УДК 519.685.3

Н.Ф. КИРИЧЕНКО, С.А. ГАВРИЛЕНКО, А.С. ГАВРИЛЕНКО

ОПТИМИЗАЦИЯ ПАРАМЕТРОВ ФУНКЦИОНАЛЬНЫХ ПРЕОБРАЗОВАНИЙ В СИСТЕМЕ КЛАССИФИКАЦИИ СИГНАЛОВ

Ключевые слова: *кластеризация, синтез систем классификации, распознавание образов, условия линейной и нелинейной отделимости и разделимости множеств, псевдообратные и проекционные матрицы.*

ВВЕДЕНИЕ

Результаты, полученные в настоящей работе, являются развитием метода синтеза линейных и нелинейных систем классификации сигналов [1, 2]. Предложены способы и методы интерпретации процесса синтеза для системы преобразования сигналов к операциям инверсии, псевдоинверсии и проецирования линейных и нелинейных преобразований. Постановки задач синтеза систем, рассматриваемых в указанных источниках, примыкают к идеям методов МГУА [3] и Support Vector Machine [4]. Использование средств псевдообращения операций и результатов по представлению изменений псевдообратных и проекционных операций при возмущении исходных данных [5, 6] позволило с новой точки зрения рассмотреть задачу оптимального синтеза как линейных, так и нелинейных систем, осуществляющих дихотомное разделение сигналов из обучающей выборки. Далее приводятся алгоритмы и критерии линейной отделимости и разделимости множества точек в конечномерном пространстве и некоторое обобщение этих результатов на бесконечномерное пространство, а также алгоритмы оптимального синтеза нелинейных операций, обладающих параллельной и последовательной структурой с использованием базовых нелинейных преобразований в класс полиномов третьего порядка. Выбор конкретного примера базовых преобразований позволяет пользователю иметь более точное представление о возможностях предлагаемых средств.

УСЛОВИЯ ОТДЕЛИМОСТИ И РАЗДЕЛИМОСТИ ТОЧЕК В ПРОСТРАНСТВЕ ПРИЗНАКОВ

Известные условия линейной отделимости выпуклого множества теоремы Хана–Банаха допускают за счет использования операций псевдоинверсии матриц и оптимизации квадратичных форм конструктивное расширение на конечные невыпуклые множества, рассматриваемые обычно в задачах классификации информации. Условия линейной отделимости множества можно затем интерпретировать в форме условий линейной разделимости двух множеств между собой. Ниже приведем формулировку этих свойств.

Вначале рассмотрим условия линейной отделимости конечного множества точек $x(j) \in R^m$, $j=\overline{1,n}$. Так как условие отделимости этого множества от начала координат в смысле

$$\exists\, a \in R^m,\ \|a\| > 0,\ a^{\mathrm{T}} x(j) \ge 1 \ \ \forall j=\overline{1,n}$$

эквивалентно разрешимости системы линейных алгебраических уравнений

$$X^{\mathrm{T}} a = y,\ X=(x(1)\dots x(n))=\begin{pmatrix} x_{(1)}^{\mathrm{T}} \\ \vdots \\ x_{(m)}^{\mathrm{T}} \end{pmatrix},\ y=\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

при некоторых значениях $y_j \in R^1$, $j=\overline{1,n}$, для практического использования удобно его переформулировать в виде

$$\min_{y \in D} y^{\mathrm{T}} Z(X) y = y_*^{\mathrm{T}} Z(X) y_* = 0, \; D = \{ y : y = (y_1 \ldots \; y_n)^{\mathrm{T}}, \; y_j \geq 1, \; j = \overline{1, n} \},$$

при этом $a = X^{\mathrm{T}^+} y_*$ и расстояние от отделяющей гиперплоскости $a^{\mathrm{T}} x = 1$ до начала координат определяется величиной $h = \left(\sqrt{y_*^{\mathrm{T}} R(X) y_*} \right)^{-1}$; $Z(X) = I_n - X^+ X$, $R(X) = X^+ X^{+\mathrm{T}}$ — соответствующие проекционные матрицы. Оптимальная отделяющая гиперплоскость, обладающая максимальным расстоянием h, удовлетворяет условиям

$$a_{\mathrm{opt}} = X^{+\mathrm{T}} y_{\mathrm{opt}}, \tag{1}$$

$$y_{\mathrm{opt}}^{\mathrm{T}} R(X) y_{\mathrm{opt}} = \min_{y \in D_*} y^{\mathrm{T}} R(X) y, \tag{2}$$

$$h_{\mathrm{opt}} = \left(\sqrt{y_{\mathrm{opt}}^{\mathrm{T}} R(X) y_{\mathrm{opt}}} \right)^{-1}, \tag{3}$$

$$D_* = \{ y : y^{\mathrm{T}} Z(X) y = y_*^{\mathrm{T}} Z(X) y_*, \; e_j^{\mathrm{T}} y \geq 1, \; j = \overline{1, n} \}, \tag{4}$$

где e_j — j-й единичный орт в R^n.

Используя SVD для матрицы X, т.е. раскладывая ее по соответствующим собственным вектор-столбцам и вектор-строкам,

$$X = \sum_{j=1}^{r} u_j v_j^{\mathrm{T}} \lambda_j, \; r = \mathrm{rank}\, X,$$

оптимальные значения представим следующим образом:

$$y_{\mathrm{opt}} = (v_1 \vdots \cdots \vdots v_r) \alpha_{\mathrm{opt}}, \; a_{\mathrm{opt}} = (\lambda_1^{-1} u_1 \vdots \cdots \vdots \lambda_r^{-1} u_r) \alpha_{\mathrm{opt}},$$

$$h_{\mathrm{opt}} = \left(\sqrt{\alpha_{\mathrm{opt}}^{\mathrm{T}} \mathrm{diag}\, (\lambda_1^{-2}, \ldots, \lambda_r^{-2}) \alpha_{\mathrm{opt}}} \right)^{-1},$$

$$\alpha_{\mathrm{opt}}^{\mathrm{T}} \mathrm{diag}\, (\lambda_1^{-2}, \ldots, \lambda_r^{-2}) \alpha_{\mathrm{opt}} = \min_{\alpha \in D_\alpha} \alpha^{\mathrm{T}} \mathrm{diag}\, (\lambda_1^{-2}, \ldots, \lambda_r^{-2}) \alpha,$$

$$D_\alpha = \{ \alpha : e_j^{\mathrm{T}} (v_1 \vdots \cdots \vdots v_r) \alpha \geq 1, \; j = \overline{1, n} \}, \; \alpha \in R^r.$$

Учитывая свойство собственных векторов и значений $v_j = X^{\mathrm{T}} u_j \lambda_j^{-1}$, множество D_α выразим в виде $D_\alpha = \{ \alpha : x^{\mathrm{T}}(j)(u_1 \lambda_1^{-1} \ldots u_r \lambda_r^{-1}) \alpha \geq 1, \; j = \overline{1, n} \}$.

В случае, когда точки x заполняют некоторое ограниченное (не обязательно выпуклое) компактное множество Γ в пространстве R^1, выражения (1)–(4) можно распространить на множество точек $x(i) \in R^1$, $i = \overline{1, N}$, достаточно плотно распределенных на Γ. Очевидно, что отделимость множества Γ от начала координат следует из отделимости этого конечного множества и условий

$$\min_{j = \overline{1, N}} || x - x(j) ||^2 < h_{\mathrm{opt}}^2 \quad \forall x \in \Gamma.$$

Используя интегральные псевдообратные операции, по аналогии с предыдущими соотношениями получаем соответственно условие отделимости множества Γ

$$\min_{\{ y : \, y(x) \geq 1, \, x \in \Gamma \}} \left\{ \int_\Gamma y^2(x) dx - \int_\Gamma y(x) x^{\mathrm{T}} dx \left(\int_\Gamma x x^{\mathrm{T}} dx \right)^+ \int_\Gamma x y(x) \, dx \right\} = 0.$$

Применение собственных векторов, функций и значений позволяет определить оптимальную отделяющую гиперплоскость $a_{\mathrm{opt}}^{\mathrm{T}} x \geq 1 \;\; \forall x \in \Gamma$:

$$\int_\Gamma xx^{\mathrm T} dx u_j = \lambda_j^2 u_j, \ j=\overline{1,r}, \ r = \operatorname{rank} \int_\Gamma xx^{\mathrm T} dx, \ u_i^{\mathrm T} u_j = \delta_{ij}, \ i,j=\overline{1,N},$$

$$\alpha_{\mathrm{opt}}^{\mathrm T} \operatorname{diag}(\lambda_1^{-2},\ldots,\lambda_r^{-2})\alpha_{\mathrm{opt}} = \min_{\alpha\in D_\alpha} \alpha^{\mathrm T} \operatorname{diag}(\lambda_1^{-2},\ldots,\lambda_r^{-2})\alpha,$$

$$D_\alpha = \{\alpha : x^{\mathrm T}(\lambda_1^{-1}u_1 \vdots \cdots \vdots \lambda_r^{-1}u_r)\alpha \ge 1 \ \forall x\in\Gamma\},$$

$$a_{\mathrm{opt}} = (\lambda_1^{-1}u_1 \vdots \cdots \vdots \lambda_r^{-1}u_r)\alpha_{\mathrm{opt}}, \quad h_{\mathrm{opt}}^2 = (\alpha_{\mathrm{opt}}^{\mathrm T} \operatorname{diag}(\lambda_1^{-2},\ldots,\lambda_r^{-2})\alpha_{\mathrm{opt}})^{-1}.$$

Подобным образом определяются условия отделимости конечного множества, элементами которого являются функции непрерывного аргумента (временны́е сигналы, изображения лиц и т.п.), а значит, в качестве векторов $x(j)\in R^m$, $j=\overline{1,n}$, рассматриваются функции

$$x(1,\xi),\ldots,x(n,\xi), \ \xi\in\Gamma\subset R^m, \ x(j,\xi)\in R^1, \ j=\overline{1,n}.$$

Условие отделимости можно использовать в виде

$$\min_{y\in\{y:\, y=(y_1 \ldots y_n)^{\mathrm T},\, y_j\ge 1,\, j=\overline{1,n}\}} y^{\mathrm T} Z(P)\, y = 0, \quad P = \int_\Gamma \begin{pmatrix} x(1,\xi) \\ \vdots \\ x(n,\xi) \end{pmatrix} (x(1,\xi) \cdots x(n,\xi))\, d\xi .$$

Применение SVD дает возможность получить представление оптимальной отделяющей линейной операции для таких объектов:

$$Pv_j = \lambda_j^2 v_j \quad \forall j=\overline{1,r}, \ r=\operatorname{rank} P,$$

$$\alpha_{\mathrm{opt}}^{\mathrm T} \operatorname{diag}(\lambda_1^{-2},\ldots,\lambda_r^{-2})\alpha_{\mathrm{opt}} = \min_{\alpha\in D_\alpha} \alpha^{\mathrm T} \operatorname{diag}(\lambda_1^{-2},\ldots,\lambda_r^{-2})\alpha,$$

$$D_\alpha = \{\alpha : e_j^{\mathrm T}(v_1 \vdots \cdots \vdots v_r)\alpha \ge 1, \ j=\overline{1,n}\},$$

$$a_{\mathrm{opt}}(\xi) = (x^{\mathrm T}(\xi) v_1 \lambda_1^{-2} \vdots \cdots \vdots x^{\mathrm T}(\xi) v_r \lambda_r^{-2})\,\alpha_{\mathrm{opt}},$$

$$x^{\mathrm T}(\xi) = (x(1,\xi) \cdots x(n,\xi)), \qquad h_{\mathrm{opt}}^2 = (\alpha_{\mathrm{opt}}^{\mathrm T} \operatorname{diag}(\lambda_1^{-2},\ldots,\lambda_r^{-2})\alpha_{\mathrm{opt}})^{-1}.$$

Следовательно, в этом случае оптимальная линейная отделяющая принимает вид

$$\int_\Gamma a_{\mathrm{opt}}(\xi)\, x(j,\xi)\, d\xi \ge 1 \quad \forall j=\overline{1,n}.$$

Рассмотренные условия линейной отделимости множества несложно распространить на исследование задачи разделимости множеств, лежащей в основе современных средств классификации сигналов. Для двух конечных множеств точек $x(i_k)\in R^m$, $k=\overline{1,n_1}$, $x(j_s)\in R^m$, $s=\overline{1,n_2}$, линейную полосную разделимость будем определять при существовании такого вектора $\begin{pmatrix} a_1 \\ \vdots \\ a_{m+1} \end{pmatrix} \in R^{m+1}$, $a_{m+1}\in R^1$, для которого имеют место соотношения

$$\begin{pmatrix} a_1 \\ \vdots \\ a_{m+1} \end{pmatrix}^{\mathrm T} \begin{pmatrix} x(i_k) \\ 1 \end{pmatrix} \ge 1, \ k=\overline{1,n_1}, \quad \begin{pmatrix} a_1 \\ \vdots \\ a_{m+1} \end{pmatrix}^{\mathrm T} \begin{pmatrix} x(i_s) \\ 1 \end{pmatrix} \le -1, \ s=\overline{1,n_2}.$$

Условие наличия линейной полосной разделимости рассматриваемых множеств вытекает из ранее полученных рассуждений о линейной отделимости множеств и имеет вид

$$\min_{y\in D} y^{\mathrm T} Z\begin{pmatrix} X \\ \hline J_n^{\mathrm T}\end{pmatrix} y = 0\,,\;\; D=\{\, y:\; e_{i_k}^{\mathrm T}\, y \ge 1,\; e_{j_s}^{\mathrm T}\;\; y\le -1,\;\; k=\overline{1,n_1},\, s=\overline{1,n_2}\,\}\,, \tag{5}$$

$$X=(x(1)\ldots x(n)),\;\; x(i)\in\{x: x(i_k),\; k=\overline{1,n_1}\}\cup\{x: x(j_s),\, s=\overline{1,n_2}\},\;\; i=\overline{1,n}\,,$$

$$J_n=(1\vdots\cdots\vdots 1)^{\mathrm T}\;\in R^{\,n},$$

а максимальная толщина полосы разделимости достигается при

$$\begin{pmatrix} a_{\mathrm{opt}} \\ \hline a_{m+1\ \mathrm{opt}}\end{pmatrix}=\begin{pmatrix} X \\ \hline J_n^{\mathrm T}\end{pmatrix}^{+\mathrm T} y_{\mathrm{opt}},\quad y_{\mathrm{opt}}=\arg\min_{D_*} y^{\mathrm T} R\begin{pmatrix} X \\ \hline J_n^{\mathrm T}\end{pmatrix} y\,, \tag{6}$$

$$D_*=\left\{ y:\, y^{\mathrm T} Z\begin{pmatrix} X \\ \hline J_n^{\mathrm T}\end{pmatrix} y=0,\; e_{i_k}^{\mathrm T}\, y\ge 1,\; e_{j_s}^{\mathrm T}\;\; y\le -1,\; k=\overline{1,n_1},\; s=\overline{1,n_2}\right\}.$$

Поскольку в реальности $n \gg m$, целесообразно предположить общность положения линейной независимости вектор-строки $J_n^{\mathrm T}$ от вектор-строк матрицы X, а значит, согласно обращению формул Гревиля [3] имеют место соотношения

$$J_n^{\mathrm T}\begin{pmatrix} X \\ \hline J_n^{\mathrm T}\end{pmatrix}^{+} e_{m+1}=1,\;\; R(X)=Z\left(\begin{pmatrix} X \\ \hline J_n^{\mathrm T}\end{pmatrix}^{+} e_{m+1}\right) R\begin{pmatrix} X \\ \hline J_n^{\mathrm T}\end{pmatrix} Z\left(\begin{pmatrix} X \\ \hline J_n^{\mathrm T}\end{pmatrix}^{+} e_{m+1}\right)$$

и в терминах собственных векторов и значений получим следующую форму представления формул (6):

$$\begin{pmatrix} X \\ \hline J_n^{\mathrm T}\end{pmatrix}=\sum_{j=1}^{r} u_j v_j^{\mathrm T}\lambda_j\,,\;\; \begin{pmatrix} X \\ \hline J_n^{\mathrm T}\end{pmatrix}^{+}=\sum_{j=1}^{r} v_j\, u_j^{\mathrm T}\lambda_j^{-1},$$

$$y_{\mathrm{opt}}=(v_1\vdots\cdots\vdots v_r)\alpha_{\mathrm{opt}}\,,\;\begin{pmatrix} a_{\mathrm{opt}} \\ \hline a_{m+1\ \mathrm{opt}}\end{pmatrix}=(\lambda_1^{-1}u_1\vdots\cdots\vdots\lambda_r^{-1}u_r)\alpha_{\mathrm{opt}}\,,$$

$$\left\|\mathrm{diag}\,(\lambda_1^{-1},\ldots,\lambda_r^{-1})\left(I_r-\frac{\begin{pmatrix}\lambda_1^{-1}u_1^{\mathrm T}\\ \vdots \\ \lambda_r^{-1}u_r^{\mathrm T}\end{pmatrix} e_{m+1}\, e_{m+1}^{\mathrm T}(\lambda_1^{-1}u_1\vdots\cdots\vdots\lambda_r^{-1}u_r)}{e_{m+1}^{\mathrm T}\left(\sum_{j=1}^{r}\lambda_j^{-2}u_j u_j^{\mathrm T}\right)e_{m+1}}\right)\alpha_{\mathrm{opt}}\right\|^2 =$$

$$=\min_{\alpha\in D_\alpha}\left\|\mathrm{diag}\,(\lambda_1^{-1},\ldots,\lambda_r^{-1})\left(I_r-\frac{\begin{pmatrix}\lambda_1^{-1}u_1^{\mathrm T}\\ \vdots \\ \lambda_r^{-1}u_r^{\mathrm T}\end{pmatrix} e_{m+1}\, e_{m+1}^{\mathrm T}(\lambda_1^{-1}u_1\vdots\cdots\vdots\lambda_r^{-1}u_r)}{e_{m+1}^{\mathrm T}\left(\sum_{j=1}^{r}\lambda_j^{-2}u_j u_j^{\mathrm T}\right)e_{m+1}}\right)\alpha\right\|^2,$$

$$D_\alpha=\{\alpha: e_{i_k}^{\mathrm T}\,(v_1\vdots\cdots\vdots v_r)\alpha\ge 1,\; e_{j_s}^{\mathrm T}\;(v_1\vdots\cdots\vdots v_r)\alpha\le -1,\; k=\overline{1,n_1},\; s=\overline{1,n_2}\}\,.$$

ОПТИМИЗАЦИЯ ПАРАМЕТРОВ В СИСТЕМЕ КЛАССИФИКАЦИИ СИГНАЛОВ С ПАРАЛЛЕЛЬНОЙ СТРУКТУРОЙ НЕЛИНЕЙНЫХ ФУНКЦИОНАЛЬНЫХ ПРЕОБРАЗОВАНИЙ

Рассмотрим сначала задачу синтеза системы нелинейных функциональных преобразований с параллельной структурой безотносительно к исследованию задачи классификации. Пусть известны значения входных сигналов $x(j)\in R^{\,m}$,

$j=\overline{1,n}$, и выходных сигналов $y(j)\in R^1$, $j=\overline{1,n}$. При этом будем предполагать, что в классе линейных математических моделей описать взаимосвязь между значениями $x(j)$ и $y(j)$ $\forall j=1,n$ невозможно, т.е.

$$y^{\mathrm{T}}Z\begin{pmatrix}X\\ \overline{J_n^{\mathrm{T}}}\end{pmatrix}y>0,\;\; y=(y(1),\dots,y(n))^{\mathrm{T}},\;\; X=(X(1)\vdots\dots\vdots X(n))\in R^{m\times n},\;\; J_n^{\mathrm{T}}=(1,\dots,1)\in R^n.$$

В работах [5, 6] описаны псевдоинверсные средства синтеза нелинейных трансформаций состояний входного вектора $x(j)$ для уменьшения отклонения значений выходных сигналов $\hat{y}(j)$ математической модели от заданных $y(j)$. В настоящей работе предлагается осуществить оптимальный выбор значений для параметров тех линейных комбинаций, для которых и выполняются соответствующие нелинейные трансформации.

В первом приближении указанные линейные комбинации могут представлять или проекцию вектора y на линейную оболочку вектор-строк матрицы $\begin{pmatrix}X\\ \overline{J_n^{\mathrm{T}}}\end{pmatrix}$, или основные факторные линейные комбинации этих вектор-строк. В качестве примера нелинейные преобразования выбраны в классе полиномов третьего порядка. Для таких нелинейных преобразований структура синтезируемой системы приведена на рис. 1. Здесь a_i, $i=\overline{1,s}$, — векторы, компоненты которых изначально могут принимать значения

$$a_{(1)}=\begin{pmatrix}X\\ \overline{J_n^{\mathrm{T}}}\end{pmatrix}^{+}y,\;\; a_{(2)}=u_1,\;\dots,\;a_{(s)}=u_{s-1},\;\;\begin{pmatrix}X\\ \overline{J_n^{\mathrm{T}}}\end{pmatrix}=\sum_{j=1}^{r}u_j v_j^{\mathrm{T}}\lambda_j.$$

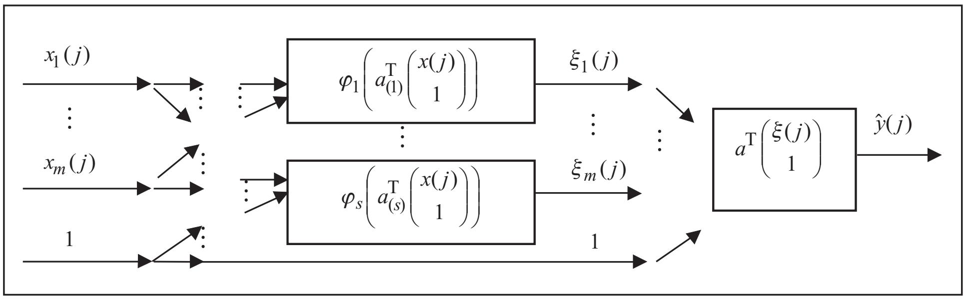

Рис. 1

Тогда, ограничиваясь полиномиальными функциями

$$\varphi_k\left(a_{(k)}^{\mathrm{T}}\begin{pmatrix}x(j)\\1\end{pmatrix}\right)=\left(a_{(k)}^{\mathrm{T}}\begin{pmatrix}x(j)\\1\end{pmatrix},\left(a_{(k)}^{\mathrm{T}}\begin{pmatrix}x(j)\\1\end{pmatrix}\right)^2,\left(a_{(k)}^{\mathrm{T}}\begin{pmatrix}x(j)\\1\end{pmatrix}\right)^3\right)\alpha_{(k)},\;\;\alpha_{(k)}\in R^3,$$

получаем идеальную ситуацию синтеза, когда $\hat{y}(j)=y(j)$ $\forall j=\overline{1,n}$ при условии существования решений системы алгебраических линейных уравнений

$$B\alpha=$$
$$=\begin{pmatrix} a_{(1)}^{\mathrm{T}}\begin{pmatrix}x(1)\\1\end{pmatrix} & \left(a_{(1)}^{\mathrm{T}}\begin{pmatrix}x(1)\\1\end{pmatrix}\right)^2 & \left(a_{(1)}^{\mathrm{T}}\begin{pmatrix}x(1)\\1\end{pmatrix}\right)^3 & \cdots & a_{(s)}^{\mathrm{T}}\begin{pmatrix}x(1)\\1\end{pmatrix} & \left(a_{(s)}^{\mathrm{T}}\begin{pmatrix}x(1)\\1\end{pmatrix}\right)^2 & \left(a_{(s)}^{\mathrm{T}}\begin{pmatrix}x(1)\\1\end{pmatrix}\right)^3 & 1\\ \cdots\;\vdots & \cdots & \vdots\;\cdots\;\vdots & \cdots & \vdots & \cdots & \vdots\;\cdots & \vdots\vdots\\ a_{(1)}^{\mathrm{T}}\begin{pmatrix}x(n)\\1\end{pmatrix} & \left(a_{(1)}^{\mathrm{T}}\begin{pmatrix}x(n)\\1\end{pmatrix}\right)^2 & \left(a_{(1)}^{\mathrm{T}}\begin{pmatrix}x(n)\\1\end{pmatrix}\right)^3 & \cdots & a_{(s)}^{\mathrm{T}}\begin{pmatrix}x(n)\\1\end{pmatrix} & \left(a_{(s)}^{\mathrm{T}}\begin{pmatrix}x(n)\\1\end{pmatrix}\right)^2 & \left(a_{(s)}^{\mathrm{T}}\begin{pmatrix}x(n)\\1\end{pmatrix}\right)^3 & 1 \end{pmatrix}\alpha=y, \quad (7)$$

где $\alpha=(\alpha_{(1)},\dots,\alpha_{(s)},\alpha_{(s+1)})$, $\alpha_{(j)}\in R^3$, $j=\overline{1,s}$, $\alpha_{(s+1)}\in R^1$.

Если (в силу выбора структуры) при естественном условии линейной независимости вектор-столбцов матрицы B необходимо минимизировать невязку, то последовательно можно определить

$$\min_{\alpha} \| y - B\alpha \|^2 = y^{\mathrm{T}} Z(B^{\mathrm{T}})\, y = y^{\mathrm{T}} (I_n - B^{+^{\mathrm{T}}} B^{\mathrm{T}})\, y = y^{\mathrm{T}} (I_n - B(B^{\mathrm{T}} B)^{-1} B^{\mathrm{T}})\, y$$

и значение $\operatorname{grad}_{a_{(j)}} y^{\mathrm{T}} Z(B^{\mathrm{T}})\, y$, где $\alpha = (B^{\mathrm{T}} B)^{-1} B^{\mathrm{T}} y$.

При этом дифференцирование проекционных матриц по параметру удобно выполнять согласно формулам

$$\frac{\partial}{\partial \beta} (y^{\mathrm{T}} Z(B^{\mathrm{T}}(\beta))\, y) = \frac{\partial}{\partial \beta} (y^{\mathrm{T}} (I_n - B(\beta)(B^{\mathrm{T}}(\beta) B(\beta))^{-1} B^{\mathrm{T}}(\beta))\, y) =$$
$$= -2\, y^{\mathrm{T}} Z(B^{\mathrm{T}}(\beta)) \frac{\partial B(\beta)}{\partial \beta} B^{+}(\beta)\, y, \quad B(\beta) = \begin{pmatrix} b_{(1)}^{\mathrm{T}}(\beta) \\ \vdots \\ b_{(n)}^{\mathrm{T}}(\beta) \end{pmatrix}, \quad \beta \in R^1, \tag{8}$$

$$\operatorname*{grad}_{\beta} y^{\mathrm{T}} Z(B^{\mathrm{T}}(\beta))\, y = -2 \begin{pmatrix} y^{\mathrm{T}} Z(B^{\mathrm{T}}(\beta)) \dfrac{\partial B(\beta)}{\partial \beta_1} B^{+}(\beta)\, y \\ \vdots \\ y^{\mathrm{T}} Z(B^{\mathrm{T}}(\beta)) \dfrac{\partial B(\beta)}{\partial \beta_l} B^{+}(\beta)\, y \end{pmatrix}, \quad \beta \in R^l, \tag{9}$$

$$\operatorname*{grad}_{B} y^{\mathrm{T}} Z(B^{\mathrm{T}})\, y = -2 Z(B^{\mathrm{T}})\, y y^{\mathrm{T}} B^{+^{\mathrm{T}}}. \tag{10}$$

Кратко вывод формулы (8) приводится ниже, а справедливость формул (9) и (10) очевидно следует из формулы (8). Итак, при $\beta \in R^1$ имеют место такие соотношения:

$$\frac{\partial}{\partial \beta} y^{\mathrm{T}} (B(\beta)(B^{\mathrm{T}}(\beta) B(\beta))^{-1} B^{\mathrm{T}}(\beta))\, y = y^{\mathrm{T}} \frac{\partial}{\partial \beta} (B(\beta)(B^{\mathrm{T}}(\beta) B(\beta))^{-1} B^{\mathrm{T}}(\beta))\, y =$$
$$= y^{\mathrm{T}} \left(\frac{\partial B(\beta)}{\partial \beta} (B^{\mathrm{T}}(\beta) B(\beta))^{-1} B^{\mathrm{T}}(\beta) - B(\beta)(B^{\mathrm{T}}(\beta) B(\beta))^{-1} \times \right.$$
$$\left. \times \frac{\partial}{\partial \beta} ((B^{\mathrm{T}}(\beta) B(\beta))(B^{\mathrm{T}}(\beta) B(\beta))^{-1} B^{\mathrm{T}}(\beta)) + B(\beta)(B^{\mathrm{T}}(\beta) B(\beta))^{-1} \frac{\partial B^{\mathrm{T}}(\beta)}{\partial \beta} \right) y =$$
$$= y^{\mathrm{T}} \left(\frac{\partial B(\beta)}{\partial \beta} B(\beta)^{+} - B(\beta)^{+^{\mathrm{T}}} \frac{\partial B^{\mathrm{T}}(\beta)}{\partial \beta} B(\beta)\, B(\beta)^{+} - \right.$$
$$\left. - B(\beta)^{+^{\mathrm{T}}} B^{\mathrm{T}}(\beta) \frac{\partial B^{\mathrm{T}}(\beta)}{\partial \beta} B(\beta)^{+} + B(\beta)^{+^{\mathrm{T}}} \frac{\partial B^{\mathrm{T}}(\beta)}{\partial \beta} \right) y =$$
$$= y^{\mathrm{T}} \left(Z(B^{\mathrm{T}}(\beta)) \frac{\partial B(\beta)}{\partial \beta} B(\beta)^{+} + B(\beta)^{+^{\mathrm{T}}} \frac{\partial B^{\mathrm{T}}(\beta)}{\partial \beta} Z(B^{\mathrm{T}}(\beta)) \right) y =$$
$$= 2\, y^{\mathrm{T}} Z(B^{\mathrm{T}}(\beta)) \frac{\partial B(\beta)}{\partial \beta} B(\beta)^{+}\, y.$$

Тогда вследствие зависимости элементов матрицы B в системе (7) от компонент вектора $a_{(i)} = (a_{i,1}, \ldots, a_{i,m+1})$ получим в соответствии с (8) следующие соотношения:

$$\operatorname*{grad}_{a_{(j)}} y^{\mathrm{T}} Z(B^{\mathrm{T}})\, y = -2 \begin{pmatrix} y^{\mathrm{T}} Z(B^{\mathrm{T}}) \dfrac{\partial B}{\partial a_{i,1}} B^{+}\, y \\ \vdots \\ y^{\mathrm{T}} Z(B^{\mathrm{T}}) \dfrac{\partial B}{\partial a_{i,m+1}} B^{+}\, y \end{pmatrix},$$

$$\frac{\partial B}{\partial a_{1,j}} = \begin{pmatrix} x_j(1) & 2a_{(1)}^{\mathrm{T}} \begin{pmatrix} x(1) \\ 1 \end{pmatrix} x_j(1) & 3 \left(a_{(1)}^{\mathrm{T}} \begin{pmatrix} x(1) \\ 1 \end{pmatrix} \right)^2 x_j(1) & 0 \\ \cdots\ \vdots & \cdots\ \vdots & \cdots & \vdots\ \cdots \\ x_j(n) & 2a_{(1)}^{\mathrm{T}} \begin{pmatrix} x(n) \\ 1 \end{pmatrix} x_j(n) & 3 \left(a_{(1)}^{\mathrm{T}} \begin{pmatrix} x(n) \\ 1 \end{pmatrix} \right)^2 x_j(n) & 0 \end{pmatrix} \quad \forall j = \overline{1, m},$$

$$\frac{\partial B}{\partial a_{s,j}} = \begin{pmatrix} x_j(1) & 2a_{(s)}^{\mathrm{T}} \begin{pmatrix} x(1) \\ 1 \end{pmatrix} x_j(1) & 3\left(a_{(s)}^{\mathrm{T}} \begin{pmatrix} x(1) \\ 1 \end{pmatrix} \right)^2 x_j(1) & 0 \\ \cdots \ \vdots & \cdots & \vdots \quad \cdots & \vdots \cdots \\ x_j(n) & 2a_{(s)}^{\mathrm{T}} \begin{pmatrix} x(n) \\ 1 \end{pmatrix} x_j(n) & 3\left(a_{(s)}^{\mathrm{T}} \begin{pmatrix} x(n) \\ 1 \end{pmatrix} \right)^2 x_j(n) & 0 \end{pmatrix} \quad \forall j = \overline{1, m},$$

$$\frac{\partial B}{\partial a_{1,m+1}} = \begin{pmatrix} 1 & 2a_{(1)}^{\mathrm{T}} \begin{pmatrix} x(1) \\ 1 \end{pmatrix} & 3\left(a_{(1)}^{\mathrm{T}} \begin{pmatrix} x(1) \\ 1 \end{pmatrix} \right)^2 & 0 \\ \cdots \vdots & \cdots & \vdots \quad \cdots & \vdots \cdots \\ 1 & 2a_{(1)}^{\mathrm{T}} \begin{pmatrix} x(n) \\ 1 \end{pmatrix} & 3\left(a_{(1)}^{\mathrm{T}} \begin{pmatrix} x(n) \\ 1 \end{pmatrix} \right)^2 & 0 \end{pmatrix},$$

$$\frac{\partial B}{\partial a_{s,m+1}} = \begin{pmatrix} 1 & 2a_{(s)}^{\mathrm{T}} \begin{pmatrix} x(1) \\ 1 \end{pmatrix} & 3\left(a_{(s)}^{\mathrm{T}} \begin{pmatrix} x(1) \\ 1 \end{pmatrix} \right)^2 & 0 \\ \cdots \vdots & \cdots & \vdots \quad \cdots & \vdots \cdots \\ 1 & 2a_{(s)}^{\mathrm{T}} \begin{pmatrix} x(n) \\ 1 \end{pmatrix} & 3\left(a_{(s)}^{\mathrm{T}} \begin{pmatrix} x(n) \\ 1 \end{pmatrix} \right)^2 & 0 \end{pmatrix}.$$

Используя градиентные вычислительные средства минимизации невязки системы $y^{\mathrm{T}} Z(B^{\mathrm{T}}(a_{(1)}, \ldots, a_{(s)}))\, y$, за счет выбора значений $a_{(i)}$, $i = \overline{1, s}$, осуществляем фактически оптимальный синтез системы с выбранной структурой параллельных нелинейных преобразований компонент вектора входных сигналов.

Описанную идею параллельного преобразования компонент вектора входных сигналов можно интерпретировать как постановку и решение задачи синтеза нелинейной системы полосной разделимости множеств при классификации сигналов в пространстве признаков. В том случае, если линейная полосная разделимость исследуемых множеств неосуществима, а значит, условие (5) не выполняется и при этом

$$\min_{y \in D} y^{\mathrm{T}} Z \begin{pmatrix} X \\ J_n^{\mathrm{T}} \end{pmatrix} y = y_*^{\mathrm{T}} \begin{pmatrix} X \\ J_n^{\mathrm{T}} \end{pmatrix} y_* > 0,$$

вектор y_* принимается за допустимый выход синтезируемой нелинейной системы и все параллельные преобразования осуществляются так, как описано ранее. Естественно, в новом пространстве входных сигналов задачу минимизации невязки за счет выбора следует переопределить, что, в свою очередь, улучшит качество синтезируемой системы.

ОПТИМИЗАЦИЯ ПАРАМЕТРОВ В СИСТЕМЕ КЛАССИФИКАЦИИ СИГНАЛОВ С ПОСЛЕДОВАТЕЛЬНОЙ СТРУКТУРОЙ НЕЛИНЕЙНЫХ ФУНКЦИОНАЛЬНЫХ ПРЕОБРАЗОВАНИЙ

В отличие от рассматриваемой ранее задачи формирования параллельной структуры нелинейных преобразований в системе классификации сигналов далее описывается некоторое математическое средство последовательных, или суперпозиционных, нелинейных преобразований компонент вектора входного сигнала и интерпретации задачи оптимального выбора параметров этих преобразований как задачи оптимального управления конкретной системой с дискретным аргументом и матричным состоянием системы для каждого значения ее аргумента. Такая интерпретация позволяет формализовать решение исследуемой задачи оптимизации параметров синтезируемой системы классификации сигналов развитыми современными численными методами теории оптимального управления с применением функций Гамильтона и системы уравнений сопряженных переменных в пространстве матричных состояний.

В схеме последовательных нелинейных преобразований, представленной на рис. 2, нелинейные преобразования $\varphi_N(\ldots)$ рассматриваются в качестве примера в таком виде:

$$\varphi_k(x(k,j)) = (1, a_{(k)}^{\mathrm{T}} x(k,j), (a_{(k)}^{\mathrm{T}} x(k,j))^2, (a_{(k)}^{\mathrm{T}} x(k,j))^3)\, u_{(k)},$$

$$x(k,j) = (x_1(k,j), \ldots, x_m(k,j)),\ j = \overline{1,n},\ k = \overline{1, N-1}.$$

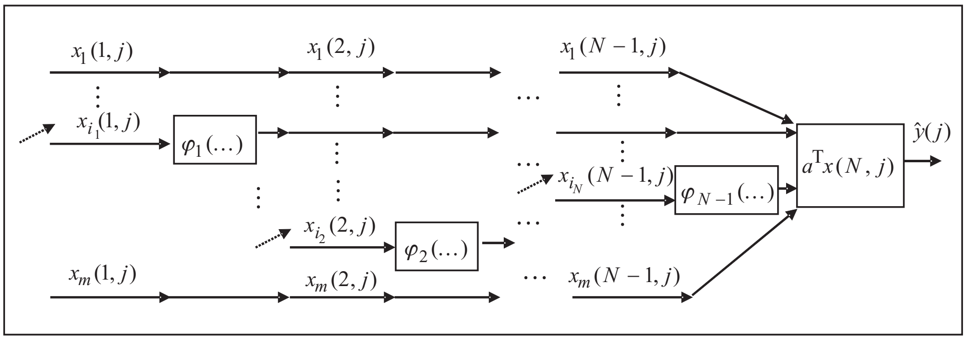

Рис. 2

Будем предполагать, что $a(k) \in R^m$, $i_k \in \{i : i = \overline{1,m}\}$ предварительно выбраны и изменяться не будут. Оптимизационные средства выбора этих параметров описаны в работах [6, 7]. Тогда взаимосвязь между матрицами $X(k)$, $X(k+1)$, где $X(j) = (x(j,1) \vdots \ldots \vdots x(j,n))\ \forall j = \overline{0,n}$, $X(0) = X = (x(1) \vdots \ldots \vdots x(n))$, имеет вид следующей математической модели системы управления с матричным состоянием:

$$X(k+1) = (I_m - e_{i_k} e_{i_k}^{\mathrm{T}}) X(k) + e_{i_k} \varphi^{\mathrm{T}}(X(k), u(k)),\ k = \overline{0, N-1}, \tag{11}$$

$$X(0) = X, \tag{12}$$

где

$$\varphi(X(k), u(k)) = \begin{pmatrix} 1 & a_{(k)}^{\mathrm{T}} x(k,1) & (a_{(k)}^{\mathrm{T}} x(k,1))^2 & (a_{(k)}^{\mathrm{T}} x(k,j))^3 \\ \cdots \vdots & \cdots \quad \vdots & \cdots \quad \vdots & \cdots \\ 1 & a_{(k)}^{\mathrm{T}} x(k,n) & (a_{(k)}^{\mathrm{T}} x(k,n))^2 & (a_{(k)}^{\mathrm{T}} x(k,j))^3 \end{pmatrix} u_k,$$

e_{i_k} — i_k-й единичный орт в R^m, т.е. $e_{i_k} = (0 \ldots 0\ 1\ 0 \ldots 0)^{\mathrm{T}}$.

Функционал качества этой системы

$$J(u) = y^{\mathrm{T}} Z((X(N)))\, y \tag{13}$$

означает невязку между синтезированным выходом и заданным целевым выходом системы, т.е. $J(u) = \sum_{j=1}^{n} |\hat{y}(j) - y(j)|^2$.

Решение задачи оптимального управления по поиску параметров $u(k)$ с целью минимизации функционала (13) для системы (11), (12) можно осуществить известными вычислительными средствами с использованием $\mathrm{grad}_u J(u)$. В свою очередь, определение значения $\mathrm{grad}_u J(u)$ сводится, как известно [8], к дифференцированию функции Гамильтона, которая в данном случае имеет вид

$$\mathbf{H}(\Psi(k+1), X(k), u(k), k) =$$
$$= \mathrm{tr}\, \{\Psi^{\mathrm{T}}(k+1)[(I_m - e_{i_k} e_{i_k}^{\mathrm{T}}) X(k) + e_{i_k} \varphi^{\mathrm{T}}(X(k), u(k))]\},$$
$$\Psi(k) = \mathrm{grad}_{X(k)}\, \mathbf{H}(\Psi(k+1), X(k), u(k), k) =$$
$$= (I_m - e_{i_k} e_{i_k}^{\mathrm{T}}) \Psi(k+1) + \mathrm{grad}_{X(k)}\, \varphi^{\mathrm{T}}(X(k), u(k))\, \Psi^{\mathrm{T}}(k+1)\, e_{i_k}, \tag{14}$$

где $\Psi(k) \in R^{m \times n}$ — матричное состояние сопряженной системы. При этом

$\mathrm{grad}_{X(k)} f(X(k))$ представляет собой матрицу, в которой размерность совпадает с размерностью матрицы $X(k)$ и каждый ее элемент является частной производной функции $f(X(k))$ по соответствующему элементу матрицы $X(k)$.

Система (14) рассматривается при краевом условии

$$\Psi(N) = -\mathrm{grad}_{X(k)} y^{\mathrm{T}} Z((X(N)))\, y = 2(Z(X^{\mathrm{T}}(N))\, y y^{\mathrm{T}} X^{+}(N))^{\mathrm{T}} =$$

$$= 2X^{+^{\mathrm{T}}}(N)\, y\, y^{\mathrm{T}} Z(X(N)),$$

в котором учитывается общность положения линейной независимости вектор-строк матрицы $X(N)$. Тогда для вычисления градиента от функционала $J(u)$ по векторам $u(k)$ $\forall k = 0, N-1$ получим соотношение

$$\mathrm{grad}_{u(k)} J(k) = \mathrm{grad}_{u(k)} \mathbf{H}(\Psi(k+1), X(k), u(k), k) =$$

$$= -\mathrm{grad}_{u(k)} \varphi^{\mathrm{T}}(X(k), u(k))\, \Psi^{\mathrm{T}}(k+1) e_{i_k} =$$

$$= -\begin{pmatrix} 1 & \ldots & 1 \\ a_{(k)}^{\mathrm{T}} x(k,1) & \cdots & a_{(k)}^{\mathrm{T}} x(k,n) \\ (a_{(k)}^{\mathrm{T}} x(k,1))^2 \vdots & \ldots & \vdots (a_{(k)}^{\mathrm{T}} x(k,n))^2 \\ (a_{(k)}^{\mathrm{T}} x(k,j))^3 & \cdots & (a_{(k)}^{\mathrm{T}} x(k,j))^3 \end{pmatrix} \Psi^{\mathrm{T}}(k+1)\, e_{i_k}.$$

ЗАКЛЮЧЕНИЕ

Описаны применения условий линейной и нелинейной отделимости и разделимости множества точек в конечномерном пространстве признаков и обобщение данных результатов на бесконечномерное пространство. В результате проведенных исследований выполнена реализация решения задачи выбора структуры нелинейных функциональных преобразований за счет последовательных и параллельных структурных элементов этих преобразований. Предложенные методы и алгоритмы оптимального синтеза систем с такой структурой нелинейных функциональных преобразований в задачах классификации сигналов при распознавании конкретной информации являются новыми и перспективными.

По результатам настоящей работы можно констатировать, что применение средств сингулярного представления матриц, функций Гамильтона и уравнений для сопряженных матричных переменных в возникающей при этом задаче оптимизации позволяет сокращать количество переменных и вычислительных операций при решении конкретных задач синтеза систем классификации информации.

СПИСОК ЛИТЕРАТУРЫ

1. Кириченко Н.Ф., Кривонос Ю.Г., Лепеха Н.П. Синтез систем нейрофункциональных преобразователей в решении задач классификации // Кибернетика и системный анализ. — 2007. — № 3. — С. 47–57.
2. Кириченко Н.Ф., Донченко В.С., Сербаев Д.П. Нелинейные рекурсивные регрессионные преобразователи: динамические системы и оптимизация // Там же. — 2005. — № 3. — С. 58–68.
3. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений / Пер. с англ. — М.: Мир, 1980. — 279 с.
4. Алберт А. Регрессия, псевдоинверсия, рекуррентное оценивание / Пер. с англ. — М.: Наука, 1977. — 224 с.
5. Кириченко Н.Ф. Аналитическое представление возмущений псевдообратных матриц // Кибернетика и системный анализ. — 1997. — № 2. — С. 98–107.
6. Кириченко Н.Ф., Крак Ю.В., Полищук А.А. Псевдообратные и проекционные матрицы в задачах синтеза функциональных преобразователей // Там же. — 2004. — № 3. — С. 116–129.
7. Кириченко Н.Ф., Лепеха Н.П. Применение псевдообратных и проекционных матриц к исследованию задач управления, наблюдения и идентификации // Там же. — 2002. — № 4. — С. 107–124.
8. Беклемишев Д.В. Дополнительные главы линейной алгебры. — М.: Наука, 1983. — 336 с.

Поступила 04.02.2010