

Состоятельность оценки области определения алгоритмом спектральных вложений Грассмана-Штифеля

Янович Юрий

*Институт проблем передачи информации
им. А. А. Харкевича РАН, Москва
Национальный исследовательский университет
«Высшая школа экономики», Москва
yuanovich@hse.ru*

Аннотация

В машинном обучении при построении регрессионных зависимостей или решении задач классификации многомерные описания объектов часто являются избыточными и функционально зависимыми. Такие описания зачастую лежат около многообразий существенно меньшей размерности, чем размерность их первичной записи. Данное предположение называется гипотезой многообразия (Manifold Hypothesis). Использование такой информации может помочь в решении исходной задачи. Так возникает задача оценивания многообразий.

В последние годы был разработан ряд подходов, таких как изометрическое отображение (Isomap), локально-линейное вложение (LLE), выравнивания локальных касательных пространств (LTSA) и спектральных вложений Грассмана-Штифеля (GSE), для решения данной задачи.

В работе рассмотрена общая схема подобных алгоритмов. Одним из их шагов является расширение отображения сжатия на точки вне заданной выборки. Однако, для таких отображений необходимо задавать область определения, так как многообразии неизвестно. Основной результат работы состоит в доказательстве того, что область определения алгоритма Грассмана-Штифеля сходится по расстоянию Хаусдорфа по вероятности к неизвестному многообразию.

1. Введение

Многие задачи анализа данных, такие как распознавание образов, классификация, кластеризация, прогнозирование, восстановление регрессии и другие, связаны с реальными данными, которые лежат в пространствах высокой размерности, и “проклятие

размерности” часто является препятствием для использования алгоритмов машинного обучения при решении таких задач.

К счастью, во многих приложениях, реальные многомерные данные заполняют лишь очень малую часть многомерного пространства наблюдений \mathbb{R}^p , чья внутренняя размерность q мала (обычно, $q \ll p$) [1, 2]. Так, множество алгоритмов снижения размерности (извлечения признаков), чьей задачей является нахождение низкоразмерной параметризации многомерных данных, могут быть использованы для сведения таких “многомерных” задач к низкоразмерным с сохранением их свойств [3, 4]. После этого маломерные признаки, избавленные от “проклятия размерности”, могут быть использованы для процедур обучения вместо исходных многомерных векторов [5]: “снижение размерности может быть важно для отказа от избыточности и снижения вычислительных затрат на последующие операции” [6].

Наиболее популярной моделью многомерных данных, занимающей малую часть пространства наблюдений \mathbb{R}^p , является модель многообразия, в соответствии с которой данные лежат на или вблизи неизвестного многообразия (многообразия данных, МД) \mathcal{X} меньшей размерности $q < p$, вложенного в многомерное пространство входов \mathbb{R}^p (гипотеза многообразия [7] о многомерных данных). Обычно это предположение выполнено для “реальных” многомерных данных, полученных из “естественных” источников. В реальных примерах, размерность многообразия q обычно неизвестна и может быть оценена с помощью выборки, случайно сгенерированной на многообразии данных [8, 9, 10, 11].

Снижение размерности при условии гипотезы о многообразии для обрабатываемых данных обычно называется обучением на многообразиях [12, 13]. Задачей оценивания многообразий является построение маломерной параметризации МД (глобальных низкоразмерных координат на МД) по конечной вы-

борке из МД.

Гипотеза о многообразии означает, что малая окрестность каждой точки многообразия эквивалентна окрестности в низкоразмерном евклидовом пространстве. Поэтому большинство алгоритмов оценивания многообразий включают в себя две части: “локальную”, в которой оцениваются некоторые локальные характеристики каждой точки по её малой окрестности; и “глобальную” часть, в которой решаются выпуклые оптимизационные задачи специального вида и строятся маломерные координаты на МД (обычно, обобщенные задачи на собственные значения). Примерами таких алгоритмов служат локально-линейные вложения (LLE) [14], изометрические вложения (Isomap) [15], спектральные вложения Лапласа (LE) [16], локальные вложения касательных пространств (LTSA) [17], спектральные вложения гессиана [18], полуопределенные вложения (SDE) [19], диффузионные вложения [20]. Такие методы обучения представляют собой двухшаговые процедуры: сначала оцениваются локальные характеристики, а затем они объединяются и согласуются для построения глобального решения [14, 21, 22, 23, 24, 25].

Однако, необходимо определять подмножество многомерного пространства, на котором определены оцениваемые величины. При этом от такого подмножества требуется, чтобы с ростом выборки оно включало все многообразие и, по возможности, не включало далекие от многообразия точки.

В данной статье изучается область определения алгоритма спектральных вложений Лапласа (GSE, [22]) и доказывается, что расстояние Хаусдорфа между областью определения алгоритма и многообразием с большой вероятностью стремится к нулю.

Статья устроена следующим образом:

- в разделе 2 описана задача оценивания многообразий и типичная схема алгоритмов, решающих эту задачу;
- в разделе 3 сформулирован и доказан главный результат статьи — состоятельность оценки области определения алгоритма;
- в разделе приложения А перечислены предположения, в рамках которых производится доказательство.

2. Общая схема алгоритмов обучения на многообразиях

2.1. Оценивание многообразий как задача вложения многообразий

Рассмотрим неизвестное q -мерное многообразие данных

$$\mathbb{X} = \{X = f(b) \in \mathbb{R}^p : b \in \mathbb{B} \subset \mathbb{R}^q\}, \quad (1)$$

покрытое одной картой (\mathbb{B}, f) и вложенное в p -мерное пространство \mathbb{R}^p , $q < p$. Отображение f является взаимнооднозначным из открытого ограниченного множества $\mathbb{B} \subset \mathbb{R}^q$ на многообразии $\mathbb{X} = f(\mathbb{B})$ с гладким обратным отображением $f^{-1} : \mathbb{X} \rightarrow \mathbb{B}$. Внутренняя размерность многообразия q также является неизвестной.

Предполагается, что для МД \mathbb{X} существует положительное число обусловленности $c(\mathbb{X})$, то есть для каждой точки $X \in \mathbb{R}^p$, удаленной от \mathbb{X} не более чем на $1/c(\mathbb{X})$, существует единственная проекция на \mathbb{X} [26].

Существует обратная функция $h_f(X) = f^{-1}(X)$, чьи значения $b = h_f(X) \in \mathbb{R}^q$, можно рассматривать как низкоразмерные координаты на многообразии \mathbb{X} , задающие низкоразмерные представления (признаки) $b = h_f(X)$ многомерных данных X с многообразия.

Если отображения $h_f(X)$ и $f(b)$ гладкие, и матрица Якоби $J_f(b)$ размера $p \times q$ отображения $f(b)$ ранга q , тогда q -мерное линейное пространство

$$L(X) = \text{Span}(J_f(h_f(X))) \quad (2)$$

пространства \mathbb{R}^p является касательным пространством многообразия \mathbb{X} в точке $X \in \mathbb{X}$. Здесь и далее $\text{Span}(H)$ — это линейное пространство, построенное на столбцах матрицы H .

Пусть $\mathbb{X}_N = \{X_1, \dots, X_N\} \subset \mathbb{X}$ — случайная выборка на многообразии, сгенерированная из какой-то (неизвестной) вероятностной меры, чей носитель совпадает с \mathbb{X} . Общая задача обучения на многообразиях ставится следующим образом [13]: по заданной выборке \mathbb{X}_N построить низкоразмерную параметризацию МД, которая задает отображение

$$H : \mathbb{X} \subset \mathbb{R}^p \rightarrow \mathbb{Y}_h = h(\mathbb{X}) \subset \mathbb{R}^q \quad (3)$$

из МД \mathbb{X} в пространство признаков $\mathbb{Y}_h \subset \mathbb{R}^q$, $q < p$, которое сохраняет свойства МД.

2.2. Типичная схема алгоритмов вложения многообразий

Следуя [27], рассмотрим класс типичных алгоритмов обучения на многообразиях, который восстанавливает внутреннюю структуру МД по выборке;

этот класс включает так называемые алгоритмы с “нормированными входами” [27]. Общая схема рассматриваемых алгоритмов состоит из четырех шагов.

2.3. Шаг первый: построение окрестности

Для каждой точки выборки X_n ищутся окрестности из соседей $U_N(X_n) = \{X_n, X_{n,1}, \dots, X_{n,k(n)}\} \subset \mathbb{X}_N$, то есть подмножество элементов выборки, состоящее из близких к X_n точек. Например, $U_N(X_n) = U_N(X_n, \varepsilon)$ состоит из точек \mathbb{X}_N , попавших в шар радиуса ε с центром в X_n в пространстве \mathbb{R}^p , или $U_N(X_n, k)$ состоит из $k(n) = k$ ближайших к X_n элементов \mathbb{X}_N .

Построенные окрестности определяют граф выборки $\Gamma(\mathbb{X}_N)$, состоящий из вершин $\{X_1, \dots, X_N\}$, и между вершинами X_i и X_j есть ребро тогда и только тогда, когда $X_i \in U_N(X_j)$ или $X_j \in U_N(X_i)$.

2.4. Шаг второй: описание окрестностей

Вычисляются описания выбранных локальных свойств многообразия. Примерами таких описаний служат:

- барицентрические координаты $\{w_{n,1}, \dots, w_{n,k}\}$ “центральной” точки X_i по системе ее ближайших соседей $\{X_{n,1}, \dots, X_{n,k}\}$, то есть решение оптимизационной задачи $\|X_n - \sum_j w_{n,j} X_{n,j}\|^2$ [14];
- результатом применения метода главных компонент к окрестности $U_N(X_n, \varepsilon)$ является ортогональная матрица $Q_{PCA}(X_n)$ размера $p \times q$, чьи столбцы являются главными собственными векторами, соответствующими q наибольшим собственным значениям [17, 28]. Такие матрицы определяют q -мерные линейные подпространства $L_{PCA}(X_n) = \text{Span}(Q_{PCA}(X_n))$ в \mathbb{R}^p , которые, при определенных условиях хорошо аппроксимируют касательные пространства $L(X_n)$ (2) к МД \mathbb{X} в точке X_n [29].

2.5. Шаг третий: глобальное описание

Вычисляется глобальное описание МД. Для этого решаются выпуклые оптимизационные задачи при некоторых нормировочных ограничениях. Обычно, низкоразмерные признаки $\mathbb{Y}_N = h(\mathbb{X}_N) = \{y_1, \dots, y_N\} \subset \mathbb{Y}_h$ находятся как решения задач минимизации выбранной функции цены $L(\mathbb{Y}_N | \mathbb{X}_N)$ по \mathbb{Y}_N .

Примерами таких функций могут служить

$$L_{LLE}(\mathbb{Y}_N | \mathbb{X}_N) = \sum_{n=1}^N \|y_n - \sum_{j=1}^N w_{n,j} y_{n,j}\|_F^2;$$

$$L_{LE}(\mathbb{Y}_N | \mathbb{X}_N) = \sum_{n,j=1}^N K(X_n, X_j) \cdot \|y_n - y_j\|_2^2;$$

$$L_{LTSA}(\mathbb{Y}_N | \mathbb{X}_N) = \sum_{j=1}^N \|(I_q - Q_{PCA}(X_n)) \cdot Q_{PCA}^T(X_n) \cdot H_n \cdot Y_{(n)}\|_F^2;$$

используемые в алгоритмах LLE [14], LE [16] и LTSA [21] соответственно; здесь $Y_{(n)}$ — матрицы размера $q \times (k(n) + 1)$, состоящие из q -мерных столбцов $\{y_n, y_{n,1}, \dots, y_{n,k(n)}\}$, где пары индексов используются как в $U_N(X_n)$; $H_n = I_q - (1/k(n)) \cdot \vec{1} \cdot \vec{1}^T$ — матрица размера $q \times q$, I_q — единичная матрица порядка q , $\vec{1}$ — вектор из единиц размера q . Некоторые ограничения в задачах используются, чтобы избежать вырожденных решений.

2.6. Шаг четвертый: расширение задачи для точек вне выборки

Выборка признаков \mathbb{Y}_N задает значения отображения вложения $h(X)$ (3) только в точках выборки. Обычно, интерес представляет нахождение признаков $h(X)$ для точек вне выборки $X \in \mathbb{X} \setminus \mathbb{X}_N$. Популярные расширения для алгоритмов LLE, LE основаны на общем подходе ядерного метода главных компонент [30] и предложены в [31, 32]. Концепция функций цены также позволяет строить вложения для $X \in \mathbb{X} \setminus \mathbb{X}_N$ [33].

3. Основной результат

Множество $\mathbb{X} \setminus \mathbb{X}_N$ неизвестно. Поэтому для отображения $h(X)$ (3) необходимо задать область определения. Так как алгоритмы могут оценивать различные свойства многообразий, то, чтобы гарантировать хорошие свойства оценок, область определения задается различным образом в зависимости от рассматриваемого алгоритма. В данной работе рассматривается область определения алгоритма спектральных вложений Грассмана-Штифеля [22, 28, 22, 4, 23, 25].

Для ее определения введем дополнительные обозначения. Для точки $X \in \mathbb{X}$, гладкой неотрицательной монотонно убывающей функции одной скалярной переменной $k(\rho)$, $\rho \in [0, \infty)$ и скалярного параметра $\varepsilon = \varepsilon(N)$ определим выборочную матрицу ковариационную матрицу:

$$\Sigma(X) = \frac{\sum_{n \in U_N(X, \varepsilon)} k(\|X - X_n\|/\varepsilon) \cdot (X_n - X) \cdot (X_n - X)^T}{\sum_{n \in U_N(X, \varepsilon)} k(\|X - X_n\|/\varepsilon)}, \quad (4)$$

где $U_N(X, \varepsilon)$ — шар радиуса ε с центром в X в \mathbb{R}^p , $\|X - X_n\|$ — вторая норма в пространстве \mathbb{R}^p .

Определение 1. Область определения \mathbb{X}_h отображения сжатия $h(X)$ определяется как множество точек $X \in \mathbb{R}^p$ таких, что ровно q наибольших собственных чисел матрицы $\Sigma(X)$ больше или равны заданному скалярному параметру λ_{\min} .

Обозначим

$$C_K = \frac{\int_{Z \in \mathbb{R}^q, \|Z\| \leq 1} k(\|Z\|) \|Z\|^2 dZ}{q \cdot \int_{Z \in \mathbb{R}^q, \|Z\| \leq 1} k(\|Z\|) dZ}. \quad (5)$$

Обозначим расстояние Хаусдорфа между множествами $A, B \subset \mathbb{R}^p$ через $d_H(A, B)$:

$$d_H(A, B) = \max\{\sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|a - b\|\}. \quad (6)$$

Основной результат работы состоит в

Теорема 1. Если относительно многообразия \mathbb{X} , выборки, параметра $\varepsilon(N)$ предполагается M1-M8, S1-S3, P1-P3 соответственно, функция $k(\rho)$, $\rho \in [0, \infty)$ является неотрицательной монотонно убывающей функцией одной скалярной переменной и параметры $0 < \alpha < \beta < 1$, то для $\alpha \cdot C_K \leq \lambda_{\min} \leq \beta \cdot C_K$ выполнено:

$$d_H(\mathbb{X}_h, \mathbb{X}) \xrightarrow{p} 0.$$

Для доказательства теоремы нам потребуются дополнительные обозначения.

Обозначим \mathbb{X}_ε — множество внутренних точек многообразия, отстоящих от границы не менее чем на ε :

$$\mathbb{X}_\varepsilon = \{X \in \mathbb{X} \mid \forall V \in T_X(\mathbb{X}) \cap \|V\| < \varepsilon: \exp_X(V) \in \mathbb{X}\}, \quad (7)$$

где $\exp_X(V)$ — экспоненциальное отображение вектора V в точке X (соответствие между окрестностью нуля в касательном пространстве и окрестностью точки X многообразия).

Пусть для точки $X \in \mathbb{X}$ векторы V_1, \dots, V_q задают ортонормированный базис в касательном пространстве $L(X)$, и векторы W_1, \dots, W_{p-q} задают ортонормированный базис в ортогональном дополнении $L(X)$ до \mathbb{R}^p . Обозначим элементы матрицы ковариации в базисе $V_1, \dots, V_q, W_1, \dots, W_{p-q}$: для $E_1, E_2 \in \{V_1, \dots, V_q, W_1, \dots, W_{p-q}\}$:

$$\begin{aligned} & \Sigma(X|E_1, E_2) = \\ &= \sum_{n \in U_N(X, \varepsilon)} k(\|X - X_n\|/\varepsilon) \cdot \\ & \cdot ((X_n - X)^T \cdot E_1) \cdot ((X_n - X)^T \cdot E_2) / \\ & \sum_{n \in U_N(X, \varepsilon)} k(\|X - X_n\|/\varepsilon), \end{aligned} \quad (8)$$

Теорема 2 (об уклонении элементов матрицы выборочной матрицы ковариации). Если относительно многообразия \mathbb{X} , выборки, параметра $\varepsilon(N)$ предполагается M1-M8, S1-S3, P1-P3 соответственно, функция $k(\rho)$, $\rho \in [0, \infty)$ является неотрицательной монотонно убывающей функцией одной скалярной переменной, $\varepsilon < \varepsilon_{\max}$, то с вероятностью не меньше

$$\left(\frac{6a\sqrt{p}}{\varepsilon}\right)^p \cdot \exp(-N \cdot \varepsilon^q \cdot C_M)$$

для всех $X \in \mathbb{X}_\varepsilon$ выполнено

$$|\Sigma(X|V_i, V_j) - C_K \cdot \delta_{ij}| \leq \varepsilon^2 \cdot C_\Sigma;$$

$$|\Sigma(X|V_i, W_j)| \leq \varepsilon^2 \cdot C_\Sigma;$$

$$|\Sigma(X|W_i, W_j)| \leq \varepsilon^2 \cdot C_\Sigma,$$

где ε_{\max} , C_Σ и C_M — положительные константы, $\delta_{ij} = 1$ тогда и только тогда, когда $i = j$, a — ребро описанного вокруг \mathbb{X} p -мерного гиперкуба (9), $\Sigma(X|E_1, E_2)$ определена в (8).

Доказательство. Доказана в работе [34]. \square

Теорема 3 (Вейля). Пусть A и E — симметричные матрицы порядка p . Тогда для упорядоченных по возрастанию собственных $\lambda_1 \geq \dots \geq \lambda_p$ и $\mu_1 \geq \dots \geq \mu_p$ чисел матриц A и $A + E$ выполнено для $i = 1, \dots, p$

$$\lambda_i - \|E\|_F \leq \mu_i \leq \lambda_i + \|E\|_F,$$

где $\|B\|_F$ — норма Фробениуса матрицы B .

Доказательство. Доказана в книге [35]. \square

Доказательство теоремы 1. Для матрицы $A = (a_{ij})_{i,j=1}^p$, где $a_{i,i} = C_K$ для $i \leq p$, а остальные элементы равны нулю, и матрицы $E = (e_{ij})_{i,j=1}^p$, где $|e_{ij}| \leq \varepsilon^2 \cdot C_\Sigma$, по теореме 3 для собственных чисел матрицы $A + E$:

$$\mu_i \geq C_K - p^2 \cdot \varepsilon^2 \cdot C_\Sigma, \quad i \leq q;$$

$$\mu_i \leq p^2 \cdot \varepsilon^2 \cdot C_\Sigma, \quad i > p.$$

Следовательно, при больших N :

$$\mu_i \geq \lambda_{\min}, \quad i \leq q;$$

$$\mu_i \leq \lambda_{\min}, \quad i > p.$$

Из теоремы 2 следует, что $\mathbb{X}_\varepsilon \subset \mathbb{X}_h$ с вероятностью не менее

$$\left(\frac{6a\sqrt{p}}{\varepsilon}\right)^p \cdot \exp(-N \cdot \varepsilon^q \cdot C_M).$$

Далее, точки удаленные от \mathbb{X} больше, чем на ε не входят в \mathbb{X}_h . Поэтому

$$\sup_{X \in \mathbb{X}} \inf_{X' \in \mathbb{X}_h} \|X - X'\| \leq \varepsilon$$

с вероятностью не меньше

$$\left(\frac{6a\sqrt{p}}{\varepsilon}\right)^p \cdot \exp(-N \cdot \varepsilon^q \cdot C_M),$$

и всегда

$$\sup_{X \in \mathbb{X}_h} \inf_{X' \in \mathbb{X}} \|X - X'\| \leq \varepsilon.$$

Остается заметить, что по предположению P2

$$\left(\frac{6a\sqrt{p}}{\varepsilon}\right)^p \cdot \exp(-N \cdot \varepsilon^q \cdot C_M) \rightarrow 0.$$

Таким образом,

$$d_H(\mathbb{X}_h, \mathbb{X}) \xrightarrow{p} 0.$$

□

4. Заключение

В работе рассмотрена область определения отображения сжатия алгоритма спектральных вложений Грассмана-Штифеля. Для нее доказана состоятельность: расстояние Хаусдорфа между неизвестным многообразием и областью определения стремится к нулю по вероятности.

5. Благодарности

Работа была выполнена в ИППИ РАН при поддержке гранта РФФИ 16-29-09649 офн_м.

A. Модель данных

Будем предполагать, что

- M1. $\mathbb{X} \in \mathbb{R}^p$ является q -мерным многообразием, покрытым одной картой. То есть, для некоторого открытого $\mathbb{B} \in \mathbb{R}^q$ и гомеоморфного отображения $f: \mathbb{B} \rightarrow \mathbb{R}^p$ выполнено $\mathbb{X} = f(\mathbb{B})$;
- M2. q — известно;
- M3. \mathbb{B} — ограниченное множество;
- M4. собственные значения матрицы $J_f^T(b) \cdot J_f(b)$ размера $q \times q$ якобианов отображения f равномерно отделены от 0 и бесконечности;
- M5. гессиан f существует и поэлементно ограничен на \mathbb{X} ;
- M6. третьи производные f существуют и поэлементно ограничены \mathbb{X} ;
- M7. число обусловленности $c(\mathbb{X})$ конечно. Числом обусловленности $c(\mathbb{X})$ многообразия здесь называется такое число, что для любой точки, удаленной от \mathbb{X} не более чем на $1/c(\mathbb{X})$, существует единственная проекция на \mathbb{X} [26];

M8. многообразие \mathbb{X} геодезически выпукло, то есть кратчайшая, соединяющая любые две точки многообразия, является отрезком геодезической;

Замечание 1. Предположение M1 эквивалентно существованию на многообразии глобальной системы координат размерности q .

Свойство M3 используется для получения равномерных свойств рассматриваемых статистик.

Предположения M4-M6 представляют собой условия гладкости и применяются для связи евклидовых расстояний и объемов с расстояниями и объемами на многообразии.

Свойство M7 означает, что многообразие не содержит “коротких замыканий”: близость точек по евклидовому расстоянию влечет близость точек на многообразии.

То есть, пересечение малой евклидовой окрестности точки многообразия порождает малую окрестность точки многообразия.

Предположение M8 является техническим упрощением, которое используется для простоты записи разложения Тейлора.

Так как \mathbb{B} ограничено и ограничены производные f , то ограничено и \mathbb{X} . Обозначим a — ребро описанного около многообразия гиперкуба:

$$a = \inf_{a', a_1, \dots, a_p: \mathbb{X} \subset \otimes_{i=1}^p [a_i, a_i + a']} a'. \quad (9)$$

Обозначим c_J и C_J минимальное и максимальное собственные значения метрического тензора $J_f(b)^T J_f(b)$, $b \in \mathbb{B}$ соответственно:

$$c_J = \inf_{b \in \mathbb{B}} \min_{\lambda \in \sigma(J_f(b)^T J_f(b))} \lambda; \quad (10)$$

$$C_J = \inf_{b \in \mathbb{B}} \max_{\lambda \in \sigma(J_f(b)^T J_f(b))} \lambda. \quad (11)$$

Введем обозначение для максимального элемента матрицы Гессе отображения (\mathbb{B}, f) :

$$\begin{aligned} C_H &= \sup_{X \in \mathbb{X}, i, j=1, \dots, q} \left\| \frac{\partial^2 f(b)}{\partial b_i \partial b_j} \right\| = \\ &= \sup_{X \in \mathbb{X}, i, j=1, \dots, q} \sqrt{\sum_{k=1}^p \left| \frac{\partial^2 f_k(b)}{\partial b_i \partial b_j} \right|^2}, \end{aligned} \quad (12)$$

где $X = (f_1(b), \dots, f_p(b))^T$. Обозначим

$$\begin{aligned} C_T &= \sup_{X \in \mathbb{X}, i, j, m=1, \dots, q} \left\| \frac{\partial^3 f(b)}{\partial b_i \partial b_j \partial b_m} \right\| = \\ &= \sup_{X \in \mathbb{X}, i, j, m=1, \dots, q} \sqrt{\sum_{k=1}^p \left| \frac{\partial^3 f_k(b)}{\partial b_i \partial b_j \partial b_m} \right|^2}. \end{aligned} \quad (13)$$

Многообразие \mathbb{X} неизвестно и представлено лишь случайной выборкой точек

$$\mathbb{X}_n = \{X_1, \dots, X_N\} \subset \mathbb{X} \subset \mathbb{R}^p$$

размера N . Причем относительно способа выбора точек предполагается:

- S1. Точки для \mathbb{X}_n выбираются независимо одинаково распределенные из некоторой меры μ такой, что \mathbb{X} является ее носителем: $\mathbb{X} = \text{supp}\mu$;
- S2. Мера μ является непрерывной относительно римановой меры на многообразии с отделенной от нуля и бесконечности плотностью p_μ ;
- S3. Плотность p_μ дифференцируема на многообразии \mathbb{X} с ограниченной производной.

Замечание 2. Многообразие \mathbb{X} снабжено Римановой мерой $dV(X)$ (мерой объема), которая в локальных координатах в главном члене равна q -мерному объему. Пусть $(\Omega, \mathfrak{B}, \mathbb{P})$ — вероятностное пространство, тогда борелевская функция $\mathfrak{B} \rightarrow \mathbb{X}: X = X(\omega)$ называется случайной величиной на многообразии. Обозначим индуцированную этим отображением на \mathbb{X} меру через μ . Тогда, если любого борелевского множества

$$\mu(X \in \mathfrak{X}) = \int_{\mathfrak{X}} p(X) dV(X),$$

то функция $p_\mu(X), X \in \mathbb{X}$ называется функцией плотности [36].

Обозначим p_{\min} и p_{\max} — минимальное и максимальное значение плотности μ на многообразии \mathbb{X} (которые существуют по предположению S2):

$$p_{\min} = \inf_{X \in \mathbb{X}} p_\mu(X); \quad (14)$$

$$p_{\max} = \sup_{X \in \mathbb{X}} p_\mu(X). \quad (15)$$

Обозначим максимальные значения первых и вторых производных $p_\mu(X)$ (которые существуют, если выполнено S3):

$$C_{p,1} = \sup_{X \in \mathbb{X}, \theta \in T_X(\mathbb{X}): \|\theta\|=1} \|\nabla_\theta p_\mu(X)\|; \quad (16)$$

$$C_{p,2} = \sup_{X \in \mathbb{X}, \theta \in T_X(\mathbb{X}): \|\theta\|=1} \|\nabla_\theta \nabla_\theta p_\mu(X)\|, \quad (17)$$

где ∇_θ — ковариантная производная, которую можно понимать как обобщение производной по направлению на случай многообразий.

Замечание 3. S1 и S2 являются стандартными предположениями, гарантирующими связь выборки \mathbb{X}_n с многообразием \mathbb{X} , свойства которого изучаются.

Предположение S3 используется для доказательства равномерности результатов по многообразию.

Также, относительно параметра размера окрестности $\varepsilon = \varepsilon(N)$ предполагается:

P1. При $N \rightarrow \infty: \varepsilon \rightarrow 0$;

P2. При $N \rightarrow \infty: N\varepsilon^q \rightarrow \infty$;

P3. При $N \rightarrow \infty: N\varepsilon^{q+4} \rightarrow 0$.

Замечание 4. Предположение P1 означает, что размер окрестности стремится к нулю, а значит при разложении функций главным членом будет член с наименьшей степенью длины. Условие P2 обеспечивает попадание бесконечного числа точек из выборки, несмотря на убывание размера окрестности. Условие P3 сильнее условия P1 и гарантирует, что вклад смещений порядка ε^2 будет бесконечно мал в результатах центральной предельной теоремы для числа точек порядка $N\varepsilon^q$.

Список литературы

- [1] David L. Donoho. High-dimensional data analysis: The curses and blessings of dimensionality. *AMS conference on math challenges of 21st century*, pages 1–31, 2000.
- [2] Michel Verleysen. Learning High-Dimensional Data. *Limitations and Future Trends in Neural Computation*, 186:141–162, 2003.
- [3] Y. Bengio, A. Courville, and P. Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, aug 2013.
- [4] Alexander Bernstein and Alexander Kuleshov. Low-Dimensional Data Representation in Data Analysis. pages 47–58. 2014.
- [5] Alexander Kuleshov and Alexander Bernstein. Manifold Learning in Data Mining Tasks. pages 119–133. 2014.
- [6] John A. Lee and Michel Verleysen. *Nonlinear Dimensionality Reduction*. Information Science and Statistics. Springer New York, New York, NY, 2007.
- [7] H. S. Seung and Daniel D. Lee. COGNITION: The Manifold Ways of Perception. *Science*, 290(5500):2268–2269, dec 2000.
- [8] Elizaveta Levina and Peter J. Bickel. Maximum Likelihood Estimation of Intrinsic Dimension. In *Advances in Neural Information Processing Systems*, pages 777–784. MIT Press, 2005.
- [9] Mingyu Fan, Hong Qiao, and Bo Zhang. Intrinsic dimension estimation of manifolds by incising balls. *Pattern Recognition*, 42(5):780–787, may 2009.
- [10] Jochen Einbeck and Zakiah Kalantana. Intrinsic Dimensionality Estimation for High-dimensional Data Sets: New Approaches for the Computation of Correlation Dimension. *Journal of Emerging Technologies in Web Intelligence*, 5(2), may 2013.
- [11] Alessandro Rozza, Gabriele Lombardi, Marco Rosa, Elena Casiraghi, and Paola Campadelli. IDEA: Intrinsic Dimension Estimation Algorithm. pages 433–442. 2011.

- [12] Andrew Smith, Hongyuan Zha, and Xiao-ming Wu. Convergence and Rate of Convergence of a Manifold-Based Dimension Reduction Algorithm. *Advances in Neural Information Processing Systems 21*, pages 1529–1536, 2009.
- [13] Yunqian Ma and Yun Fu. *Manifold Learning Theory and Applications*. CRC Press, Boca Raton, 2011.
- [14] Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.
- [15] J. B. Tenenbaum, V. de Silva, and JC. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323, dec 2000.
- [16] Mikhail Belkin and Partha Niyogi. Laplacian Eigenmaps for dimensionality reduction and data representation. *Journal Neural Computation*, 15(6):1373–1396, 2003.
- [17] Zhenyue Zhang and Hongyuan Zha. Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment. *SIAM Journal on Scientific Computing*, 26(1):313–338, jan 2004.
- [18] D. L. Donoho and C. Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences*, 100(10):5591–5596, may 2003.
- [19] Kilian Q. Weinberger and Lawrence K. Saul. Unsupervised Learning of Image Manifolds by Semidefinite Programming. *International Journal of Computer Vision*, 70(1):77–90, oct 2006.
- [20] Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, jul 2006.
- [21] Zhenyue Zhang and Hongyuan Zha. Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment. *SIAM Journal on Scientific Computing*, 26(1):313–338, jan 2004.
- [22] A. Bernstein and A. Kuleshov. Manifold Learning: Generalization Ability and Tangent Proximity. *International Journal of Software and Informatics*, 7(3):359–390, 2013.
- [23] A. Bernstein, A. Kuleshov, and Y. Yanovich. Information preserving and locally isometric&conformal embedding via Tangent Manifold Learning. In *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on*, pages 1 – 9, Paris, 2015. IEEE.
- [24] Alexander Bernstein, Alexander Kuleshov, and Yury Yanovich. Information preserving and locally isometric&conformal embedding via Tangent Manifold Learning. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–9. IEEE, oct 2015.
- [25] Alexander Kuleshov and Alexander Bernstein. Extended Regression on Manifolds Estimation. In *Lecture Notes in Computer Science*, chapter Conformal, pages 208–228. 2016.
- [26] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the Homology of Submanifolds with High Confidence from Random Samples. *Discrete & Computational Geometry*, 39(1-3):419–441, mar 2008.
- [27] Y. Goldberg, A. Zakai, D. Kushnir, and Y Ritov. Manifold Learning: The Price of Normalization. *J. Mach. Learn. Re*, 9:1909–1939, 2008.
- [28] Alexander Bernstein and Alexander Kuleshov. Manifold Learning: Generalization Ability and Tangent Proximity. *International Journal of Software and Informatics*, 7(3):359–390, 2013.
- [29] A Singer and H.-T Wu. Vector diffusion maps and the connection Laplacian. *Communications on Pure and Applied Mathematics*, 65(8):1067–1144, aug 2012.
- [30] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10(5):1299–1319, jul 1998.
- [31] Yoshua Bengio, Olivier Delalleau, Nicolas Le Roux, Jean-François Paiement, Pascal Vincent, and Marie Ouimet. Learning Eigenfunctions Links Spectral Embedding and Kernel PCA. *Neural Computation*, 16(10):2197–2219, oct 2004.
- [32] Yoshua Bengio, Jean-Francois Paiement, and Pascal Vincent. Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering. In *Advances in Neural Information Processing Systems*, pages 177–184, 2003.
- [33] Kerstin Bunte, Michael Biehl, and Barbara Hammer. A General Framework for Dimensionality-Reducing Data Visualization Mapping. *Neural Computation*, 24(3):771–804, mar 2012.
- [34] Yury Yanovich. Asymptotic Properties of Local Sampling on Manifold (under review). *Journal of Mathematics and Statistics*, 2016.
- [35] Joel N. Franklin. *Matrix Theory*. Dover Publications, 2000.
- [36] Xavier Pennec. Probabilities And Statistics On Riemannian Manifolds: Basic Tools For Geometric Measurements. IN *IEEE WORKSHOP ON NONLINEAR SIGNAL AND IMAGE PROCESSING*, pages 194–198, 1999.