,il,jl}^* = \frac{l_i l_j}{k_i k_j} \quad \text{для } p \geq 1, i, j = 1, 2, \dots, n.$$

Используя для $\mathbf{F}^*(\mathbf{x}, \mathbf{W}(z))$ соотношение (48), путем прямых вычислений может быть получен результат, который формулируется следующим образом. Оценка по выборке является эффективной при выполнении условия $l_i = E\{\varphi_i(s_i)\} = 0$ для $i = 1, 2, \dots, n$.

Из результатов анализа, проведенного в этом и двух предшествующих разделах, следует, что для повышения эффективности при оценке по выборке и по алгоритму на основе м-градиента требуется выполнение одного и того же условия: $l_i = E\{\varphi_i(s_i)\} = 0$ для $i = 1, 2, \dots, n$. Но поскольку в качестве активационной функции нейронных сетей, с помощью которых реализуются алгоритмы, применяются кубическая функция или гиперболический тангенс, то обе эти функции удовлетворяют требуемому условию.

ЗАКЛЮЧЕНИЕ

Рассмотрены оценивающие функции и метод полупараметрической модели для разделения сигналов смеси (РСС), когда смесь реализована в форме многоканальной свертки первичных сигналов (сигналов источника). Для метода определения размещающей матрицы \mathbf{W} (обратной относительно \mathbf{H} — матрицы, смешивающей ПС) проанализирована сходимость и условия, обеспечивающие повышение эффективности оценок \mathbf{W} как при использовании алгоритма, обучаемого по выборке, так и алгоритма РВ на основе м-градиента.

На первом этапе задача РСС при смеси в форме многоканальной свертки ПС сформулирована в рамках концепции полупараметрической модели, семейства оценивающих функций и стандартизованных оценивающих функций. Преимущество метода, основанного на полупараметрической модели, состоит в том, что в задаче РСС она избавляет от оценки мешающих параметров, т. е. от плотностей вероятности распределения сигналов источника. Из анализа теории оценивающих функций следует, что оценки размещающей матрицы алгоритмом, обучаемым по выборке (если объ-

ем ее достаточно велик), сходятся к реальному решению. При этом, если удовлетворены условия устойчивости, то обучение алгоритма на основе м-градиента также приводит к сходимости к реальному решению независимо от плотности распределения первичных сигналов. Повышение эффективности обоих алгоритмов обеспечивается при выполнении определенных локальных условий.

В статье рассмотрены основные элементы двух математических концепций, использованных для целесообразной формы описания действий на многообразии матричных передаточных функций (МПФ) и на многообразии КИХ-фильтров. Это — операции МПФ на группе Ли и геометрические структуры многообразия КИХ-фильтров.

Идея, концепция и метод использования полупараметрической статистической модели, а также семейства оценивающих функций разработаны Цхангом, Амари, Сичоки [6], [26], Дугласом [8], Нагаоки [10], Пхамом [18], Каванабе, Муратой [21], Икеда, Тойамой [22], Тонгом [23].

СПИСОК ЛИТЕРАТУРЫ

1. *Chen T.-P., Amari S., Lin Q.* A unified algorithm for principal and minor component extraction // *Neural Networks*. 1998. V. 11. P. 385–390.
2. *Choi S., Cichocki A.* Separation of non-stationary sources in noisy mixtures // *Electronic Letters*. 2000. V. 36. P. 848–849.
3. *Малыхина Г.Ф., Меркушева А.В.* Адаптивный алгоритм на основе рекуррентной сети для задачи восстановления формы линейно смешанных сигналов // *Нейрокомпьютеры*. 2006. № 7. С. 3–13
4. *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.
5. *Малыхина Г.Ф., Меркушева А.В.* Метод анализа независимых компонент для восстановления формы сигналов по их смеси // *Информационные технологии*. 2008. № 9. С. 39–42.
6. *Zhang L., Amari S., Cichocki A.* Semi-parametric model and super efficiency in blind de-convolution // *Signal Processing*. 2001. V. 81, N 12. P. 2535–2553.
7. *Малыхин 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. 120–132.
8. *Amari S., Douglas S.C., Cichocki A.* Information geometry of source de-convolution // *Mathematical Theory of Networks and Systems*, MTNS 98, Padova, Italy, July 1998.
9. *Amari S., Kawanabe M.* Information geometry of estimating functions in semi-parametric models // *Bernulli*. 1997. V. 3, N 1. P. 29–54.
10. *Amari S., Nagaoka H.* Methods of information geometry (Translations of Mathematical Monographs. V. 191). AMS & Oxford University Press. 1999. 205 p.
11. *Jutten C., Herault J.* Blind separation of sources. I. An adaptive algorithm based on neuro-mimetic architecture // *Signal Processing*. 1991. V. 24, N 1. P. 1–10.
12. *Bell A.J., Sejnowski T.J.* Learning and higher order structure of natural sounds // *Neural computation in neural systems*. 1996. N 7. P. 261–266.
13. *Amari S., Cichocki A., Yang H.H.* A new learning algorithm for blind signal separation // *Advances in Neural Networks Information Processing Systems / D.S. Touretzky, M.C. Mozer, E.E. Hasselmo (eds)*. Cambridge: MIT Press, 1996. V. 8. P. 757–763.
14. *Cardoso J.-F., Laheld B.H.* Equivariant adaptive source separation // *IEEE Trans. Signal Processing*. 1996. V. 44, N 12. P. 3017–3030.
15. *Oja E., Karhunen J.* Signal separation by nonlinear hebbian learning // *Computational Intelligence — a Dynamic System Perspective*. NY.: IEEE Press, 1995. P. 83–97.
16. *Amari S., Cardoso J.-F.* Blind source separation: semi-parametric statistical approach / *IEEE Trans. Signal Processing*. 1997. V. 45, N 11. P. 2692–2700.
17. *Amari S., Kawanabe M.* Estimating functions in semi-parametric models // *Estimating Functions (Monograph Series)*. IMS, 1998. V. 32 / Basawa I.V., Godambe V., Teylor R. (eds). P. 65–80.
18. *Pham D.-T.* Separation of instantaneous mixture of sources via an independent analysis // *IEEE Trans. Signal Processing*. 1996. V. 44, N 11. P. 2768–2779.
19. *Amari S., Chen T.-P., Cichocki A.* Stability analysis for adaptive blind source separation // *Neural Networks*. 1997. V. 10, N 8. P. 1345–1351.
20. *Zhang L., Amari S., Cichocki A.* Equi-convergence algorithm for blind separation of sources with arbitrary distributions // *Bio-Inspired Applications of Connectionism*. Springer, Granada, Spain, 2001. V. LNCS 2085 / Mira J., Prieto A. (Eds). P. 626–833.
21. *Kawanabe M., Murata N.* Independent component analysis in presence of Gaussian noise based on estimating functions // *Proceedings of the Second Workshop on Independent Component Analysis and Blind Signal Separation*, 2000. P. 39–44.
22. *Ikeda S., Toyama K.* Independent component analysis for noisy data – MEG data analysis // *Neural Networks*. 2001. V.13, N 10. P. 1063–1074.
23. *Tong L., Liu R.W., Soon V.C., Huang Y.F.* Indeterminacy and identifiability of blind identification // *IEEE Trans. Circuits and Systems*. 1991. V. 38, N 5. P. 499–509.
24. *Common P.* Independent component analysis, a new concert // *Signal Processing*. 1998. V. 36, N 3. (Special Issue on Higher Order Statistics). P. 287–314.
25. *Amari S.* Estimating functions for independent component analysis for temporally correlated signals // *Neural Computation*. 2000. V. 12, N 9. P. 2083–2107.
26. *Zhang L., Cichocki A., Amari S.* Geometrical structures of FIR manifolds and their application to multi-channel blind de-convolution // *Proceeding of Int. IEEE Workshop on Neural Networks for Signal (NNSP'99)*, Madison, USA, August 23–25, 1999. P. 404–312.
27. *Yang H.H., Amari S.* Adaptive on line learning algorithm for blind separation: Maximum entropy and minimal mutual information // *Neural Computation*. 1997. N 9. P. 1457–1482.

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

СПбГПУ, Санкт-Петербург

Материал поступил в редакцию 24.01.2011.

METHOD OF EFFICIENCY INCREASING FOR INDEPENDENT COMPONENTS AND CONVOLUTION ANALYSIS PROCEDURES THAT ARE USED FOR SIGNAL FORM RECONSTRUCTION ON THE BASE OF MIXTURE MEASUREMENTS

A. V. Merkusheva, G. F. Malychina

SPbSPU, Saint-Petersburg

The conception is considered that generate unified view-point on different algorithms for signal separation (including de-convolution of multi-channel mixtures) on the base of signal mixture measurements. Estimating functions method is considered that give the possibility to explain the structure of different adaptive algorithms that realize de-separating matrix estimation and signal form reconstruction using only mixture measurements. The problem of signal mixture separation is formulated on the base of semi-parametric statistical model and estimating functions family. In terms of Li group formulation there is filter manifold considered.

Keywords: signal mixtures, multi-channel convolution, signal separation, algorithms, efficiency criteria, de-mixing matrix, Li group, FIR filters, geometrical structures