*Информационные процессы, Том 20, № 4, 2020, стр. 317–329*




# Функциональное снижение размерности в метамоделировании

**Е.В. Бурнаев[*,**], А.В. Бернштейн[*]**

[*] *Сколковский институт науки и технологий, Москва, Россия*
[**] *Институт проблем передачи информации, Российская академия наук*
*e-mail: e.burnaev@skoltech.ru*

Поступила в редколлегию 01.10.2020

**Аннотация**—В работе рассмотрена задача снижения размерности в метамоделировании. Показано, что для построения по редуцированным данным точной метамодели необходимо решать задачу снижения размерности в нестандартной формулировке, а именно, с учетом дополнительных функциональных ограничений. Для построения по данным касательного аффинного линейного подпространства, необходимого при решении задачи функционального снижения размерности, сформулирована и решена обобщенная задача построения главных компонент.

**КЛЮЧЕВЫЕ СЛОВА:** снижение размерности, функциональное снижение размерности, обобщенная задача построения главных компонент.

## 1. ВВЕДЕНИЕ

В процессе проектирования сравниваются различные технические решения, касающиеся структуры и параметров объекта, механизмов его функционирования и других элементов объекта. Компьютерные системы для поддержки принятия инженерных решений (Knowledge Based Engineering, Computer Aided Engineering) создаются для сокращения времени проектирования и числа дорогостоящих (в материальном и временном смыслах) натурных экспериментов и являются, по существу, системами моделирования. Значительный прогресс в области математического моделирования и возможностей вычислительной техники позволяет исследовать большое количество вариантов построения объекта (конфигурации, параметров и др.), предсказывать ожидаемые характеристики и находить наилучшие (рациональные) решения без проведения натурных экспериментов. Поэтому математическое моделирование (вычислительные эксперименты для исследования аналитических моделей создаваемого объекта и окружающей его среды) стало одним из самых распространенных методов анализа и оптимизации структуры технических объектов.

Традиционно в моделировании используются математические модели, основанные на “физике процессов” и описывающие физические процессы и явления, происходящие при функционировании объекта, сложными дифференциальными уравнениями в частных производных с граничными условиями, для которых зачастую неизвестны ни теоремы о существовании и единственности решения, ни характер зависимости решения от параметров и граничных условий. Численные методы решения таких уравнений имеют значительную вычислительную трудоемкость, как самих расчетов, так и подготовки исходных данных и расчетных сеток. Это существенно сокращает возможности использования моделей, основанных на “физике процессов”, особенно на стадии предварительного (концептуального) проектирования, на которой рассматривается очень большое количество вариантов решений и высока цена неправильно выбранного решения. Тем самым, актуальна проблема создания технологии, позволяющей в режиме реального времени проводить сравнение большого числа вариантов построения сложных технических объектов с обеспечением требуемой достоверности выводов. Такой технологией является информационная Технология быстрых расчетов.

Технология быстрых расчетов предназначена для построения математических моделей, основанных на данных, с минимальным привлечением знаний из предметной области (физики процессов). Построение основанных на данных моделей основано на идеях машинного обучения (machine learning): модели "обучаются" по множеству прототипов входных и выходных данных (результатов натурных и/или вычислительных экспериментов, проведенных с различными объектами рассматриваемого класса). Построенные модели фактически имитируют (заменяют) как источники получения данных, основанные на некоторой исходной модели, так и сами модели, созданные на основе изучения физики процессов. Поэтому, такие адаптивные модели иногда называют также метамоделями (модели над моделями) или суррогатными моделями (Surrogate Models), см. [1, 2].

Методы построения метамоделей, описанные в литературе, касаются, в основном, решения задачи восстановления по данным неизвестной функции (построения аппроксимирующей функции). Однако, при построении метамоделей необходимо решать ряд взаимосвязанных задач интеллектуального анализа данных в ситуации, когда выходные данные одной частной задачи являются входными данными для другой частной задачи, и целевые функции для частных задач нельзя определить независимо.

К решаемым задачам относятся задачи определения внутренней размерности множества данных, построения процедур снижения размерности [3], построения многомерных аппроксимирующих зависимостей [4], кластеризации и классификации данных, и др. Существенной особенностью решаемых задач является тот факт, что данные лежат, как правило, на нелинейных многообразиях существенно меньшей размерности (по сравнению с размерностью исходных данных). Однако, как мы показываем в данной работе, на практике требуется не просто учесть информацию о зависимостях между признаками входных данных при построении метамодели, а построить многообразие данных, которое обеспечит наилучшее восстановление выходных переменных в зависимости от входных переменных [5, 6] (т.н. "функциональное снижение размерности"), и оценить его касательное пространство [7]. Обычно при решении такой задачи рассматриваются модели линейных многообразий в пространстве входных признаков [8]. В данной работе формулируется общая постановка задачи функционального снижения размерности и решается задача оценки локальных линейных аффинных подпространств для соответствующего многообразия.

Статья имеет следующую структуру. В разделе 2 рассматривается постановка задачи построения метамоделей. В разделе 3 формулируется постановка задачи снижения размерности и обсуждаются требования к её решению. В разделе 4 рассматриваются особенности задачи снижения размерности при её использовании в процессе построения метамоделей. В разделе 5 авторами формулируется постановка задачи функционального снижения размерности. В разделе 6 формулируется задача построения локальных линейных аффинных подпространств, решение которой необходимо при функциональном снижении размерности, и показывается, как эта задача может быть сведена к обобщенной задаче анализа главных компонент, которая решается в разделе 7. Заключение приведено в разделе 8.

## 2. ПОСТРОЕНИЕ ПО ДАННЫМ МЕТАМОДЕЛЕЙ КАК БАЗОВЫЙ ЭЛЕМЕНТ ТЕХНОЛОГИИ БЫСТРЫХ РАСЧЕТОВ

В самом общем виде базовая задача построения метамодели формулируется следующим образом. Пусть $B = \{b\}$ есть множество объектов рассматриваемого класса. Для каждого объекта $b \in B$ имеется его цифровое описание $\mathbf{x} = \mathbf{x}(b)$ размерности $d$. Обозначим $\mathcal{X} =$

$\mathcal{X}(B) = \{\mathbf{x}(b),\ b \in B\}$ подмножество $d_x$-мерного евклидова пространства $\mathbb{R}^{d_x}$, состоящее из цифровых описаний объектов рассматриваемого класса.

Пусть $\mathbf{y} = \mathbf{y}(b)$ — некоторая числовая (в общем случае, векторная) характеристика объекта $b$, описывающая свойства (поведение) объекта в некоторых условиях. Характеристика $\mathbf{y}$ может быть описана в виде функциональной зависимости $\mathbf{y} = f(\mathbf{x}, \mathbf{u})$, в которой переменная $\mathbf{x} = \mathbf{x}(b)$ описывает сам объект, а переменная $\mathbf{u} \in \mathbb{R}^{d_u}$ описывает условия функционирования объекта (параметры управления объектом, параметры внешней среды) и принадлежит множеству различных значений $\mathcal{U}$.

Для вычисления значений функции $\mathbf{y} = f(\mathbf{x}, \mathbf{u})$ могут использоваться различные методы, например, вычислительные или натурные эксперименты. Пусть $M$ — некоторый метод (модель) получения характеристики $\mathbf{y}$, тогда $f(\mathbf{x}, \mathbf{u})$ есть результат вычисления значения характеристики $\mathbf{y}$ для объекта $\mathbf{x}$ в условиях $\mathbf{u}$. Пусть имеется некоторое количество измерений характеристики $\mathbf{y}$, то есть

$$S_n = \{(\mathbf{x}_i, \mathbf{u}_i,\ \mathbf{y}_i = f(\mathbf{x}_i, \mathbf{u}_i))\}_{i=1}^{n}, \tag{1}$$

полученных с помощью модели $M$ для разных значений аргументов, принадлежащих множеству $\{\mathbf{X}_n, \mathbf{U}_n\} = \{(\mathbf{x}_1, \mathbf{u}_1), \ldots, (\mathbf{x}_n, \mathbf{u}_n)\}$.

Множество $S_n$ (1) состоит из известных значений функции $f(\mathbf{x}, \mathbf{u})$, по которым строится аппроксимация $\hat{f}(\mathbf{x}, \mathbf{u}) = \hat{f}(\mathbf{x}, \mathbf{u}\,|S_n)$ для функции $f(\mathbf{x}, \mathbf{u})$. Если построенная по данным $S_n$ функция $\hat{f}(\mathbf{x}, \mathbf{u})$ хорошо аппроксимирует функцию $f(\mathbf{x}, \mathbf{u})$ в области изменения переменных $\mathcal{X} \times \mathcal{U}$, то есть $\hat{f}(\mathbf{x}, \mathbf{u}) \approx f(\mathbf{x}, \mathbf{u}),\ (\mathbf{x}, \mathbf{u}) \in \mathcal{X} \times \mathcal{U}$, то функция $\hat{f}(\mathbf{x}, \mathbf{u})$ определяет новую модель (метамодель) $SM$, которая может рассматриваться как заменитель (surrogate) для модели $M$.

Как правило, функция $\hat{f}(\mathbf{x}, \mathbf{u})$ имеет явный аналитический вид, и поэтому суррогатная модель может иметь существенно более высокую вычислительную эффективность по сравнению с исходной моделью $M$.

В приложениях размерность $d_x$ цифрового описания объекта может быть очень высока. Например, 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 геометрического 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 CAD-системах, компьютерной графике и других приложениях. Такие цифровые описания объекта (например, 3D-поверхности самолета) состоят из десятков тысяч чисел, каждое из которых, рассматриваемое изолированно, не несет смысловой нагрузки. Высокая размерность цифрового описания объектов существенно затрудняет аппроксимацию функции, зависящей от векторов высокой размерности.

Однако в приложениях, как правило, множество $\mathcal{X} \times \mathcal{U}$ лежит, по крайней мере, приближенно, на многообразии (в общем случае, нелинейном) $\mathcal{M}_q \subset \mathbb{R}^{d_x + d_u}$, размерность $q$ которого существенно меньше исходной размерности $d_x + d_u$. Следовательно, задачу аппроксимации необходимо решать лишь в окрестности многообразия $\mathcal{M}_q$. Нахождение этого многообразия $\mathcal{M}_q$ очень важно и по другой причине. Для проведения вычислительных экспериментов с целью получения обучающего множества данных $S_n$, необходимо генерировать входные данные $(\mathbf{x}, \mathbf{u})$, и при этом цифровые описания $\mathbf{x}$ должны генерироваться в окрестности многообразия $\mathcal{M}_q$ (в частности, необходимо "оставаться" вблизи многообразия $\mathcal{M}_q$ при генерации новых объектов в процессе оптимизации).

Таким образом, актуальна задача создания новых методов построения многообразий, аппроксимирующих данные, с учетом особенностей их дальнейшего использования при построении метамоделей. Отметим, что рассматриваемые далее проблемы не связаны с наличием/отсутствием описания $\mathbf{u}$ условий функционирования объекта. Соответственно, мы ограничимся рассмотрением функций $\mathbf{y} = f(\mathbf{x})$, зависящих только от аргумента $\mathbf{x}$, который может включать как цифровое описание объекта, так и условия его функционирования (в зависимости от контекста рассматриваемой ситуации). Здесь $\mathbf{x} \in \mathbb{R}^d$, $d(= d_x + d_u)$.

## 3. ЗАДАЧА ПОСТРОЕНИЯ МНОГООБРАЗИЙ, АППРОКСИМИРУЮЩИХ ДАННЫЕ

Задача построения многообразий, аппроксимирующих данные, может быть сформулирована как обобщенная задача снижения размерности. Пусть аппроксимирующее многообразие $\mathcal{M}_q$ состоит из одной карты и параметризовано $q$-мерным параметром $\mathbf{z} \in \mathcal{Z} \subset \mathbb{R}^q$, $q < d$. Другими словами, многообразие $\mathcal{M}_q$ имеет вид:

$$\mathcal{M}_q = \left\{ g(\mathbf{z}) \in \mathbb{R}^d : \mathbf{z} \in \mathcal{Z} \subset \mathbb{R}^q \right\}. \tag{2}$$

Пусть $\|\cdot\|$ — некоторая норма в $\mathbb{R}^d$. Рассмотрим для многообразия $\mathcal{M}_q$ трубку $U_\varepsilon(\mathcal{M}_q)$ максимального радиуса $\varepsilon$. Это означает, что для всех $\mathbf{x}$, принадлежащих $\varepsilon$-окрестности $U_\varepsilon(\mathcal{M}_q)$ многообразия $\mathcal{M}_q$, $\inf_{\mathbf{z} \in \mathcal{Z}} \|\mathbf{x} - g(\mathbf{z})\|$ достигается в единственной точке $\mathbf{z} = \mathbf{z}(\mathbf{x}) \in \mathcal{Z}$, и $\|\mathbf{x} - g(\mathbf{z}(\mathbf{x}))\| \leqslant \varepsilon$.

Если множество цифровых описаний $\mathcal{X}$ лежит в трубке $U_\varepsilon(\mathcal{M}_q)$, то для произвольного объекта $b \in B$ с цифровым описанием $\mathbf{x} = \mathbf{x}(b) \in \mathcal{M}_q$ можно построить модельный объект $b^*$ с цифровым описанием $\mathbf{x}(b^*) = g(\mathbf{z}(b)) \in \mathcal{M}_q$, отличающимся от $\mathbf{x}(b)$ не более чем на величину $\varepsilon$.

Рассмотрим множество модельных объектов $\mathbf{B}^* = \{b^* : \mathbf{x}(b^*) = g(\mathbf{z}(b)),\ b \in B\}$, цифровые описания которых параметризованы параметром $\mathbf{z} \in \mathcal{Z}$, размерность $q$ которого существенно меньше $d$, и $\mathbf{B}^* = \{b^* : \mathbf{x}(b^*) = g(\mathbf{z}),\ \mathbf{z} \in \mathcal{Z}(B) \subset \mathbb{R}^q\}$, где

$$\mathcal{Z}(B) = \{\mathbf{z}(b),\ b \in B\}. \tag{3}$$

Задача построения по данным многообразия $\mathcal{M}_q$ формулируется как задача снижения размерности $\Pi = \{q,\ \mathbf{C}_q,\ \mathbf{R}_q\}$ : для заданной размерности $q$ по заданным цифровым описаниям

$$\mathbf{X}_n = \{\mathbf{x}(b_1),\ \mathbf{x}(b_2), \ldots, \mathbf{x}(b_n)\} \equiv \{\mathbf{x}_1,\ \mathbf{x}_2, \ldots, \mathbf{x}_n\} \tag{4}$$

множества прототипов $\{b_i \in B,\ i = 1, 2, \ldots, n\}$ построить два отображения $\mathbf{C}_q$ и $\mathbf{R}_q$:

- преобразование сжатия $\mathbf{C}_q : \mathbb{R}^d \to \mathbb{R}^q$, т.ч. $\mathbf{x} \to \mathbf{z} = \mathbf{z}(\mathbf{x}) = \mathbf{C}_q(\mathbf{x}) \in \mathbb{R}^q$,
- преобразование восстановления $\mathbf{R}_q : \mathbb{R}^q \to \mathbb{R}^d$, т.ч. $\mathbf{z} \to \mathbf{x}^* = \mathbf{x}^*(\mathbf{z}) = \mathbf{R}_q(\mathbf{z}) \in \mathbb{R}^d$,

принадлежащих выбранным классам преобразований, таким образом, чтобы обеспечивалось приближенное равенство:

$$\mathbf{x} \approx \mathbf{x}^* = \mathbf{R}_q(\mathbf{C}_q(\mathbf{x})), \quad \mathbf{x} \in \mathcal{X}. \tag{5}$$

Рассматривая в качестве меры близости между векторами в $\mathbb{R}^d$ $L_2$-норму, нетрудно видеть, что при выбранном преобразовании восстановления $\mathbf{R}_q$ в качестве преобразования наилучшего сжатия $\mathbf{C}_q$ следует выбрать проекцию $\mathbf{C}_q(\mathbf{x}) = \arg\min_{\mathbf{z} \in \mathcal{Z}} \|\mathbf{x} - \mathbf{R}_q(\mathbf{z})\|$ на многообразие

$$\mathcal{M}_q = \left\{ \mathbf{x} = \mathbf{R}_q(\mathbf{z}) \in \mathbb{R}^d,\ \mathbf{z} \in \mathcal{Z} \subset \mathbb{R}^q \right\}. \tag{6}$$

Отсюда следует, что с геометрической точки зрения решение задачи снижения размерности по данным (заданному множеству точек в полноразмерном пространстве) сводится к построению многообразия меньшей размерности (аппроксимирующего многообразия) такого, что сумма квадратов расстояний между заданными точками и их проекции на это многообразие минимальна. Поэтому в литературе для рассматриваемой задачи в последние годы наряду с термином "снижение размерности" (dimension reduction) используется также термин "обучение многообразия" (manifold learning) — аналог термина "машинное обучение" (machine learning) применительно к задаче построения по данным аппроксимирующей зависимости.

Формальная постановка задачи построения по данным $\mathbf{X}_n$ преобразований $\mathbf{C}_q$ и $\mathbf{R}_q$, минимизирующих среднюю ошибку восстановления

$$\varepsilon_n = \varepsilon_n(\Pi) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i - \mathbf{R}_q(\mathbf{C}_q(\mathbf{x}_i))\|^2}.$$

Если при этом для всех $\mathbf{x} \in \mathcal{X}(B)$ имеет место соотношение,

$$\|\mathbf{x} - \mathbf{R}_q(\mathbf{C}_q(\mathbf{x}))\| \leqslant \varepsilon, \tag{7}$$

то множество $\mathcal{X}(B)$ лежит в $\varepsilon$-трубке многообразия $\mathcal{M}_q$ (2) с $g(\mathbf{z}) = \mathbf{R}_q(\mathbf{z})$ и $\mathcal{Z} = \mathcal{Z}(B)$ (3).

Отображения $\mathbf{C}_q$ и $\mathbf{R}_q$, удовлетворяющие соотношению (7), позволяют аппроксимировать $d$-мерные цифровые описания $\{\mathbf{x}(b)\}$ объектов $\{b\}$ цифровыми описаниями $\{\mathbf{x}(b^*)\}$ модельных объектов $\{b^*\}$, параметризованных $q$-мерным параметром $\mathbf{z}$, и $q$-мерный вектор $\mathbf{z}(\mathbf{x}(b)) = \mathbf{C}_q(\mathbf{x}(b))$, можно считать сжатым цифровым описанием объекта $b$, по которому восстанавливается цифровое описание $\mathbf{x}^* = \mathbf{x}(b^*) = \mathbf{R}_q(\mathbf{z}(\mathbf{x}(b)))$ модельного объекта $b^*$.

Однако такая аппроксимация объектов модельными объектами, удовлетворяющая требованиям по точности:

$$\|\mathbf{x}(b) - \mathbf{x}(b^*)\| \leqslant \varepsilon, \tag{8}$$

может оказаться недостаточной при использовании модельных объектов для построения метамоделей.

## 4. ОСОБЕННОСТИ ЗАДАЧИ СНИЖЕНИЯ РАЗМЕРНОСТИ ПРИ ЕЕ ИСПОЛЬЗОВАНИИ В ПРОЦЕССЕ ПОСТРОЕНИЯ МЕТАМОДЕЛЕЙ

Пусть по множеству данных

$$S_n = \{(\mathbf{x}_i,\ \mathbf{y}_i = f(\mathbf{x}_i)),\ i = 1, 2, \ldots, n\} \tag{9}$$

необходимо построить метамодель $\mathbf{y} = \hat{f}(\mathbf{x})$. Пусть построено аппроксимирующее многообразие $\mathcal{M}_q$ (2), (6) с $g(\mathbf{z}) = \mathbf{R}_q(\mathbf{z})$ и $\mathcal{Z} = \mathcal{Z}(B)$ (3).

Рассмотрим новую функцию $\mathbf{y} = F(\mathbf{z}) = f(\mathbf{R}_q(\mathbf{z})), \quad \mathbf{z} \in \mathcal{Z} \subset \mathbb{R}^q$, для которой в силу (5) можно также ожидать выполнение приближенного равенства:

$$f(\mathbf{x}) \approx f(\mathbf{x}^*) = f(\mathbf{R}_q(\mathbf{C}_q(\mathbf{x}))) = F(\mathbf{C}_q(\mathbf{x})) = F(\mathbf{z}(\mathbf{x})). \tag{10}$$

Рассмотрим выборку $s_n = \{(\mathbf{z}_i,\ \mathbf{y}_i) :\ \mathbf{z}_i = \mathbf{C}_q(\mathbf{x}_i),\ \mathbf{y}_i = f(\mathbf{x}_i),\ \ i = 1, 2, \ldots, n\}$, состоящую из редуцированных данных, полученных путем проектирования аргументов (4) исходной выборки $S_n$ (9) на многообразие (6). По этой выборке можно построить аппроксимационную зависимость $\hat{\mathbf{y}} = \hat{F}(\mathbf{z}) = \hat{F}(\mathbf{z}\,|s_n)$.

Если для новой аппроксимационной зависимости (10) выполняется соотношение

$$\hat{F}(\mathbf{z}) \approx F(\mathbf{z}), \tag{11}$$

то в совокупности с (10) это обеспечивает цепочку приближенных равенств $\hat{F}(\mathbf{C}_q(\mathbf{x})) \approx F(\mathbf{C}_q(\mathbf{x})) \approx f(\mathbf{x})$. Тем самым, исходную задачу построения аппроксимационной зависимости для функции $f$, зависящей от $d$-мерного аргумента, мы заменили задачей построения аппроксимационной зависимости для функции $F$, зависящей уже от $q$-мерного аргумента, $q < d$.

Однако для корректности такой замены необходимо не только обеспечение точности аппроксимации (11), но и выполнение приближенных равенств (10). Следовательно, процедура снижения размерности, используемая для снижения размерности аргумента в задаче построения метамодели, должна обеспечивающую не только близость в пространстве аргументов (близость цифровых описаний $\mathbf{x}(b)$ и $\mathbf{x}(b^*)$ (8)), но и функциональную близость (близость в пространстве значений функций $f(\mathbf{x}(b))$ и $f(\mathbf{x}(b^*))$ (10)). Поэтому в данной работе поставлена и изучена задача функционального снижения размерности.

## 5. ЗАДАЧА ФУНКЦИОНАЛЬНОГО СНИЖЕНИЯ РАЗМЕРНОСТИ

Задача функционального снижения размерности формулируется следующим образом: по заданной выборке $\mathbf{X}_n = \{\mathbf{x}_1,\ \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ из $d$-мерного множества $\mathcal{X} \subset \mathbb{R}^d$ для заданной размерности $q < d$ и заданной векторной функции

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}),\ f_2(\mathbf{x}), \ldots, f_k(\mathbf{x}))^\top \tag{12}$$

построить два преобразования: $\mathbf{C}_q$ (преобразование сжатия) и $\mathbf{R}_q$ (преобразование восстановления), принадлежащих выбранным классам преобразований, таким образом, чтобы обеспечивались приближенное равенство $\mathbf{x} \approx \mathbf{x}^* = \mathbf{R}_q(\mathbf{C}_q(\mathbf{x}))$ (5) и приближенное равенство

$$f_j(\mathbf{x}) \approx f_j(\mathbf{R}_q(\mathbf{C}_q(\mathbf{x}))), \quad \mathbf{x} \in \mathcal{X}, \quad j = 1, 2, \ldots, k. \tag{13}$$

Здесь и далее векторы записываются как векторы-столбцы.

Соответствующая формальная математическая задача состоит в минимизации средней ошибки восстановления

$$\varepsilon_n(\mathbf{f}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\theta_0^2\|\mathbf{x}_i - \mathbf{R}_q(\mathbf{C}_q(\mathbf{x}_i))\|^2 + \sum_{j=1}^{k}\theta_j^2|f_j(\mathbf{x}_i) - f_j(\mathbf{R}_q(\mathbf{C}_q(\mathbf{x}_i)))|^2},$$

где $\theta_0,\ \theta_1, \ldots, \theta_k$ — заданные неотрицательные веса, т.ч. $\theta_0^2 + \theta_1^2 + \ldots + \theta_k^2 = 1$, $\theta_0 > 0$.

При $\theta_0 = 1$ или при $f_j(\mathbf{x}) \equiv \mathbf{x}$ для всех $j = 1, 2, \ldots, k$, сформулированная задача совпадает с обычной задачей снижения размерности. Далее в отчете будем рассматривать только положительные веса $\theta_0,\ \theta_1, \ldots, \theta_k$, общий случай сводится к изменению значения $k$.

Очевидно, что требование (13) накладывает дополнительные функциональные ограничения на аппроксимирующее многообразие $\mathcal{M}_q$.

## 6. ЛОКАЛЬНЫЕ ЛИНЕЙНЫЕ АФФИННЫЕ ПОДПРОСТРАНСТВА В ЗАДАЧЕ ФУНКЦИОНАЛЬНОГО СНИЖЕНИЯ РАЗМЕРНОСТИ

Пусть $p = d + k$. Рассмотрим $p$-мерное пространство $\mathbb{R}^p = \mathbb{R}^p(\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\theta_0,\ \theta_1, \ldots, \theta_k)^\top$, состоящее из $p$-мерных векторов $\mathbf{v} = (\mathbf{x} : \mathbf{f})$, составленных из двух подвекторов: $d$-мерного подвектора $\mathbf{x}$ и $k$-мерного подвектора $\mathbf{f} = (f_1,\ f_2, \ldots, f_k)^\top$, в котором введем норму

$$\|\mathbf{v}\|_\theta = \|(\mathbf{x} : \mathbf{f})\|_\theta = \sqrt{\theta_0^2\|\mathbf{x}\|^2 + \sum_{j=1}^{k}\left(\theta_j^2|f_j|^2\right)}, \tag{14}$$

где $\|\cdot\|$ — обычная евклидова норма в $\mathbb{R}^d$.

Рассмотрим в $\mathbb{R}^p\,(\theta)$ $p$-мерное расширенное многообразие $\mathcal{H}_q$:

$$\mathcal{H}_q = \left\{(\mathbf{R}_q(\mathbf{z}):\ (f_1(\mathbf{R}_q(\mathbf{z}))), (f_2(\mathbf{R}_q(\mathbf{z}))), \ldots, (f_k(\mathbf{R}_q(\mathbf{z}))))^\top : \mathbf{z} \in \mathcal{Z} \subset \mathbb{R}^q\right\} \subset \mathbb{R}^p. \qquad (15)$$

В дальнейшем будем называть вектор

$$\mathbf{v}\,(\mathbf{x}\,(b)) = (\mathbf{x}\,(b) : \mathbf{f}\,(\mathbf{x}\,(b))) \qquad (16)$$

обобщенным цифровым описанием объекта $b$, а многообразие $\mathcal{H}_q$ — обобщенным многообразием модельных объектов. Имеет место

**Теорема 1.** *Для любой процедуры снижения размерности* $\Pi = \{q,\ \mathbf{C}_q,\ \mathbf{R}_q\}$ *при выбранном преобразовании восстановления* $\mathbf{R}_q$ *наилучшим преобразованием сжатия* $\mathbf{C}^*_q$ *является*

$$\mathbf{z}^*\,(\mathbf{x}) = \arg\min_{\mathbf{z}} \|\mathbf{v}\,(\mathbf{x}) - \mathbf{v}\,(\mathbf{R}_q\,(\mathbf{z}))\|_\theta, \qquad (17)$$

*и для величины ошибки* $\varepsilon_n\,(\Pi)$ (6) *имеет место неравенство* $\varepsilon_n\,(\Pi^*) \leqslant \varepsilon_n\,(\Pi)$, *где* $\Pi^* = \left\{q,\ \mathbf{C}^*_q,\ \mathbf{R}_q\right\}$ *— построенная мажорирующая процедура снижения размерности.*

Отсюда следует, что если $\mathbf{x}\,(b)$ является цифровым описанием объекта, то цифровое описание $\mathbf{R}_q\,(\mathbf{z}^*\,(\mathbf{x}\,(b)))$ модельного объекта $b^*$ получается путем проектирования обобщенного цифрового описания $\mathbf{v}\,(\mathbf{x}\,(b))$ (16) исходного объекта $b$ на многообразие $\mathcal{H}_q$ (17) (нахождения точки многообразия, являющейся ближайшей точкой к $\mathbf{v}\,(\mathbf{x}\,(b))$ в норме $\|\cdot\|_\theta$ (14)).

Задача функционального снижения размерности может быть переформулирована следующим образом. Рассмотрим выборку

$$V_n = \{\mathbf{v}_1,\ \mathbf{v}_2, \ldots, \mathbf{v}_n\}, \qquad (18)$$

состоящую из обобщенных описаний $\mathbf{v}_i = (\mathbf{x}_i : \mathbf{f}\,(\mathbf{x}_i))$, $\quad i = 1, 2, \ldots, n$. Точки $\mathbf{v} = (\mathbf{x} : \mathbf{f}) \in \mathbb{R}^p$ (16) лежат на известном $d$-мерном многообразии $\mathcal{V} = \left\{\mathbf{v} = (\mathbf{x} : \mathbf{f}\,(\mathbf{x})) \in \mathbb{R}^p : \mathbf{x} \in \mathbb{R}^d\right\}$, в котором мы хотим по данным $V_n$ (18) построить $q$-мерное подмногообразие $\mathcal{H}_q$ (15).

Рассмотрим сформулированную задачу функционального снижения размерности с топологической точки зрения. Для каждого значения $\mathbf{z} \in \mathbf{Z}$ рассмотрим в $\mathbb{R}^p$ $(p-q)$-мерное аффинное линейное подпространство $\mathrm{L}_q\,(\mathbf{z})$, проходящее через точку $\mathbf{v}\,(\mathbf{R}_q\,(\mathbf{z})) \in \mathcal{H}_q$ и ортогональное $q$-мерному касательному линейному подпространству $\mathrm{T}_q\,(\mathbf{z})$ к многообразию $\mathcal{H}_q$ (17) в точке $\mathbf{v}\,(\mathbf{R}_q\,(\mathbf{z}))$.

По построению, ошибка восстановления $\Delta_q\,(\mathbf{x}) = \mathbf{v}\,(\mathbf{x}) - \mathbf{v}\,(\mathbf{R}_q\,(\mathbf{z}^*\,(\mathbf{x})))$ принадлежит аффинному подпространству $L_q\,(\mathbf{z}^*\,(\mathbf{x}))$. Тем самым, многообразие $\mathcal{H}_q$ является гладким оснащенным многообразием.

Одним из наиболее известных методов решения задачи снижения размерности является Метод Главных Компонент (Principal Component Analysis), который решает задачу снижения размерности в классе линейных преобразований $\mathbf{C}_q$ и $\mathbf{R}_q$ (на самом деле достаточно требовать линейности только преобразования восстановления $\mathbf{R}_q$, а наилучшее преобразование сжатия $\mathbf{C}_q$, см. теорему 1, будет линейным по построению). Однако в случае нелинейной вектор-фукнции $\mathbf{f}\,(\mathbf{x})$ (12), многообразие $\mathcal{H}_q$ (15) будет заведомо нелинейным.

Тем не менее, большинство современных методов построения нелинейных аппроксимирующих многообразий основаны на исследования локальной структуры аппроксимирующего многообразия, которое в рассматриваемой точке $\mathbf{v}_0 \in \mathcal{H}_q$ локально аппроксимируется касательным линейным подпространством $\mathrm{T}_q\,(\mathbf{v}_0)$ размерности $q$.

Обозначим $\mathbf{V}(\mathbf{z})$ матрицу порядка $q \times p$, $j$-й столбец которой есть градиент $j$-й компоненты вектор-функции $\mathbf{v} = \mathbf{v}(\mathbf{R}_q(\mathbf{z}))$ (19) в точке $\mathbf{z}$, $j = 1, 2, \ldots, d$. Тогда

$$(\mathbf{V}(\mathbf{z}))^\top = \left((\mathbf{r}(\mathbf{z}))^\top \times \nabla \mathbf{f}(\mathbf{R}_q(\mathbf{z}))\right)^\top = A(\mathbf{x}) \times (\mathbf{r}(\mathbf{z}))^\top, \tag{19}$$

где

- $A(\mathbf{x})$ — матрица порядка $p \times d$, имеющая вид: $A(\mathbf{x}) = \begin{pmatrix} \mathbf{I}_d \\ \nabla^\top \mathbf{f}(\mathbf{x}) \end{pmatrix}, \quad \mathbf{x} = \mathbf{R}_q(\mathbf{z})$;
- $\nabla \mathbf{f}(\mathbf{x})$ — матрица порядка $p \times d$, $j$-й столбец которой есть градиент $j$-й компоненты $k$-мерной вектор-функции $\mathbf{f}(\mathbf{x})$ (12) в точке $\mathbf{x}$, $j = 1, 2, \ldots, k$;
- $\mathbf{r}(\mathbf{z})$ — матрица порядка $p \times d$, $j$-й столбец которой есть градиент $j$-й компоненты $q$-мерной вектор-функции $\mathbf{R}_q(\mathbf{z})$ в точке $\mathbf{z}$, $j = 1, 2, \ldots, q$.

Пусть точка $\mathbf{v}_0^* = (\mathbf{R}_q(\mathbf{z}_0) : \mathbf{f}(\mathbf{R}_q(\mathbf{z}_0))) \in \mathcal{H}_q$ соответствует значению $\mathbf{z}_0$ параметра $\mathbf{z}$, которым параметризовано многообразие $\mathcal{H}_q$. Тогда касательное линейное подпространство $\mathrm{T}_q(\mathbf{z}_0)$ размерности $q$ натянуто на векторы, являющиеся столбцами матрицы $(\mathbf{V}(\mathbf{z}_0))^\top$ (19).

Большинство современных методов построения нелинейных аппроксимирующих многообразий основаны на исследования локальной структуры аппроксимирующего многообразия, которое в рассматриваемой точке $\mathbf{v}_0^* \in \mathcal{H}_q$ локально аппроксимируется касательным линейным подпространством $\mathrm{T}_q(\mathbf{z}_0)$ размерности $q$.

Для исследования локальной структуры многообразия для каждой точки

$$\mathbf{v}_0 = (\mathbf{x}_0 : \mathbf{f}(\mathbf{x}_0)) \in V_n \tag{20}$$

рассматривается подвыборка $V_n(\mathbf{v}_0)$ выборки $V_n$, состоящая из $n(\mathbf{v}_0)$ элементов выборки $V_n$, ближайших к точке $\mathbf{v}_0$ (включая точку $\mathbf{v}_0$). Величина $n(\mathbf{v}_0)$ либо задана, либо равна числу элементов выборки $V_n$, находящихся в $\varepsilon$-окрестности точки $\mathbf{v}_0$ для заданного числа $\varepsilon$. Для удобства записи будем считать, что подвыборка $V_n(\mathbf{v})$ состоит из первых $n(\mathbf{v})$ элементов $\left\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n(\mathbf{v})}\right\}$ выборки $V_n$, при необходимости перенумеровав для этого элементы выборки $V_n$.

Пусть $\mathbf{v}_i^* = (\mathbf{R}_q(\mathbf{C}_q(\mathbf{x}_i)) : \mathbf{f}(\mathbf{R}_q(\mathbf{C}_q(\mathbf{x}_i))))$ — проекция точки $\mathbf{v}_i = (\mathbf{x}_i : \mathbf{f}(\mathbf{x}_i)) \in V_n(\mathbf{v}_0)$ на многообразие $\mathcal{H}_q$, и пусть $\mathbf{z}_i = \mathbf{C}_q(\mathbf{x}_i)$, $i = 1, 2, \ldots, n(\mathbf{v}_0)$, а $\mathbf{z}_0 = \mathbf{C}_q(\mathbf{x}_0)$ для компоненты $\mathbf{x}$ точки $\mathbf{v}_0$ (20) (напомним, что $\mathbf{v}_0$ — одна из точек подвыборки $V_n(\mathbf{v}_0)$). Тогда из определения матрицы $\mathbf{V}(\mathbf{z})$ следует, что:

$$\begin{aligned} &\mathbf{v}_i^* - \mathbf{v}_0^* = (\mathbf{V}(\mathbf{z}_0))^\top \times (\mathbf{z}_i - \mathbf{z}_0) + \mathrm{o}(\|\mathbf{z}_i - \mathbf{z}_0\|) = \\ &= A(\mathbf{x}_0) \times (\mathbf{r}(\mathbf{z}_0))^\top \times (\mathbf{z}_i - \mathbf{z}_0) + \mathrm{o}(\|\mathbf{z}_i - \mathbf{z}_0\|) = \mathbf{v}_i^{**} - \mathbf{v}_0^* + \mathrm{o}(\|\mathbf{z}_i - \mathbf{z}_0\|), \end{aligned} \tag{21}$$

где $\mathbf{v}_i^{**} = \mathbf{v}_0^* + A(\mathbf{x}_0) \times (\mathbf{r}(\mathbf{z}_0))^\top \times (\mathbf{z}_i - \mathbf{z}_0), \quad i = 1, 2, \ldots, n(\mathbf{v}_0)$, и нетрудно видеть, что точки $\mathbf{v}_i^{**}$, $i = 1, 2, \ldots, n(\mathbf{v}_0)$, лежат в $q$-мерном касательном линейном подпространстве $\mathrm{T}_q(\mathbf{z}_0)$. Если для рассматриваемой процедуры снижения размерности процедура сжатия имеет вид (17), то точки $\mathbf{v}_i^{**}$ являются проекциями (в норме $\|\cdot\|_\theta$ (14)) точек $\mathbf{v}_i$ на касательное линейное подпространство $\mathrm{T}_q(\mathbf{z}_0)$: $\mathbf{v}_i^{**} = \mathbf{v}_0^* + P(\mathbf{z}_0) \times (\mathbf{v}_i - \mathbf{v}_0^*)$, здесь $P(\mathbf{z}_0)$ — линейный оператор проектирования на аффинное линейное подпространство $\mathrm{T}_q(\mathbf{z}_0)$.

Если подвыборка $V_n(\mathbf{v}_0)$ взята из достаточно малой окрестности точки $\mathbf{v}_0$ (20), то касательное линейное подпространство $\mathrm{T}_q(\mathbf{z}_0)$ может быть аппроксимировано аффинным линейным подпространством $\mathrm{L}(\mathbf{v})$, средняя сумма квадратов расстояний до которого точек подвыборки $V_n(\mathbf{v})$ минимальна (в норме $\|\cdot\|_{\boldsymbol{\theta}}$ (14)).

Из определения матрицы $\mathbf{r}(\mathbf{z})$ следует, что: $\mathbf{x}_i^* - \mathbf{x}_0^* = (\mathbf{r}(\mathbf{z}_0))^\top \times (\mathbf{z}_i - \mathbf{z}_0) + \mathrm{o}(\|\mathbf{z}_i - \mathbf{z}_0\|)$, и, следовательно, $\mathbf{v}_i^{**} - \mathbf{v}_0^* = A(\mathbf{x}_0) \times (P(\mathbf{z}_0) \times (\mathbf{x}_i - \mathbf{x}_0)) + \mathrm{o}(\|\mathbf{z}_i - \mathbf{z}_0\|)$.

Для элементов $\mathbf{v}_i = (\mathbf{x}_i : \mathbf{f}(\mathbf{x}_i)) \in V_n(\mathbf{v}_0)$, $i = 1, 2, \ldots, n(\mathbf{v}_0)$, имеет место соотношения:

$$\mathbf{f}(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_0) = (\nabla \mathbf{f}(\mathbf{x}_0))^\top \times (\mathbf{x}_i - \mathbf{x}_0) + \mathrm{o}(\|\mathbf{x}_i - \mathbf{x}_0\|), \tag{22}$$

$$\mathbf{f}(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_i^*) \approx (\nabla \mathbf{f}(\mathbf{x}_0))^\top \times (\mathbf{x}_i - \mathbf{x}_i^*) + \mathrm{o}(\|\mathbf{x}_i - \mathbf{x}_0\|), \tag{23}$$

и, следовательно, $\mathbf{v}_i - \mathbf{v}_i^* = A(\mathbf{x}_0) \times (\mathbf{x}_i - \mathbf{x}_i^*) + \mathrm{o}(\|\mathbf{x}_i - \mathbf{x}_0\|)$.

Поэтому имеет место приближенное равенство:

$$\begin{aligned} &\|\mathbf{v}_i - \mathbf{v}_i^*\|_\theta^2 = \theta_0^2 \|\mathbf{x}_i - \mathbf{R}_q(\mathbf{C}_q(\mathbf{x}_i))\|^2 + \sum_{j=1}^{k} \theta_j^2 |f_j(\mathbf{x}_i) - f_j(\mathbf{R}_q(\mathbf{C}_q(\mathbf{x}_i)))|^2 = \\ &= \theta_0^2 \|\mathbf{x}_i - \mathbf{x}_i^*\|^2 + \sum_{j=1}^{k} \theta_j^2 |f_j(\mathbf{x}_i) - f_j(\mathbf{x}_i^*)|^2 = \\ &= \theta_0^2 \|\mathbf{x}_i - \mathbf{x}_i^*\|^2 + \left\| \left(\mathrm{diag}\left(\theta_1^2, \theta_2^2, \ldots, \theta_k^2\right)\right)^\top \times (\mathbf{f}(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_i^*)) \right\|^2 \approx \|A_\theta(\mathbf{x}_0) \times (\mathbf{x}_i - \mathbf{x}_i^*)\|^2, \end{aligned} \tag{24}$$

где матрица $A_\theta(\mathbf{x}) \in \mathbb{R}^{p \times d}$ имеет блочный вид:

$$A_\theta(\mathbf{x}) = \begin{pmatrix} \theta_0^2 \mathbf{I}_d \\ \mathrm{diag}\left(\theta_1^2,\ \theta_2^2, \ldots, \theta_k^2\right) \times \nabla^\top \mathbf{f}(\mathbf{x}) \end{pmatrix},$$

$\mathbf{I}_d$ — единичная матрица размера $d$, а $\mathrm{diag}\left(\theta_1^2,\ \theta_2^2, \ldots, \theta_k^2\right)$ — диагональная матрица размера $k$, на диагонали которой стоят веса $\theta_1^2,\ \theta_2^2, \ldots, \theta_k^2$.

Поэтому задача нахождения линейного аффинного подпространства средняя сумма квадратов расстояний до которого точек подвыборки $V_n(\mathbf{v})$ минимальна (в норме $\|\cdot\|_\theta$ (14)), сводится к оптимизационной задаче в пространстве $\mathbb{R}^d$, являющейся обобщенной задачей анализа главных компонент и сформулированной в следующем разделе.

## 7. ОБОБЩЕННАЯ ЗАДАЧА АНАЛИЗА ГЛАВНЫХ КОМПОНЕНТ

Пусть L — аффинное подпространство размерности $q < d$, имеющее вид

$$\mathrm{L} = \left\{ \mathbf{x} \in \mathbb{R}^d : \mathbf{x} = \mathbf{x}_0 + \sum_{j=1}^{q} z_j \mathbf{e}_j, z_1, z_2, \ldots, z_q \in \mathbb{R}^1 \right\},$$

где $\mathbf{x}_0 \in \mathbb{R}^d$, а $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_q \in \mathbb{R}^d$ – ортонормированные векторы. Обозначим $\mathbf{C}_L$ матрицу размера $d \times q$, столбцы которой составляют векторы $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_q$.

Пусть $\mathbf{B}$ — матрица размера $q \times d$, и $\mathbf{B}_L : \mathbf{x} \to \mathbf{x}_0 + \mathbf{C}_L \times \mathbf{B} \times (\mathbf{x} - \mathbf{x}_0)$ аффинное линейное преобразование, переводящее произвольный вектор $\mathbf{x} \in \mathbb{R}^d$ в точку, принадлежащую подпространству L.

Пусть $\mathbf{x}_1,\ \mathbf{x}_2, \ldots, \mathbf{x}_n$ — заданное множество $d$-мерных векторов, и $\mathbf{A}$ — заданная матрица размера $p \times d$, $p \geqslant d$. Ищутся аффинное подпространство L и аффинное линейное преобразование $\mathbf{B}_L$, минимизирующие величину

$$E = E(\mathbf{x}_0,\ \mathbf{C}_L,\ \mathbf{B}) = \frac{1}{n} \sum_{i=1}^{n} \|\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{B}_L \mathbf{x}_i\|^2. \tag{25}$$

Если $\mathbf{A}$ — единичная матрица размера $d \times d$, то решение этой задачи дается обычным методом главных компонент:

- аффинное подпространство $\mathrm{L} = \left\{ \mathbf{x} \in \mathbb{R}^d : \ \mathbf{x} = \overline{\mathbf{x}}_n + \sum\limits_{j=1}^{q} z_j \mathbf{e}_{jx}, \ z_1, z_2, \ldots, z_q \in \mathbb{R}^1 \right\}$, где $\overline{\mathbf{x}}_n = \frac{1}{n} \sum\limits_{i=1}^{n} \mathbf{x}_i$, а $\mathbf{e}_{1x}, \ \mathbf{e}_{2x}, \ldots, \mathbf{e}_{qx} \in \mathbb{R}^d$ — ортонормированные собственные векторы выборочной ковариационной матрицы $\mathbf{S}_x = \frac{1}{n} \sum\limits_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}}_n) \times (\mathbf{x}_i - \overline{\mathbf{x}}_n)^\top$, соответствующие $q$ наибольшим собственным числам матрицы $\mathbf{S}_x$;
- оператор $\mathbf{B}_L$ есть оператор проектирования на L, и строки матрицы $\mathbf{B}$ размера $q \times d$ состоят из транспонированных собственных векторов $\mathbf{e}_{1x}, \ \mathbf{e}_{2x}, \ldots, \mathbf{e}_{qx}$.

Пусть $\mathbf{S}_A = \mathbf{A}^\top \mathbf{A}$, $\mathbf{C}_{0L}$ — матрица размера $d \times q$, столбцы которой состоят из первых $q$ ортонормированных собственных векторов $\{\mathbf{e}_{1x,A}, \ \mathbf{e}_{2x,A}, \ldots, \mathbf{e}_{qx,A}\} \in \mathbb{R}^d$ матрицы

$$\mathbf{S}_{x,A} = \mathbf{S}_x \times \mathbf{S}_A, \tag{26}$$

записанных в порядке убывания соответствующих собственных чисел этой матрицы, и $\mathbf{B}_0 = (\mathbf{C}_{0L})^\top \times \mathbf{S}_A$ матрица размера $q \times d$.

**Теорема 2.** *Пусть матрица $\mathbf{A}$ имеет ранг $d$, тогда вектор $\overline{\mathbf{x}}_n$, $q$-мерное аффинное подпространство*

$$\mathrm{L}_0 = \left\{ \mathbf{x} \in \mathbb{R}^d : \ \mathbf{x} = \overline{\mathbf{x}}_n + \sum_{j=1}^{q} z_j \mathbf{e}_{jx,A}, \ z_1, z_2, \ldots, z_q \in \mathbb{R}^1 \right\}$$

*и аффинное линейное преобразование $\mathbf{x} \to \mathbf{x}_n + \mathbf{C}_{0,L(q)} \times \mathbf{B}_0 \times (\mathbf{x} - \overline{\mathbf{x}}_n)$, переводящее произвольный вектор $\mathbf{x} \in \mathbb{R}^d$ в точку, принадлежащую подпространству $\mathrm{L}_0$, минимизируют по $\mathbf{x}_0$, $\mathbf{C}_L$ и $\mathbf{B}$ величину*

$$E = E(\mathbf{x}_0, \ \mathbf{C}_L, \ \mathbf{B}) = \frac{1}{n} \sum_{i=1}^{n} \|\mathbf{A}\mathbf{x}_i - \mathbf{A}(\mathbf{x}_0 + \mathbf{C}_L \times \mathbf{B} \times (\mathbf{x}_i - \mathbf{x}_0))\|^2.$$

**Доказательство.** Запишем выражение (25) в виде:

$$\begin{aligned} E(\mathbf{x}_0, \ \mathbf{C}_L, \ \mathbf{B}) &= \frac{1}{n} \sum_{i=1}^{n} \|\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{B}_L\mathbf{x}_i\|^2 = \frac{1}{n} \sum_{i=1}^{n} \Big\{ (\mathbf{x}_i - \mathbf{B}_L\mathbf{x}_i)^\top \times \mathbf{S}_A \times (\mathbf{x}_i - \mathbf{B}_L\mathbf{x}_i) \Big\} = \\ &= \frac{1}{n} \sum_{i=1}^{n} \Big\{ (\mathbf{x}_i - \mathbf{x}_0)^\top (\mathbf{I} - \mathbf{C}_L \times \mathbf{B})^\top \times \mathbf{S}_A \times (\mathbf{I} - \mathbf{C}_L \times \mathbf{B}) \times (\mathbf{x}_i - \mathbf{x}_0) \Big\}, \end{aligned} \tag{27}$$

где $\mathbf{S}_A = \mathbf{A}^\top \mathbf{A}$. Будем последовательно минимизировать $E(\mathbf{x}_0, \ \mathbf{C}_L, \ \mathbf{B})$ по $\mathbf{x}_0$, $\mathbf{C}_L$ и $\mathbf{B}$.

Величина (27) представляет собой неотрицательно определенную квадратическую форму от компонент вектора $\mathbf{x}_0$, и, следовательно, имеет единственный минимум, являющийся решением

$$\frac{\partial E(\mathbf{x}_0, \ \mathbf{C}_L, \ \mathbf{B})}{\partial \mathbf{x}_0} = 0.$$

Так как

$$\frac{\partial E(\mathbf{x}_0, \mathbf{C}_L, \mathbf{B})}{\partial \mathbf{x}_0} = 2 \times (\mathbf{I} - \mathbf{C}_L \times \mathbf{B})^\top \times \mathbf{S}_A \times (\mathbf{I} - \mathbf{C}_L \times \mathbf{B}) \times (\overline{\mathbf{x}}_n - \mathbf{x}_0),$$

то единственный минимум функции $E$ достигается в точке $\mathbf{x}_0 = \overline{\mathbf{x}}_n$.

Матрица $\mathbf{S}_A$ является неотрицательно определенной, и пусть $\lambda_{1A} \geqslant \lambda_{2A} \geqslant \ldots \geqslant \lambda_{dA} > 0$ ее неотрицательные собственные числа, записанные в порядке убывания. Для простоты будем считать, что все собственные числа различны, рассмотрение общего случая требует дополнительных несложных технических модификаций.

Обозначим $\mathbf{e}_{1A},\ \mathbf{e}_{2A}, \ldots, \mathbf{e}_{dA} \in \mathbb{R}^d$ — ортонормированные собственные векторы, соответствующие указанным собственным числам. Обозначим $\mathbf{C}_A$ квадратную матрицу размера $d \times d$, столбцы которой составляют векторы $\mathbf{e}_{1A},\ \mathbf{e}_{2A}, \ldots, \mathbf{e}_{dA}$. Очевидно, что матрица $\mathbf{C}_A$ является ортогональной:

$$(\mathbf{C}_A)^\top \times \mathbf{C}_A = \mathbf{I}_d. \tag{28}$$

Имеем:

$$\mathbf{S}_A = \mathbf{C}_A \times \operatorname{diag}(\lambda_{1A},\ \lambda_{2A}, \ldots, \lambda_{dA}) \times (\mathbf{C}_A)^\top = \mathbf{U}_A \times (\mathbf{U}_A)^\top, \tag{29}$$

где

$$\mathbf{U}_A = \mathbf{C}_A \times (\operatorname{diag}(\lambda_{1A},\ \lambda_{2A}, \ldots, \lambda_{qA}))^{1/2}. \tag{30}$$

Итак,

$$E(\overline{\mathbf{x}}_n,\ \mathbf{C}_L,\ \mathbf{B}) = \frac{1}{n}\sum_{i=1}^{n}\Big\{(\mathbf{z}_i - \mathbf{R}_q\mathbf{T}_q\mathbf{x}_i)^\top \times (\mathbf{x}_i - \mathbf{R}_q\mathbf{T}_q\mathbf{x}_i)\Big\},$$

где $\mathbf{z}_i = (\mathbf{U}_A)^\top(\mathbf{x}_i - \overline{\mathbf{x}}_n)\,, \quad i = 1, 2, \ldots, n,$

$$\mathbf{R}_q = (\mathbf{U}_A)^\top \times \mathbf{C}_L, \tag{31}$$

$$\mathbf{T}_q = \mathbf{B} \times \Big((\mathbf{U}_A)^\top\Big)^{-1}, \tag{32}$$

и очевидно, что $\frac{1}{n}\sum\limits_{i=1}^{n}\mathbf{z}_i = 0$. Обозначим

$$\mathbf{T}_{0q} = \Big((\mathbf{R}_q)^\top \times (\mathbf{R}_q)^{-1}\Big) \times (\mathbf{R}_q)^\top. \tag{33}$$

Имеем:

$$\begin{gathered}E(\mathbf{x}_n,\ \mathbf{C}_L,\ \mathbf{B}) = \frac{1}{n}\sum_{i=1}^{n}\Big\{(\mathbf{z}_i - \mathbf{R}_q\mathbf{T}_{0q}\mathbf{z}_i)^\top \times (\mathbf{z}_i - \mathbf{R}_q\mathbf{T}_{0q}\mathbf{z}_i)\Big\} + \\ + \frac{1}{n}\sum_{i=1}^{n}\Big\{(\mathbf{R}_q(\mathbf{T}_q - \mathbf{T}_{0q})\mathbf{T}_q\mathbf{z}_i)^\top \times (\mathbf{R}_q(\mathbf{T}_q - \mathbf{T}_{0q})\mathbf{T}_q\mathbf{z}_i)\Big\}.\end{gathered} \tag{34}$$

Так как второе слагаемое в (34) неотрицательно, то величина $E(\overline{\mathbf{x}}_n,\ \mathbf{C}_L,\ \mathbf{B})$ будет минимальна при $\mathbf{T}_q = \mathbf{T}_{0q}$, и из (32) следует, что она будет минимальна при

$$\mathbf{B} = \mathbf{B}_0 = \mathbf{B}_0(\mathbf{C}_L) = \mathbf{T}_{0q} \times (\mathbf{U}_A)^\top. \tag{35}$$

Обозначим $\mathbf{e}_{1A,L},\ \mathbf{e}_{2A,L}, \ldots, \mathbf{e}_{qA,L} \in \mathbb{R}^d$ столбцы матрицы $\mathbf{R}_q$, и пусть $\mathrm{L}(q, A)$ — $q$-мерное линейное подпространство, натянутое на векторы $\mathbf{e}_{1A,L},\ \mathbf{e}_{2A,L}, \ldots, \mathbf{e}_{qA,L}$.

Нетрудно видеть, что вектор $\mathbf{R}_q \times \mathbf{T}_{0q} \times \mathbf{z}$ есть проекция вектора $\mathbf{z} \in \mathbb{R}^d$ на линейное подпространство $\mathrm{L}(q, A)$, и задача минимизации величины $E(\overline{\mathbf{x}}_n,\ \mathbf{C}_L,\ \mathbf{B}_0(\mathbf{C}_L))$ по $\mathbf{C}_L$ есть задача нахождения $q$ -мерного линейного подпространства $\mathrm{L}(q, A)$, минимизирующего сумму квадратов расстояний векторов $\mathbf{z}_1,\ \mathbf{z}_2, \ldots, \mathbf{z}_n$ до их проекций на $\mathrm{L}(q, A)$.

Решение этой оптимизационной задачи дается методом главных компонент, и столбцы оптимальной матрицы $\mathbf{R}_{0q}$ состоят из векторов $\mathbf{e}_{01A,L},\ \mathbf{e}_{02A,L}, \ldots, \mathbf{e}_{0qA,L}$, являющихся ортонормированными собственными векторами матрицы

$$\mathbf{S}_z = \frac{1}{n}\sum_{i=1}^{n} (\mathbf{z}_i) \times (\mathbf{z}_i)^\top,$$

отвечающим $q$ наибольшим корням (собственным числам) $\lambda_{01A,L},\ \lambda_{02A,L}, \ldots, \lambda_{0qA,L}$ уравнения $\operatorname{diag}(\mathbf{S}_z - \lambda \times \mathbf{I}_d) = 0$.

Из определения (30) получаем

$$\begin{aligned}\mathbf{S}_z &= \frac{1}{n}\sum_{i=1}^{n} \Big[(\mathbf{z}_i) \times (\mathbf{z}_i)^\top\Big] = \frac{1}{n}\sum_{i=1}^{n} \Big[\mathbf{U}_A^\top (\mathbf{x}_i - \overline{\mathbf{x}}_n) \times \Big(\mathbf{U}_A^\top (\mathbf{x}_i - \overline{\mathbf{x}}_n)\Big)^\top\Big] = \\ &= \mathbf{U}_A^\top \times \frac{1}{n}\sum_{i=1}^{n} \Big[(\mathbf{x}_i - \overline{\mathbf{x}}_n) \times (\mathbf{x}_i - \overline{\mathbf{x}}_n)^\top\Big] \times \mathbf{U}_A = (\mathbf{U}_A)^\top \times \mathbf{S}_x \times \mathbf{U}_A.\end{aligned} \tag{36}$$

Имеем:

$$\begin{aligned}\mathbf{S}_z - \lambda \mathbf{I}_d &= (\mathbf{U}_A)^\top \mathbf{S}_x \mathbf{U}_A - \lambda \mathbf{I}_d = (\mathbf{U}_A)^\top \Big\{\mathbf{S}_x \mathbf{U}_A - \lambda \Big((\mathbf{U}_A)^\top\Big)^{-1}\Big\} = \\ &= (\mathbf{U}_A)^\top \Big\{\mathbf{S}_x - \lambda \Big((\mathbf{U}_A)^\top\Big)^{-1} (\mathbf{U}_A)^{-1}\Big\} \mathbf{U}_A = (\mathbf{U}_A)^\top \Big\{\mathbf{S}_x - \lambda (\mathbf{S}_A)^{-1}\Big\} \mathbf{U}_A.\end{aligned} \tag{37}$$

С учетом соотношений (28) – (30) и обозначения (26) отсюда следует, что

$$\det(\mathbf{S}_z - \lambda) = \det\Big(\mathbf{S}_x - \lambda \times (\mathbf{S}_A)^{-1}\Big) = \det(\mathbf{S}_{x,A} - \lambda) \times \det(\mathbf{S}_A)^{-1}.$$

Следовательно, у матрицы $\mathbf{S}_z$ собственные числа $\lambda_{1A,L},\ \lambda_{2A,L}, \ldots, \lambda_{dA,L}$ совпадают с собственными числами матрицы $\mathbf{S}_{x,A}$ и с собственными числами пучка матриц $\Big\{\mathbf{S}_x,\ (\mathbf{S}_A)^{-1}\Big\}$.

Из определения матрицы $\mathbf{R}_{0q}$ имеем

$$\Big((\mathbf{U}_A)^\top \times \mathbf{S}_x \times \mathbf{U}_A\Big) \times \mathbf{R}_{0q} = \mathbf{R}_{0q} \times \operatorname{diag}(\lambda_{01A,L},\ \lambda_{02A,L}, \ldots, \lambda_{0qA,L}), \tag{38}$$

$$(\mathbf{R}_{0q})^\top \times \mathbf{R}_{0q} = \mathbf{I}_q. \tag{39}$$

Из соотношения (31) определим $\mathbf{C}_{0L} = \Big((\mathbf{U}_A)^\top\Big)^{-1} \times \mathbf{R}_{0q}$, тогда из (38) получаем

$$(\mathbf{U}_A)^\top \times \mathbf{S}_x \times \mathbf{U}_A \times (\mathbf{U}_A)^\top \times \mathbf{C}_{0L} = (\mathbf{U}_A)^\top \times \mathbf{C}_{0L} \times \operatorname{diag}(\lambda_{01A,L},\ \lambda_{02A,L}, \ldots, \lambda_{0qA,L}),$$

и, следовательно, с учетом (29) получаем: $\mathbf{S}_{x,A} \times \mathbf{C}_{0L} = \mathbf{C}_{0L} \times \operatorname{diag}(\lambda_{01A,L},\ \lambda_{02A,L}, \ldots, \lambda_{0qA,L})$.

Следовательно, матрица $\mathbf{C}_{0L}$ состоит из $q$ собственных векторов матрицы $\mathbf{S}_{x,A}$, записанных в порядке убывания соответствующих собственных чисел этой матрицы.

Из соотношений (32), (33), (35) и (39) следует, что оптимальная матрица $\mathbf{B}_0 = \mathbf{B}_0(\mathbf{C}_{0L})$ равна: $\mathbf{B}_0 = \mathbf{T}_{0q} \times (\mathbf{U}_A)^\top = (\mathbf{R}_{0q})^\top \times (\mathbf{U}_A)^\top = (\mathbf{C}_{0L})^\top \times \mathbf{U}_A \times (\mathbf{U}_A)^\top = (\mathbf{C}_{0L})^\top \times \mathbf{S}_A$. Теорема 2 доказана.

## 8. ЗАКЛЮЧЕНИЕ

В работе рассмотрена задача снижения размерности в метамоделировании. Показано, что решение стандартной задачи снижения размерности (построения аппроксимирующего многообразия для носителя аргумента неизвестной зависимости) в общем виде не гарантирует качества метамодели, построенной по редуцированным данным. Сформулирована задача функционального снижения размерности, решение которой позволяет строить по данным аппроксимирующие многообразия, учитывающие особенности дальнейшего использования этих многообразий при построении метамоделей. Показано, что наилучшим преобразованием сжатия цифрового описания объекта в рассматриваемой задаче функционального снижения размерности является проектирование введенного обобщенного цифрового описания объекта на введенное расширенное многообразие. Для построения по данным касательного аффинного линейного подпространства сформулирована и решена обобщенная задача построения главных компонент. Полученные результаты могут быть использованы при построении алгоритмов функционального снижения размерности.

## СПИСОК ЛИТЕРАТУРЫ

1. A. Kuleshov and A. Bernstein. Cognitive technologies in adaptive models of complex plants. *IFAC Proceedings Volumes*, 42(4):1441 – 1452, 2009.

2. Evgeny Burnaev. Algorithmic foundations of predictive analytics in industrial engineering design. *J. Commun. Technol. Electron.*, 64, 2019.

3. A. Bernstein and A. Kuleshov. Manifold learning: Generalization ability and tangent proximity. *Int. J. Software and Informatics*, 7(3):359–390, 2013.

4. Mikhail Belyaev, Evgeny Burnaev, Ermek Kapushev, and et al. Gtapprox: surrogate modeling for industrial design. *Advances in Engineering Software*, 102, 2016.

5. A. Kuleshov, A. Bernstein, and E. Burnaev. Kernel regression on manifold valued data. In *Proc. of IEEE 5th Int. Conf. on Data Science and Advanced Analytics*, pages 120–129, 2018.

6. A. Kuleshov, A. Bernstein, and E. Burnaev. Manifold learning regression with non-stationary kernels. In *Artificial Neural Networks in Pattern Recognition*, pages 152–164, Cham, 2018. Springer.

7. H. Tyagi, E. Vural, and P. Frossard. Tangent space estimation for smooth embeddings of riemannian manifold. *CoRR*, abs/1208.1065v2, 2013.

8. Yanyuan Ma and Liping Zhu. Efficient estimation in sufficient dimension reduction. *Ann. Statist.*, 41(1):250–268, 02 2013.

# Functional Dimension Reduction in Predictive Modeling

**Burnaev E.V., Bernstein A.V.**

The paper considers the problem of dimension reduction in metamodeling. It is shown that to construct an exact metamodel based on reduced data, it is necessary to solve the problem of dimension reduction in a non-standard formulation, namely, taking into account additional functional constraints. To construct a tangent affine linear subspace from the data, required for solving the functional dimension reduction problem, a generalized principal component construction problem is formulated and solved.

**KEYWORDS:** manifold learning, dimension reduction, functional dimension reduction, principal components.