

# Когнитивные технологии в проблеме снижения размерности описания геометрических объектов

А.В. Бернштейн, А.П. Кулешов

**Аннотация.** Одной из важнейших составляющих описания технического объекта является представление его геометрических компонентов (кривых и поверхностей), применяемых в САД-системах, компьютерной графике и других приложениях. При решении различных расчетных и оптимизационных задач, использующих геометрии объектов в качестве входных данных, необходимо решать задачу снижения размерности модели объекта, в том числе размерности геометрических составляющих. Когнитивные технологии являются основным инструментом решения таких задач. В статье подробно описана постановка задачи снижения размерности, приведен краткий обзор основных используемых подходов и методов и предложен общий подход к построению процедур снижения размерности описания геометрических объектов.

## Введение

Детальные описания геометрических объектов (кривых, поверхностей) или их компонентов в общем случае задаются набором 2D или 3D координат точек поверхности объекта, лежащих в выбранных узлах объекта. Другие точки объекта восстанавливаются, как правило, с использованием сплайнов - например, кривых Безье (Bezier curves, сплайнов Бежье), поверхностей Безье, B-сплайнов (base-splines, базовых сплайнов - например, рациональных B-сплайнов, задаваемых на неравномерной сетке – Non-Uniform Rational B-splines, NURBS) и др. Такого рода детальные описания кривых и поверхностей используются в САД-системах, компьютерной графике и других приложениях. Тем самым, детальные описания объекта состоят из десятков тысяч чисел, каждое из которых, рассматриваемое изолированно, не несет смысловой нагрузки.

В процессе проектирования и создания сложного технического объекта рассматриваются и сравниваются различные технические решения, касающиеся структуры объекта, ме-

ханизмов его функционирования, выбора параметров и других элементов объекта, в том числе геометрической формы объекта. При этом решаются различные расчетные и оптимизационные задачи, например:

- расчет характеристик объектов как значений различных функций  $F(X, Y)$ , зависящих от описания объекта  $X$  (в том числе, его геометрии) и условий функционирования  $Y$ ;
- оптимизация объекта, то есть максимизация или минимизация некоторых функций от описания объекта при ограничениях на значения других функций;
- генерация объектов, включая генерацию геометрий, и др.

Если в задачах расчета характеристик объектов возможно использование детальных описаний геометрий объектов в качестве входных данных (например, CFD-коды в задачах аэродинамического проектирования самолета используют описания поверхности самолета, получаемые из САД-систем), то решение задач оптимизации функций, зависящих от тысяч аргументов, не представляется возможным. При

построении быстрых (суррогатных) моделей, основанных на данных, решаются задачи аппроксимации функций, в которых размерность аргументов является существенным ограничивающим фактором [1–4].

Поэтому одной из важнейших задач, которую необходимо решать в процессе проектирования, является задача снижения размерности описания объекта, в том числе размерности геометрического описания [5].

## 1. Математическая постановка задачи снижения размерности

Существуют различные, по сути эквивалентные, постановки задач снижения размерности. В данной статье мы будем придерживаться определений, введенных в работе [1], основные положения которых перечислены ниже.

1. Пусть  $V = \{b\}$  есть множество геометрических объектов рассматриваемого класса. Для каждого объекта  $b \in V$  имеется его детальное описание  $X = X(b)$ , размерность  $N$  которого может достигать тысяч чисел. Рассмотрим множество  $X = \{X(b), b \in V\}$   $N$ -мерного евклидова пространства  $R^N$ , состоящего из детальных описаний всех объектов рассматриваемого класса.

Фундаментальным предположением является предположение, что это множество  $X$  лежит, по крайней мере, приближенно, на многообразии (в общем случае, нелинейном), размерность  $n$  которого существенно меньше  $N$ . Целью снижения размерности является нахождение представления для этого многообразия (параметризации, координатной системы), которое позволит «спроектировать» множество  $X$  на него и получить компактное описание данных  $X(b)$  в пространстве меньшей размерности.

2. Наряду с детальным описанием, можно рассмотреть различные обобщенные описания  $h = H(b)$  объекта  $b$ , состоящие из набора параметров (интегральных характеристик) объекта  $b$ . В инженерных приложениях в качестве таких характеристик обычно выбирают величины, которыми оперирует инженер в процессе разработки (которые, например, являются предметом принятия решений при выборе варианта геометрии объекта).

Зафиксируем некоторый набор параметров объекта  $H_{mod}(b)$ , определяющий отображение

$$H_{mod}(b): V \rightarrow G_{mod}, \quad (1)$$

где множество

$$G_{mod} = \{H_{mod}(b), b \in V\} \quad (2)$$

- образ множества  $V$  при отображении  $H_{mod}$  и является фактор-пространством множества объектов  $V$ , определяемое отображением  $H_{mod}$ .

Очевидно, что в общем случае существует целое множество объектов:

$$V_{mod}(h) = \{b \in V: H_{mod}(b) = h\}$$

с одним и тем же набором параметров  $h$ , и отображение  $H_{mod}$  определяет разбиение пространства  $V$  на непересекающиеся подмножества  $V_{mod}(h)$ ,  $h \in G_{mod}$ .

3. Для каждого объекта  $b \in V$  выберем (определим некоторым образом) **единственный** объект

$$b_{mod} = b_{mod}(b) \in V_{mod}(H_{mod}(b)),$$

называемый **модельным объектом**, соответствующим исходному объекту  $b$ , и обозначим

$$V_{mod} = \{b_{mod}(b), b \in V\} \subset V$$

множество всех модельных объектов. По построению, между множествами  $V_{mod}$  и  $G_{mod}$  существует взаимно-однозначное соответствие, определяемое прямым (1) и обратным отображениями:

$$H_{mod}^{-1}: G_{mod} \rightarrow V_{mod}, \quad (3)$$

с помощью которых модельный объект определяется как

$$b_{mod}(b) = H_{mod}^{-1}(H_{mod}(b)). \quad (4)$$

Модельный объект (4), построенный по объекту  $b$  с помощью пары отображений  $\{H_{mod}, H_{mod}^{-1}\}$ , будем называть также **модельным представлением** или **модельным аналогом** объекта  $b$ .

4. По построению, множество модельных объектов  $V_{mod}$  является многообразием в пространстве объектов  $V$ , параметризованным с помощью отображения  $H_{mod}^{-1}$ , определенного на фактор-пространстве  $G_{mod}$ . Обратное отображение  $H_{mod}^{-1}$  определяет также «алгоритм восстановления», позволяющий для каждого исходного объекта  $b \in V$  строить детальное описание

$$X_{\text{mod}} = X(b_{\text{mod}}(b))$$

соответствующего модельного объекта  $b_{\text{mod}}(b)$  (4).

5. Если вектор  $H_{\text{mod}}(b)$  имеет небольшую (по сравнению с размерностью  $N$  детального описания  $X(b)$ ) размерность  $n$ , и если все объекты  $b \in B_{\text{mod}}(h)$ , имеющие одно и то же значение вектора параметров модели  $h$ , имеют близкие детальные описания, то есть если

$$X(b') \approx X(H_{\text{mod}}^{-1}(h)),$$

для всех  $h \in G_{\text{mod}}$  и  $b' \in B_{\text{mod}}(h)$ , то пара отображений  $\{H_{\text{mod}}, H_{\text{mod}}^{-1}\}$  определяет процедуру снижения размерности (сжатия) описания объекта:

- отображение  $H_{\text{mod}}$  определяет процедуру снижения размерности (сжатия) детального описания объекта:

$$X(b) \rightarrow H_{\text{mod}}(b), \quad (5)$$

и величину  $H_{\text{mod}}(b)$  можно считать «сжатым» описанием объекта;

- отображение  $H_{\text{mod}}^{-1}$  определяет процедуру восстановления детального описания объекта по его сжатому описанию:

$$H_{\text{mod}}(b) \rightarrow X(H_{\text{mod}}^{-1}(H_{\text{mod}}(b))), \quad (6)$$

и погрешность приближенного равенства

$$X(b) \approx X(H_{\text{mod}}^{-1}(H_{\text{mod}}(b))) \quad (7)$$

определяет точность процедуры сжатия.

## 2. Геометрическая интерпретация задачи снижения размерности

Геометрическая интерпретация сформулированной задачи снижения размерности заключается в следующем. Рассмотрим множество  $X = \{X(b), b \in B\}$   $N$ -мерного евклидова пространства  $R^N$ , состоящее из детальных описаний всех объектов рассматриваемого класса.

Фундаментальным предположением является предположение о том, что множество  $X$  лежит, по крайней мере, приближенно, на многообразии  $Y$ , являющимся  $n$ -мерным параметрическим многообразием в пространстве  $R^N$  (в общем случае нелинейном), параметризованным вектором  $h \in R^n$ . Другими словами,

$$Y = \{y = H(h) \in R^N: h \in G \subset R^n\}.$$

В терминах (обозначениях) предыдущего раздела, множество  $Y$  совпадает с подмножеством

$$X_{\text{mod}} = \{X_{\text{mod}} = X(b_{\text{mod}}), b_{\text{mod}} \in B_{\text{mod}}\} \subset R^N, \quad (8)$$

состоящим из детальных описаний модельных объектов и параметризованным вектором  $h = H_{\text{mod}}(b) \in G_{\text{mod}}$  размерности  $n$  (обозначения (1) - (4)).

Задача снижения размерности включает в себя определение размерности  $n$  многообразия  $Y$ , которое с заданной точностью аппроксимирует исходное множество  $X$ , и нахождение представления для этого многообразия (координатную систему), которое позволит получить компактное описание данных в пространстве меньшей размерности.

Как правило, класс рассматриваемых объектов  $B = \{b\}$  не имеет точного описания и определяется конечным множеством его «представителей» (прототипов), задаваемых множеством их детальных описаний

$$X_T = \{X(b_t), t = 1, 2, \dots, T\}. \quad (9)$$

Обозначив  $X_t = X(b_t)$ ,  $t = 1, 2, \dots, T$ , и используя стандартное обозначение

$$X_T = \{X_t, t = 1, 2, \dots, T\} \quad (10)$$

для выборки данных (9), переформулируем задачу снижения размерности в том виде, в котором она формулируется в большей части публикаций по анализу и обработке данных. Задача снижения размерности включает в себя:

- определение размерности  $n < N$  многообразия, аппроксимирующего с заданной точностью выборку данных (10);

- построение множества (пространства)  $G \subset R^n$ , являющегося фактор-пространством, параметризующим аппроксимирующее многообразие;

- построение снижающего размерность отображения

$$F: X \rightarrow G$$

результат  $h = F(X)$  применения которого к вектору  $X \in X$  будем называть представлением исходного вектора  $X$  в виде сжатого вектора меньшей размерности;

- построение гладкого восстанавливающего отображения

$$F: G \rightarrow Y \subset X,$$

результат  $Y = F(h)$  применения которого к вектору  $h \in G$  есть  $N$ -мерный вектор, восстановленный по сжатому вектору.

Качество процедуры снижения размерности оценивается величиной ошибки восстановления, определяемой некоторой мерой близости между исходной выборкой  $\mathbf{X}_T$  и «восстановленным» множеством  $\mathbf{Y}_T = \{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_T\}$ , где  $\mathbf{Y}_t = F(F(\mathbf{X}_t))$ ,  $t = 1, 2, \dots, T$ . Примерами таких мер близости могут быть величины:

$$\delta_1(n, \rho) = \frac{1}{T} \sum_{t=1}^T \rho(X_t, Y_t), \quad (11)$$

или

$$\delta_2(n, \rho) = \sqrt{\frac{1}{T} \sum_{t=1}^T (\rho(X_t, Y_t))^2}, \quad (12)$$

где в качестве расстояния  $\rho(\mathbf{X}, \mathbf{Y})$  между векторами  $\mathbf{X} = (X_1, X_2, \dots, X_N)$  и  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)$  могут быть метрики:

$$\rho_\infty(\mathbf{X}, \mathbf{Y}) = \max\{|X_j - Y_j|, j = 1, 2, \dots, N\}, \quad (13)$$

$$\rho_1(\mathbf{X}, \mathbf{Y}) = \frac{1}{N} \sum_{j=1}^N |X_j - Y_j|, \quad (14)$$

$$\rho_2(\mathbf{X}, \mathbf{Y}) = \sqrt{\frac{1}{N} \sum_{j=1}^N |X_j - Y_j|^2}. \quad (15)$$

Заметим, что значения мер близости (11) или (12) (ошибок) зависят от выбранной размерности  $n$  сжатых векторов.

Малость ошибок (11) или (12) означает, что множество  $\mathbf{Y} = F(G)$  приближенно содержит все точки выборки (10). Целью процедуры снижения размерности является минимизация размерности  $n = n(D_0)$  при условии выполнения заданных ограничений

$$\delta(n, \rho) \leq D_0 \quad (16)$$

на величину ошибки.

В терминах (обозначениях) предыдущего раздела, введенные понятия означают:

- условие (7), являющееся аналогом условия (16), означает, что множество  $\mathbf{X}$  лежит в окрестности  $n$ -мерного многообразия  $\mathbf{X}_{\text{mod}}$  (8), где топология в множестве подмножеств  $N$ -мерного евклидова пространства  $\mathbb{R}^N$  определяется метрикой Хаусдорфа  $\rho(\mathbf{X}, \mathbf{X}_{\text{mod}})$ , записываемой в нашем случае как:

$$\rho(\mathbf{X}, \mathbf{X}_{\text{mod}}) = \sup\{\|\mathbf{X}(b) - \mathbf{X}(H_{\text{mod}}^{-1}(H_{\text{mod}}(b)))\|, b \in B\}$$

или некоторым ее аналогом, здесь  $\|\cdot\|$  - некоторая норма в  $\mathbb{R}^N$  (например,  $l_1$ ,  $l_2$  или  $l_\infty$ ).

- процедура сжатия (5) задает проектирование точки  $\mathbf{X}(b)$  из множества  $\mathbf{X}$  в многообразии  $\mathbf{X}_{\text{mod}}$  (определяет значение вектора  $h_{\text{mod}} = H_{\text{mod}}(b) \in G_{\text{mod}}$ , соответствующего проекции  $\mathbf{X}(H_{\text{mod}}^{-1}(h_{\text{mod}})) \in \mathbf{X}_{\text{mod}}$ ).

- процедура восстановления (6) определяет способ построения детального описания модельного объекта  $\mathbf{X}(H_{\text{mod}}^{-1}(h)) \in \mathbf{X}_{\text{mod}}$ , лежащего на многообразии  $\mathbf{X}_{\text{mod}}$ , по значению параметра  $h \in G_{\text{mod}}$ .

Априори дискретная выборка имеет нулевую размерность, и в совокупности с известным феноменом «пустого пространства», заключающимся в существенной разреженности многомерных пространств, это существенно затрудняет определение размерности аппроксимирующего многообразия. Для этих целей введено понятие «внутренней размерности» (Intrinsic Dimension, ID) данных, и определение ID по выборке (10) является одной из центральных задач снижения размерности, поскольку использование точного значения ID при применении какого-нибудь алгоритма снижения размерности позволяет избежать эффектов «переобучения» или «недообучения».

Согласно классификации, предложенной в [6], существует два подхода для оценки ID. Первый (глобальный) подход позволяет оценить размерность множества данных в целом. Второй (локальный) подход оценивает ID на основе информации, содержащейся в выборке, и не использует проекцию данных на пространство низкой размерности.

С использованием метода проекций или других топологических методов может быть вычислена так называемая топологическая размерность данных [7]. Основная идея различных методов проекций (линейных и нелинейных), которые используются в глобальных методах, заключается в том, чтобы найти наилучшее подпространство для проекции на него данных, где термин «наилучшее» обычно понимают в смысле минимизации ошибки восстановления. Топологические методы, являющиеся локальными, нацелены на оценку топологической

размерности многообразия, в котором лежат данные.

В качестве оценки внутренней размерности можно использовать фрактальную размерность [8, 9], равную величине.

$$D_B = \lim_{r \rightarrow 0} \frac{\ln(v(r))}{\ln(r)},$$

где  $v(r)$  есть число кубиков размера  $r$ , необходимых для покрытия множества (10). К сожалению, величину  $D_B$  легко подсчитать только для множеств, размерность которых невелика. Поэтому вместо фрактальной размерности можно использовать корреляционную размерность [10], являющейся оценкой снизу для фрактальной размерности (в реальных приложениях разница между ними незначительна). Определим корреляционный интеграл  $C_m(r)$  как

$$C_m(r) = \lim_{T \rightarrow \infty} \frac{2}{T(T-1)} \sum_{1 \leq i < j \leq T} I\{\|X(b_j) - X(b_i)\| \leq r\},$$

где  $I(A)$  - индикатор события  $A$ . Тогда корреляционная размерность вычисляется как

$$D_C = \lim_{r \rightarrow 0} \frac{\ln(C_m(r))}{\ln(r)}.$$

В работе [11] предложен простой алгоритм для оценки корреляционной размерности многомерного множества (этот метод принадлежит классу глобальных методов).

### 3. Используемые методы и подходы в задачах снижения размерности

Универсальные подходы, обычно рассматриваемые для решения задачи снижения размерности, могут быть условно разбиты на два класса:

- подходы, основанные на параметризации описания геометрии,
- подходы, основанные на данных.

**Параметризация описания объекта** заключается в попытке описать объект в целом или отдельные его компоненты при помощи аналитических формул, зависящих от небольшого числа параметров. Значительная часть усилий при параметризации описания объекта лежит в предметной области, хотя при этом возникают и решаются отдельные математические задачи. Однако во многих случаях при использовании такого подхода не удается учесть

всего многообразия описаний объектов рассматриваемого класса. Кроме того, такой подход основан на существенном использовании предметной области объекта, требует достаточно подробной идентификации класса рассматриваемых объектов и требует участия в этом процессе высококвалифицированных специалистов. Это затрудняет создание универсального инструментария для параметризации описания объекта.

**Подходы, основанные на данных**, заключаются в применении математических алгоритмов снижения размерности (оценивания внутренней размерности данных, сжатия и восстановления данных). При этом подходе процедуры снижения размерности строятся на основании имеющейся выборки (10), состоящей из конечного множества детальных описаний некоторых типичных объектов. Естественно, предполагается, что имеющаяся выборка достаточно «репрезентативна», иначе построенные процедуры снижения размерности могут иметь низкое качество при их применении к объектам, лежащим «далеко» от выборки.

Наиболее популярным линейным методом снижения размерности является Анализ Главных Компонент (Principal Component Analysis, PCA) [12, 13]. Одной из причин этой популярности является простота и в то же время высокая эффективность метода. Метод сводится к построению новой ортогональной системы координат, в которой осями являются собственные векторы выборочной ковариационной матрицы, называемые главными компонентами. Преобразование к главным компонентам задает систему отсчета, в которой выборка имеет нулевое среднее и диагональную ковариационную матрицу  $\Lambda$ , содержащую на диагонали собственные числа выборочной ковариационной матрицы. Нумерация главных компонент строится по уменьшению собственных чисел, то есть, в порядке убывания разброса данных вдоль главных компонент. После построения главных компонент, можно спроектировать выборку на подпространство, натянутое на несколько первых главных компонент, и получить достаточно хорошую аппроксимацию выборки, пренебрегая теми переменными (в построенной координатной системе), у кото-

рых дисперсия мала. Ключевое свойство PCA в том, что этот метод позволяет получить снижение размерности с помощью линейного отображения, обладающего различными свойствами оптимальности.

Геометрически, гиперплоскость, натянутая на несколько первых главных компонент, является регрессионной гиперплоскостью, минимизирующей ортогональные расстояния до точек выборки. Таким образом, PCA является симметричным регрессионным методом, и пространство главных компонент как раз и является линейным многообразием, проекция на которое определяет процедуру снижения размерности.

Недостаток PCA заключается в том, что с помощью этого метода можно найти только наилучшее линейное аппроксимирующее многообразие. Если данные лежат на нелинейном многообразии, то оценка ID на основе PCA дает завышенное значение внутренней размерности, и эффективность метода значительно снижается.

Универсальным методом снижения размерности является использование Искусственных Нейронных Сетей (Artificial Neural Networks, ANN) «с перетяжкой», называемых также репликативными нейронными сетями [14-18].

Топология репликативных нейронных сетей представлена на Рис. 1.

На вход сети подаются векторы  $\{X_1, X_2, \dots, X_T\}$ , и, при заданном числе  $n$  нейронов в среднем слое, коэффициенты нейронной сети подбираются таким образом, чтобы обеспечить максимальную близость (например, в смысле метрик (11) или (12)) между входными векторами и выходными (восстановленными) векторами  $\{Y_1, Y_2, \dots, Y_T\}$ . Варьируя числом нейронов в среднем слое, можно выбрать минимальное значение  $n$ , при котором удовлетворяются заданные требования по точности. После построения такой сети, левая часть сети (от входного до среднего слоя) определяет процедуру сжатия, а правая часть сети (от среднего слоя до выходного слоя) определяет процедуру восстановления.

Заметим, что процедура снижения размерности, основанная на PCA, может быть рассмотрена как частный случай ANN с перетяжкой при выборе линейных функций активации нейронов [19].

Другой метод снижения размерности основан на ядерном PCA (Kernel PCA) [20], в котором для вычисления скалярных произведений

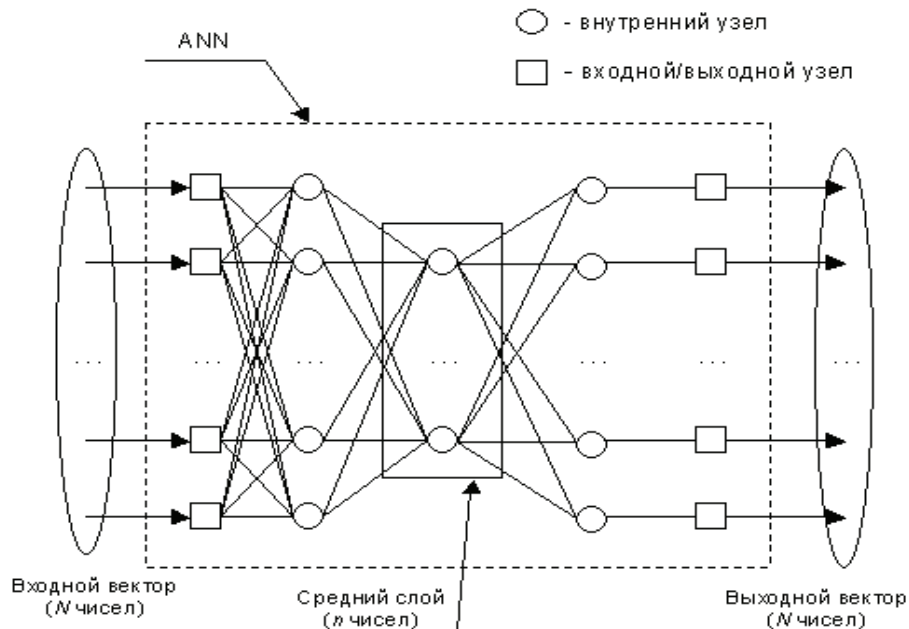


Рис. 1. Стандартная репликативная нейронная сеть

при построении выборочной ковариационной матрицы используются ядерные функции в  $X$ . Нелинейная природа этих функций означает, что получившиеся главные компоненты, по сути, будут нелинейными. Воспроизводящие ядра (полиномиальное, гауссовское, сигмоидальное), обычно используемые в ядерных функциях, используются также в методах [21, 22], основанных на системах поддерживающих векторов (Support Vector Machine, SVM).

Метод главных компонент выбирает проекции точек выборки в соответствии с критерием максимума дисперсии каждой из проекций при условии их взаимной ортогональности. Однако можно искать такие проекции, которые удовлетворяют другим свойствам. Это является основной идеей метода Projection Pursuit (PP) [23–25]. PP представляет собой метод обучения «без учителя», с помощью которого можно выделить «интересные» линейные ортогональные проекции многомерных данных в пространстве низкой размерности за счет оптимизации некоторого критерия, называемого **Projection Index (PI)**.

Отметим также методы факторного анализа [26 – 29], PP-регрессии [30], главных кривых (Principal curves) [31] и другие, используемые для снижения размерности.

#### **4. Решение задачи снижения размерности геометрического описания объекта**

В приложениях, как правило, используются геометрические объекты, являющиеся либо одномерными кривыми в двумерном пространстве, либо двумерными поверхностями в трехмерном пространстве. Так как предлагаемый подход одинаково применим к этим обоим типам геометрических объектов, то для простоты изложения и сокращения чисто технических деталей, мы изложим этот подход применительно к одномерным кривым в двумерном пространстве.

Одним из важных примеров геометрических объектов, описываемых одномерными кривыми в двумерном пространстве, являются профили (Рис. 2), которыми задаются плоские сечения многих инженерных объектов (лопастей винта, крыльев самолета и т.п.). Возможность снижения размерности описания профилей во многом

определяет возможности моделирования и оптимизации указанных объектов.

По методам описания профили делятся на две группы:

- профили, контуры которых имеют аналитические описания,
- профили, контуры которых заданы дискретным массивом координат.

Например, в самолетостроении контуры профилей первой группы могут быть получены по заданным аэродинамическим характеристикам в результате решения задачи обтекания кругового цилиндра с использованием конформных отображений (профили Жуковского, Кармана-Трефтца, Мизеса, Карафоли) [32]. Так как уравнения контуров этих профилей достаточно сложны, используются аппроксимации в виде функций простого вида (например, степенных). Известные семейства аэродинамических профилей NACA, предложенные National Advisory Committee for Aeronautics (NACA) [33, 34] дают примеры таких упрощенных описаний.

Универсальным способом детального описания профилей, как, впрочем, и других геометрических объектов, являются описания второй группы, задаваемые набором координат выбранных точек профиля (Рис. 2). При описании аэродинамических профилей крыла число  $N$  имеет порядок  $50 \div 100$ .

Однако представление геометрического объекта  $N$ -мерным вектором, состоящим из координат некоторой сетки (в общем случае, неравномерной) на поверхности объекта имеет ряд очевидных недостатков:

- один и тот же геометрический объект имеет разные цифровые представления в зависимости от выбранных точек контура;
- элементы выборки (10), состоящие из детальных описаний геометрических объектов, могут иметь разную размерность, что не позволяет применять известные методы снижения размерности, в которых предполагается одинаковая размерность элементов выборки;
- формальные математические процедуры снижения размерности, применяемые к наборам данных, формально не используют информацию о гладкости и других свойствах объекта (например, «аэродинамической форме» профилей).

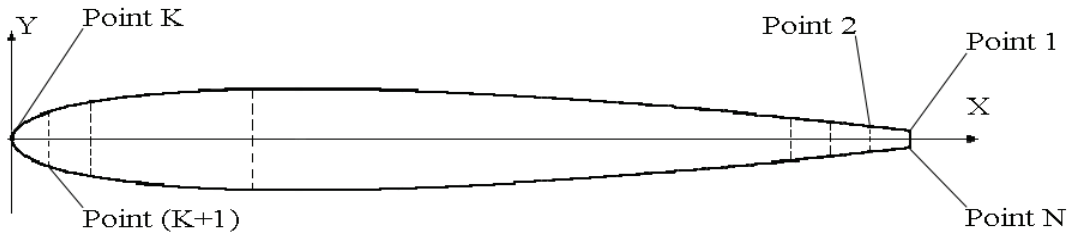


Рис. 2. Форма профиля

В статье предлагается общий подход к построению процедур снижения размерности описания геометрических объектов, иллюстрируемый на примере одномерных геометрических объектов в двумерном пространстве, в частности, профилей. Этот подход был успешно применен в процедурах снижения размерности описания профилей при решении различных задач аэродинамического проектирования [35-39].

Общая процедура снижения размерности состоит из нескольких шагов:

**Шаг 1.** Построение унифицированного описания геометрического объекта, позволяющего строить однородную выборку на базе имеющихся разнородных описаний различных объектов выборки.

**Шаг 2.** Построение наилучшей линейной процедуры снижения размерности до уровня размерности  $n_{lin} = n_{lin}(D_0)$ , обеспечивающего заданный уровень  $D_0$  (16) величины ошибки восстановления (7).

**Шаг 3.** Итеративное построение наилучшей (в общем случае, нелинейной) процедуры снижения размерности до финального уровня размерности  $n_{fin} = n_{fin}(D_0)$ , обеспечивающего заданный уровень величины ошибки восстановления.

Далее мы коротко опишем эти шаги.

**Построение унифицированного описания геометрического объекта.** Кривая на плоскости  $(X, Y)$  задается парой уравнений  $\{x(\tau), y(\tau)\}$  для координат точек кривой, параметризованных некоторым одномерным параметром  $\tau$ . В каждом «неэкзотическом» конкретном случае при помощи процедур репараметризации и преобразования координат, описание плоской кривой удастся задать в виде заданной на отрезке  $[-1, 1]$  функции  $U(t)$ , зависящей от одномерного параметра  $t$ .

Например, пусть для профилей, изображенных на Рис. 2, координата  $X$  изменяется от 0 до некоторого значения  $x_{max}$  (в аэродинамических приложениях эта величина называется длиной хорды профиля). Пусть параметр  $\tau$  параметризует движение вдоль контура профиля, которое начинается в крайней правой верхней точке (точка Point 1 на Рис. 2), проходя через «носик профиля» (точка Point K на Рис. 2) и заканчивается в крайней правой нижней точке (точка Point N на Рис. 2). Тем самым, координата  $X$  сначала изменяется (убывает) от значения  $x_{max}$  до нуля, а затем изменяется (возрастает) от нуля до значения  $x_{max}$ .

Без ограничения общности параметризацию с помощью параметра  $\tau$  можно считать равномерной вдоль абсциссы. Формально это соответствует изменению параметра  $\tau$  на отрезке  $[-x_{max}, x_{max}]$ , а уравнение для  $x(\tau)$  имеет вид:

$$x(\tau) = \begin{cases} \tau & \text{при } 0 \leq \tau \leq x_{max} \\ (-\tau) & \text{при } (-x_{max}) \leq \tau \leq 0 \end{cases} \quad (17)$$

Нормировав обе координаты на величину  $x_{max}$ , введем функцию

$$U(t) = \frac{1}{x_{max}} y(tx_{max}), \quad (18)$$

определенную при  $t \in [-1, 1]$  и изображенную на Рис 3. Эта функция при  $t > 0$  описывает верхний контур профиля (Рис. 2), а при  $t < 0$  описывает нижний контур профиля.

Пусть исходный профиль в плоскости  $(X, Y)$  был задан набором  $N$  пар координат  $\{x(\tau_j), y(\tau_j)\}$ ,  $j = 1, 2, \dots, N$ , для которых, в силу определения системы координат профиля (Рис. 2), выполняются соотношения  $x(\tau_k) = y(\tau_k) = 0$ ;  $x(\tau_1) = x(\tau_N) = x_{max}$  и  $y(\tau_1) = -y(\tau_N)$ . Тогда значения функции



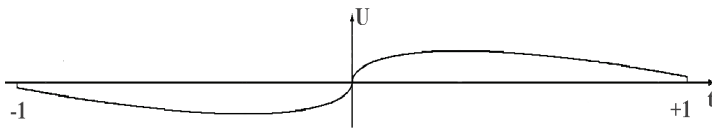


Рис. 3. Форма профиля в виде функции

$U(t)$  также заданы в  $N$  точках  $\{(u_j, t_j), j=1, \dots, N\}$ , определяемых с помощью соотношений (17) и (18).

Для того чтобы построить унифицированное описание кривой, с учетом особенности геометрического объекта строится непрерывная параметрическая аппроксимация  $U^*(t)$  для функции  $U(t)$ . Например, все точки контура профиля имеют ограниченную положительную непрерывную кривизну, поэтому при использовании полиномиальных (или сплайновых полиномиальных) аппроксимаций необходимо предварительно построить монотонное непрерывное преобразование  $t \rightarrow \varphi(t)$  отрезка  $[-1, 1]$  в себя, у которого  $\varphi(t) = O(|t|^{1/2})$  в окрестности точки  $t = 0$ .

Далее построенная аппроксимационная функция  $U^*(t)$ , рассматриваемая как элемент гильбертова пространства  $L_2[-1, 1]$  функций, интегрируемых с квадратом, разлагается в ряд Фурье по некоторому заранее выбранному базису  $\mathbf{P}$  в  $L_2[-1, 1]$ . В качестве такого базиса может быть выбрана, например, система ортонормированных полиномов  $\mathbf{P} = \{P_r(t), r = 0, 1, \dots\}$ , являющихся полиномами Лежандра, Чебышева или Эрмита [40].

Пусть  $\{z_r, r = 0, 1, \dots\}$  – коэффициенты разложения функции  $U^*(t)$  по базису  $\mathbf{P}$ :

$$z_r = \int_{-1}^1 U^*(t) P_r(t) dt, \quad r = 0, 1, \dots \quad (19)$$

Примем бесконечномерный вектор  $\mathbf{z} = (z_0, z_1, z_2, \dots)$ , рассматриваемый как элемент пространства  $l_2$  последовательностей чисел с конечной суммой квадратов, в качестве унифицированного описания геометрического объекта. В частности:

$$\|\mathbf{z}\|^2 = \sum_{r=0}^{\infty} z_r^2 = \int_{-1}^1 (U^*(t))^2 dt < \infty.$$

Рассмотрение геометрического объекта как элемента  $L_2[-1, 1]$ , определяемое соответствующим вектором  $\mathbf{z} \in l_2$ , не зависит от способа задания исходного набора координат точек объекта и дает унифицированное описание объекта «в целом».

Заметим, что при выборе полиномиального базиса и использовании сплайновых полиномиальных аппроксимаций, вычисление коэффициентов (19) может быть сделано в явном аналитическом виде, и эти коэффициенты зависят от ординат  $\{u_j, j = 1, \dots, N\}$  (при фиксированных абсциссах  $\{t_j, j = 1, \dots, N\}$ ) линейным образом.

**Построение наилучшей линейной процедуры снижения размерности.** В классическом методе главных компонент (PCA) выборка состоит из конечномерных векторов. В нашем случае выборка состоит из набора функций  $U^*_{\mathbf{T}} = \{U^*_{1}(t), U^*_{2}(t), \dots, U^*_{\mathbf{T}}(t)\}$ , или из набора соответствующих бесконечномерных векторов  $Z_{\mathbf{T}} = \{z_1, z_2, \dots, z_{\mathbf{T}}\}$ , состоящих из коэффициентов разложения этих функций по ортонормированному базису.

В качестве базового метода может быть предложен Функциональный Анализ Главных Компонент (ФГК), описанный в работах [41–45] и являющийся аналогом конечномерного PCA. В методе ФГК в соответствующем функциональном пространстве строится ортонормированный базис  $\{\gamma_0(t), \gamma_1(t), \dots\}$ , обладающий следующим свойством, определяемым рекуррентным способом:

1) разброс проекций центрированных функций из выборки  $U^*_{\mathbf{T}}$  на  $\gamma_0(t)$  является максимальным (имеет максимальную дисперсию).

Пусть  $L_n$  линейное подпространство, натянутое на первые  $n$  построенные базисные функции  $\{\gamma_0(t), \gamma_1(t), \dots, \gamma_{n-1}(t)\}$ , и  $U^*_{\mathbf{T}n} = \{U^*_{1n}(t), U^*_{2n}(t), \dots, U^*_{\mathbf{T}n}(t)\}$  – ортогональные дополнения к проекции центрированных функций из выборки  $U^*_{\mathbf{T}}$  на это подпространство. Тогда

2) разброс проекций функций из выборки  $U^*_{\mathbf{T}n}$  на  $\gamma_n(t)$  является максимальным (имеет максимальную дисперсию).

Выбрав такой ортонормированный базис, процедура сжатия исходной функции  $U^*(t)$

состоит в проектировании ее на подпространство  $L_n$ , и соответствующий  $n$ -мерный сжатый вектор  $\mathbf{c}$  состоит из коэффициентов разложения

$$c_r = \int_{-1}^1 U^*(t) \gamma_r(t) dt, \quad r = 0, 1, \dots, (n-1).$$

Размерность  $n = n_{\text{lin}}(D_0)$  сжатого вектора выбирается таким образом, чтобы обеспечивался заданный уровень  $D_0$  (16) величины ошибки восстановления (7).

Заметим, что если каждая функция в исходном наборе функций  $U^*_T$  была построена по  $N$  значениям координат точек, имеющих один и тот же набор абсцисс, то гильбертово пространство  $L_2[-1, 1]$  будет конечномерным, и базис из главных компонент  $\{\gamma_0(t), \gamma_1(t), \dots\}$  будет содержать не более  $N$  функций.

**Итеративное построение нелинейной процедуры снижения размерности.** На Шаге 2 построено наилучшее  $n$ -мерное линейное подпространство  $L_n$ ,  $n = n_{\text{lin}}(D_0)$ , аппроксимирующее исходную выборку. Однако возможно существование нелинейного многообразия меньшей размерности, аппроксимирующего исходную выборку с той же заданной точностью. Построению такого многообразия и посвящен Шаг 3.

Шаг 3 состоит в выборе наилучшей процедуры снижения размерности из набора процедур  $\{\Pi_0, \Pi_1, \dots, \Pi_n\}$ ,  $n = n_{\text{lin}}(D_0)$  (параметризованных параметром  $k$ ) – «кандидатом» на наилучшую процедуру построения нелинейного многообразия, аппроксимирующего исходную выборку с заданной точностью  $D_0$ . При  $k = n = n_{\text{lin}}(D_0)$  процедура  $\Pi_n$  совпадает с построенной на Шаге 2 наилучшей линейной процедурой.

Каждая процедура  $\Pi_k$ ,  $k = 0, 1, \dots, n_{\text{lin}}(D_0) - 1$ , состоит из двух последовательных этапов:

- На этапе 1 строится линейное подпространство  $L_k$ , натянутое на первые  $k$  построенных базисных функций  $\{\gamma_0(t), \gamma_1(t), \dots, \gamma_{k-1}(t)\}$ , являющееся наилучшим линейным подпространством размерности  $k$ , аппроксимирующим исходную выборку. Далее строятся ортогональные дополнения  $U^*_{Tk} = \{U^*_{1k}(t), U^*_{2k}(t), \dots, U^*_{Tk}(t)\}$  к проекциям централизованных функций из выборки  $U^*_T$  на подпространство  $L_k$  (разности между элементами вы-

борки и их проекциями на подпространство  $L_k$ ). Заметим, что при  $k = 0$  Шаг 1 отсутствует.

- На этапе 2 строится нелинейное многообразие с помощью нелинейной репликативной нейронной сети (Рис. 2), в которой размер среднего слоя  $n_{\text{ANN}}(k, D_0)$  обеспечивает требуемую точность аппроксимации  $D_0$ . Если выполняется неравенство

$$k + n_{\text{ANN}}(k, D_0) < n_{\text{lin}}(D_0),$$

то процедура  $\Pi_k$  может рассматриваться как возможный кандидат на финальную процедуру снижения размерности.

Финальная процедура снижения размерности позволяет построить многообразие размерности  $n_{\text{fin}}(D_0) = \min \{k + n_{\text{ANN}}(k, D_0),$

$$k = 0, 1, \dots, n_{\text{lin}}(D_0)\}, \quad (20)$$

где при  $k = n_{\text{lin}}(D_0)$  мы полагаем  $n_{\text{ANN}}(n_{\text{lin}}(D_0), D_0) = 0$ . Если минимум в (20) достигается при нескольких  $k$ , то в качестве финальной процедуры выбирается процедура с максимальным значением  $k$ .

## 5. Некоторые особенности задачи снижения размерности

В задачах снижения размерности описания геометрических объектов наряду с требованиями близости между исходной выборкой  $X_T$  и «восстановленным» множеством  $Y_T = \{Y_1, Y_2, \dots, Y_T\}$ , где  $Y_t = F(F(X_t))$ ,  $t = 1, 2, \dots, T$ , могут присутствовать дополнительные (функциональные) требования близости между указанными множествами. А именно, должна быть обеспечена близость множества

$$G_k(X_T) = \{g_k(X_1), g_k(X_2), \dots, g_k(X_T)\}$$

и множества

$$G_k(Y_T) = \{g_k(Y_1), g_k(Y_2), \dots, g_k(Y_T)\}$$

значений множества функционалов  $G = \{g_1, g_2, \dots, g_k\}$  от компонентов исходного множества  $X_T$  и «восстановленного» множества  $Y_T$ ,  $k = 1, 2, \dots, K$ . В качестве используемых мер близости могут быть выбраны величины (11), (12).

Например, в задачах снижения размерностей описания профилей могут быть дополнительные требования к близости кривизны исходного и восстановленного профилей в заданных точках или близость некоторых аэродинамических характеристик профилей.

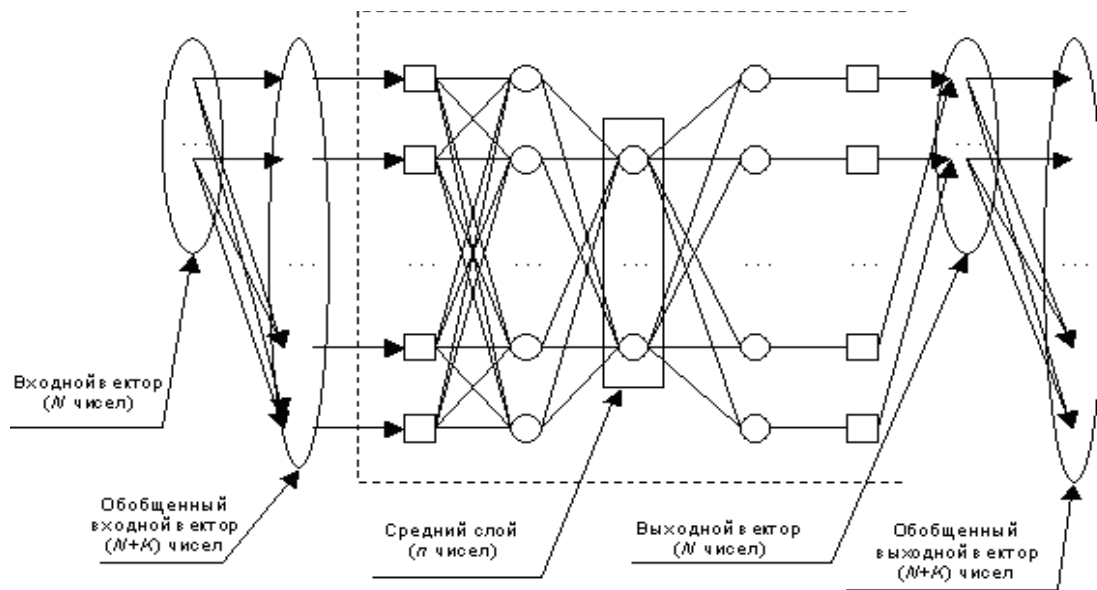


Рис. 4. Обобщенная репликативная нейронная сеть

Множество функционалов  $\mathbf{G}$  порождает новые расстояния  $\rho(\mathbf{X}, \mathbf{Y}, \mathbf{g}_k)$  между векторами  $\mathbf{X}$  и  $\mathbf{Y}$ ,  $k = 1, 2, \dots, K$ , являющиеся аналогами расстояний (13) – (15), но в которых вместо компонент векторов стоят значения соответствующих функционалов от этих компонент. В итоге, формальная постановка задачи снижения размерности состоит в минимизации размерности  $m$  при условии выполнения ограничений:

$$\delta(m, \rho(\cdot, \cdot, \mathbf{g}_k)) \leq D_k; k = 0, 1, 2, \dots, K, \quad (21)$$

где в качестве  $g_0$  выбран функционал  $g_0(x) = x$ . В качестве меры близости  $\delta$  может быть выбрана одна из мер (11) или (12).

Предлагается развитие метода репликативных нейронных сетей, позволяющее обеспечить выполнение требований (21). Как отмечалось выше [19], при линейных функциях активации нейронов такие сети реализуют процедуры метода главных компонент.

Рассмотрим обобщенную нейронную сеть, изображенную на Рис. 4. По сравнению с сетью на Рис. 2, после входного слоя добавлен обобщенный входной слой, состоящий из  $(N + K)$  чисел, в котором подвектор из первых  $N$  чисел совпадает со входным вектором  $\mathbf{X}$ , а оставшийся  $K$ -мерный подвектор  $\mathbf{g}(\mathbf{X})$  состоит из  $K$  значений функционалов  $g_k(\mathbf{X})$ ,  $k = 1, 2, \dots, K$ .

После выходного  $N$ -мерного вектора также добавлен обобщенный  $(N + K)$ -мерный вектор, вычисляемый по тем же правилам.

При заданном числе  $p$  нейронов в среднем слое, коэффициенты нейронной сети подбираются таким образом, чтобы обеспечить максимальную близость (например, в смысле метрик (18) или (19)) между обобщенными входными и выходными векторами, при этом метрика  $\rho$ , участвующая в определении меры близости  $\delta(n, \rho)$ , строится специальным образом, например:

$$\rho_{\infty}(\mathbf{X}, \mathbf{Y}, \mathbf{G}) = \max \{ |X_j - Y_j|, j = 1, 2, \dots, N; \lambda_k \times |g_k(\mathbf{X}) - g_k(\mathbf{Y})|, k = 1, 2, \dots, K \},$$

где  $\lambda_k = D_0/D_k$ ,  $k = 1, 2, \dots, K$ .

Варьируя числом  $p$  нейронов в среднем слое, можно выбрать минимальное значение  $p$ , при котором удовлетворяются заданные требования (21) по точности. После этого часть сети, изображенная на Рис. 5, определяет процедуру сжатия, а часть сети, изображенная на Рис. 6, определяет процедуру восстановления.

## Заключение

При решении различных расчетных и оптимизационных задач, использующих детальное описание геометрических объектов в качестве входных данных, необходимо в качестве пер-

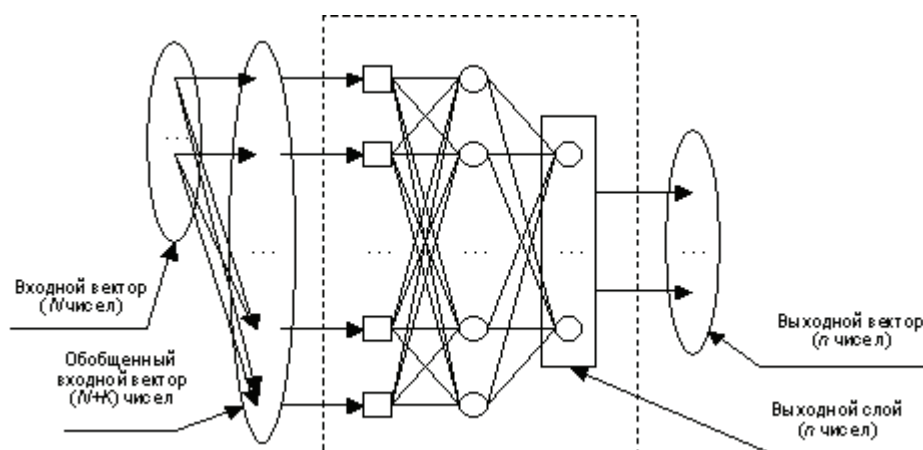


Рис. 5. Процедура сжатия в обобщенной нейронной сети

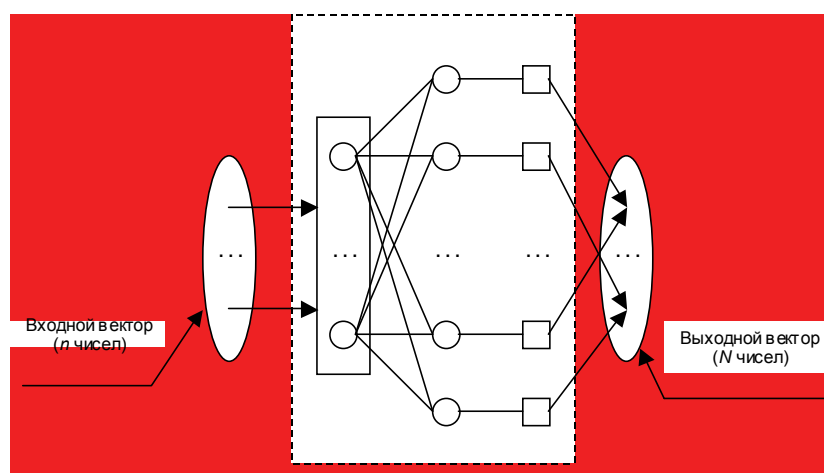


Рис. 6. Процедура восстановления в обобщенной нейронной сети

воочередной важнейшей задачи решать задачу снижения размерности описания объекта, в том числе размерности геометрического описания. Когнитивные технологии являются основным инструментом решения таких задач. В статье подробно описана постановка задачи снижения размерности, приведен краткий обзор основных используемых подходов и методов и предложен общий подход к построению процедур снижения размерности описания геометрических объектов.

## Литература

1. Кулешов А.П. Когнитивные технологии в основанных на данных адаптивных моделях сложных объектов//Информационные технологии и вычислительные системы, № 1, 2008.
2. Кулешов А.П. Технология быстрого вычисления характеристик сложных технических объектов // Информационные технологии. Прил. 2006. № 3. С. 4-11.
3. Кулешов А.П. Основные принципы технологии быстрых расчетов, основанной на знаниях//Тезисы докладов Международной научно-технической конференции «Информационные технологии и математическое моделирование в науке, технике и образовании». Сицилия (Италия), 28 сентября - 5 октября 2006 г.
4. Кулешов А.П. Информационные технологии в проблеме создания моделей сложных объектов//Вторая Международная конференция «Системный анализ и информационные технологии» САИТ-2007 (10 – 14 сентября 2007 г., г. Обнинск, Россия): Труды конференции. Т. 1, стр. 14 – 16. 2007.
5. А.В. Бернштейн, А.П. Кулешов. Задачи снижения размерности моделей сложных объектов//Вторая Международная конференция «Системный анализ и информационные технологии» САИТ-2007 (10 – 14 сентября

- 2007 г., г. Обнинск, Россия): Труды конференции. Т. 1, стр. 243 – 247. 2007.
6. A.K. Jain, R.C. Dubes. Algorithms for Clustering Data. Englewood Cliffs, NJ: Prentice Hall, 1988.
  7. J. Bruske, G. Sommer. Intrinsic Dimensionality Estimation with Optimally Topology Preserving Maps. IEEE Transactions of Pattern Analysis and Machine Intelligence, vol. 20, no. 5, pp. 572-575, 1998.
  8. M.F. Barnsley. Fractals Everywhere. Academic Press, New York, 1988.
  9. H.-O. Peitgen, H. Jurgensm, D. Saupe. Chaos and Fractals: New Frontiers of Science. Springer-Verlag, New York, 1992.
  10. P. Grassberger, I. Procaccia. Measuring the Strangeness of Strange Attractors. Physica vol. D9, pp. 189-208, 1983.F.
  11. Camastro, A. Vinciarelli. Estimating the Intrinsic Dimension of Data with a Fractal-Based Method. IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 24, num. 10, 2002, p. 1404-1407.
  12. J.E. Jackson. A User's Guide to Principal Components. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, New York, London, Sydney, 1991.
  13. I.T. Jolliffe. Principal Component Analysis. Springer Series in Statistics. Springer-Verlag, Berlin, 1986.
  14. Hornick K., Stinchcombe M., White H. Multilayer Feedforward Networks are Universal Approximators//Neural Networks, 1989, v. 2, № 5, P. 359-366.
  15. Saund. Dimensionality-reduction using connectionist networks. IEEE Trans. On Pattern Anal. And Machine Intel., 11(3):304-314, 1989.
  16. M.K. Fleming, G.W. Cottrell. Categorization of faces using unsupervised feature extraction. In proc. Int. J. Conf. on Neural Networks (IJCNN90), vol. II, pp. 65-70, San Diego, CA, June 17-21, 1990.
  17. M.A. Kramer. Nonlinear principal component analysis using autoassoaitive neural networks. Jopurnal of the American institute of Chemical Engineers, 37(2):233-243, Feb. 1991.
  18. D. DeMers, G.W. Cottrell. Nonlinear dimensionality reduction. In S.J. Hanson, J.D. Cowan, C.L. Giles, eds. Advances in Neural Information processing Systems, vol. 5, pp. 580-587. Morgan Kaufman, San Mateo, 1993.
  19. K.I. Diamantaras, S.-Y. Kung. Principal component neural network. Theory and applications. Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications and Control. John Wiley and Sons, New Yorkm London , Sydney, 1996.
  20. Sirovich, M. Kirby. Low-dimensional procedure for the identification of human faces. J. Opt. Soc. Amer. A, 4(3):519-524, Mar. 1987.
  21. N. Cristiani, J. Shawe-Taylor. An introduction to Support vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, 2000.
  22. B. Scholkopf, C.J.C. Burges, A.J. Smola. Advances in Kernel Methods. Support Vector Learning. MIT Press, 1999.
  23. J.H.Friedman, J.W. Tukey. A projection pursuit algorithm for exploratory data analysis. IEEE trans. Computers, C-23: 881-889, 1974.
  24. P.J. Huber. Projection pursuit. Annals of Statistics, 13(2): 435-475, May 1985.
  25. M.C. Jones, R. Sibson. What is projection pursuit? Journal of the Royal Statistical Society, A, 150(1): 1-18, 1987.
  26. K.V. Mardia, J.T. Kent, J.M. Bibby. Multivariate analysis. Probability and Mathematical Statistics. Academic Press, 1995.
  27. J. Friedman, T. Hastie, R. Tibshirani. Elements of Statistical Learning: Prediction, Inference and Data Mining. Springer. 2001.
  28. С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. Прикладная статистика: классификация и снижение размерности. М.: Финансы и статистика, 1989.
  29. I.K. Fodor. A survey of dimension reduction technique. LLNL technical report, June 2002. UCRL-ID-148494.
  30. J.H. Friedman, W. Stuetzle. Projection pursuit regression. J. Amer. Stat. Assoc., 76(376):817-823, Dec. 1981.
  31. T.J. Hastie, W. Stuetzle. Principal curves. J. Amer. Stat. Assoc., 84(406):502-516, 1989.
  32. Давыдов Ю.В., Злыгарев В.А. Геометрия крыла: методы и алгоритмы проектирования несущих поверхностей. М., Машиностроение, 1987, 136 с.
  33. E. N. Jacobs, K. E. Ward, & R. M. Pinkerton 1933 The characteristics of 78 related airfoil sections from tests in the variable-density wind tunnel, NACA Report No. 460.
  34. E. N. Jacobs & R. M. Pinkerton 1936 Test in the variable-density wind tunnel of related airfoils having the maximum camber unusually far forward, NACA Report No. 537.
  35. Dorofeyev Ye.A., Sviridenko Yu.N. The application of replicator neural networks to the problems of aerodynamic design of aircraft components//International Symposium on Aerospace Technologies of the XXI century: New Challenges in Aeronautics. Zhukovsky, Russia, 14 – 19 August 2001.
  36. Alexander V. Bernstein, Alexander P. Kuleshov, Yuri N. Sviridenko, Victor V. Vyshinsky. Fast Aerodynamic Model for Design Technology//Proceedings of West-East High Speed Flow Field Conference. November 19-22, 2007, Moscow, Russia, <http://wehsff.imamod.ru/pages/s7.htm>.
  37. Alexander V. Bernstein, Alexander P. Kuleshov, Yuri N. Sviridenko, Victor V. Vyshinsky. Fast Aerodynamic Model for Design Technology//Workbook “West-East High Speed Flow Field Conference. November 19-22, 2007, Moscow, Russia”, p. 125 – 126.
  38. Бернштейн А.В., Вышинский В.В., Кулешов А.П., Свириденко Ю.Н. Применение искусственных нейронных сетей для определения нагрузок по крылу пассажирского самолета на режиме крейсерского полета. Труды Центрального аэрогидродинамического института им. проф. Н.Е. Жуковского, Выпуск №2678 «Применение искусственных нейронных сетей в задачах прикладной аэродинамики», Москва, 2008.
  39. Бернштейн А.В., Вышинский В.В., Кулешов А.П., Свириденко Ю.Н. Быстрый метод аэродинамического расчета для задач проектирования. Труды Центрального аэрогидродинамического института им. проф.

- Н.Е. Жуковского, Выпуск №2678 «Применение искусственных нейронных сетей в задачах прикладной аэродинамики», Москва, 2008.
40. Суетин П.К. Классические ортогональные многочлены. М.: ФИЗМАТЛИТ, 2005.
41. Dauxois J., Pousse A., Romain Y. Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *Journal of Multivariate analysis*, 12, P. 136-154, 1982.
42. Ramsay J., Silverman B. *Functional Data Analysis*. Springer, NY, 1997.
43. Ramsay J., Silverman B. *Applied functional data analysis*. Springer, NY, 2002.
44. Ferraty F., P. Vieu. *Nonparametric functional data analysis: theory and practice*. Springer, NY, 2006.
45. Ocaña F.A., Aguilera A.M., Valderrama M.J. Functional Principal Components Analysis by choice of Norm. *J. of Multivariate Analysis*, 71, 262 – 276, 1999.

**Бернштейн Александр Владимирович.** Заведующий лабораторией анализа и обработки данных Института системного анализа РАН. Окончил механико-математический факультет МГУ им. Ломоносова в 1969 г. Доктор физико-математических наук, профессор. Автор 60 печатных научных работ. Область научных интересов: математическая статистика и анализ данных, информационные технологии, моделирование.

**Кулешов Александр Петрович.** Директор Института проблем передачи информации им. А.А. Харкевича РАН, заместитель академика-секретаря Отделения нанотехнологий и информационных технологий РАН. Окончил механико-математический факультет МГУ им. Ломоносова в 1970 г. Член-корреспондент РАН, доктор технических наук, профессор. Автор 75 печатных научных работ. Область научных интересов: информационные и телекоммуникационные технологии, моделирование.