УДК 51-74; 519.246

Б. Г. Кухаренко, канд. физ.-мат. наук, ст. науч. сотр., вед. науч. сотр.,
Институт машиноведения РАН, г. Москва, e-mail: kukharenkobg@gmail.com,
М. О. Солнцева-Чалей, аспирант,
Московский физико-технический институт (ГУ), e-mail: solnceva.chalei@gmail.com

Моделирование центроидов для пучков многомерных траекторий

Оптимизация целевой функции алгоритма k-средних с условием асимптотического схождения пучка позволяет определить центроиды для пучков пространственных траекторий. Для полного определения сходящихся пучков используются оценки их геометрических асимптот, как наиболее правдоподобные ортогональные линейные регрессии рассеянных данных траекторий. Пучки траекторий выделяются на основе меры косинуса до соответствующих асимптот, с последующим определением их центроидов. В качестве примера определены центроиды пучков траекторий посадки самолетов на полосы аэродрома.

Ключевые слова: *моделирование, оптимизация, пучки траекторий, центроиды, геометрические асимптоты, линейная регрессия, рассеянные данные*

Введение

Рассматриваются подходы к моделированию центроидов (оптимальных траекторий) для сходящихся пучков многомерных траекторий. Примером таких пучков являются наборы трехмерных пространственных траекторий посадки самолетов при заходе на заданную полосу аэродрома. Формально, для векторов $\{\mathbf{x}[i] \in \mathbb{R}^{3\times L}, i \in \mathbb{N}_k\}$ ($L \gg 1$ — максимальная длина траектории), представляющих пучок траекторий $\mathbb{N}_k$, $k = \overline{1, K_0}$ (K_0 — эмпирический параметр), выполняется условие (асимптотического) схождения пучка

$$\forall (i, j) \in \mathbb{N}_k, \|\mathbf{x}[L_i; i] - \mathbf{x}[L_j; j]\|_2 < \varepsilon, \quad (1)$$

где ($\forall i \in \mathbb{N}_k$, $\mathbf{x}[L_i; i]$ — координаты точек траекторий, которые почти совпадают, т. е. параметры L_i, $i \in \mathbb{N}_k$, подлежат определению; $\|..\|_2$ — Евклидова мера расстояния в трехмерном пространстве $\mathbb{R}^3$; ε — порог (порядка ширины полосы в случае посадки самолетов). Пучки траекторий похожи на пучки функций и пучки решений дифференциальных уравнений (см. [1—3]). Траектории в анализируемых сходящихся пучках имеют типичный профиль и геометрическую асимптоту — линию в $\mathbb{R}^3$, удовлетворяющую условию (1) и касательную в окрестности конечных точек $\forall i \in \mathbb{N}_k$, $\mathbf{x}[L_i; i]$ всех траекторий пучка (в рассматриваемом случае с порогом ε (1)) [4]. Тогда центроиды пучков $\{\boldsymbol{\mu}[k] \in \mathbb{R}^{3\times d}, k = \overline{1, K_0}\}$ также удовлетворяют условию типа (1) в виде

$$\forall k = \overline{1, K}, \forall i \in \mathbb{N}_k, \|\boldsymbol{\mu}[L_k; k] - \mathbf{x}[L_i; i]\|_2 < \varepsilon, \quad (2)$$

где параметры L_i, $i \in \mathbb{N}_k$ и L_k, $k = \overline{1, K}$ подлежат определению. Если считать, что пучки $\mathbb{N}_k$, $k = \overline{1, K}$ определяются кластерами $\mathbb{C}_k$, $k = \overline{1, K}$, полученными в результате геометрического разбиения пространства векторов $\{\mathbf{x}[n] \in \mathbb{R}^{3\times d}, n = \overline{1, N}\}$, $d \gg 1$ на ячейки в стиле диаграмм Воронова (*Voronoi diagram* [5]), то центроиды пучков определяются алгоритмом разбиения на заданное число кластеров (*K-means algorithm*) [6]. Ограничение этого подхода связано с тем, что: 1) выделяемые области пространства состояния $\mathbb{R}^{3\times d}$ имеют различное число граней; 2) Евклидова мера расстояния между траекториями (в целом) в пространстве состояний $\mathbb{R}^{3\times d}$ плохо отображает подобие (и различие) профилей траекторий в трехмерном пространстве $\mathbb{R}^3$; 3) на практике в $\mathbb{R}^3$ пучки траекторий пересекаются.

1. Методы моделирования центроидов для пучков многомерных траекторий

1.1. Определение центроидов и пучков траекторий посредством алгоритма K-средних

Для многомерных векторов $\{\mathbf{x}[i] \in \mathbb{R}^{3\times L}, i = \overline{1, N}\}$, $L \gg 1$ с Евклидовой мерой расстояния алгоритм K-средних решает задачу идентификации центроидов $\{\boldsymbol{\mu}[k] \in \mathbb{R}^{3\times d}, k = \overline{1, K}\}$, ассоциируемых с кластерами $\mathbb{C}_k$, $k = \overline{1, K}$. Для набора центроидов сумма квадратов Евклидовых расстояний до векторов в соответствующих кластерах $\mathbb{C}_k$, $k = \overline{1, K}$, является минимальной. Введем назначения векторов кластерам посредством набора бинарных индикаторных переменных $\{r[i; k] \in \{0, 1\}, k = \overline{1, K}\}$, $i = \overline{1, N}$ (то есть, если вектор $\mathbf{x}[i]$ назначен кластеру k, то $r[i; k] = 1$, а в противном случае — $r[i; k] = 0$ (т. е. $\mathbb{C}_k = \{\mathbf{x}[i] | r[i; k] = 1\}$, $k = \overline{1, K}$). Целевая функция алгоритма K-средних имеет вид

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} r[i; k]((\mathbf{x}[i] - \boldsymbol{\mu}[k]) \cdot (\mathbf{x}[i] - \boldsymbol{\mu}[k])), \quad (3)$$

где "$\cdot$" обозначает скалярное произведение векторов в пространстве состояний $\mathbb{R}^{3 \times L}$. Для определения кластеров, представляющих пучки траекторий, минимизацию целевой функции (3) необходимо проводить с учетом условия типа (2) в виде

$$\begin{aligned} &\forall k = \overline{1, K}, \exists i \in \{j | r[j; k] = 1\}, \|\boldsymbol{\mu}[L; k] - \mathbf{x}[L; i]\|_2 < \varepsilon, \\ &\forall k,\ l = \overline{1, K},\ \|\boldsymbol{\mu}[L; k] - \boldsymbol{\mu}[L; l]\|_2 \gg \varepsilon, \end{aligned} \quad (4)$$

где $\|..\|_2$ — Евклидова мера расстояния в трехмерном пространстве $\mathbb{R}^3$. В алгоритме K-средних для инициализации $\{\boldsymbol{\mu}[k], k = \overline{1, K}\}$ используются $\{\mathbf{x}[i], i = \overline{1, N}\}$. Минимизация J (3) осуществляется последовательными итерациями, состоящими из двух шагов: оценки $\{r[n; k], n = \overline{1, N}, k = \overline{1, K}\}$ при фиксированных $\{\boldsymbol{\mu}[k], k = \overline{1, K}\}$ в замкнутой форме

$$\langle r[n; k] \rangle = \\ = \begin{cases} 1, \text{ если } k = \arg\min\limits_{J} ((\mathbf{x}[n] - \boldsymbol{\mu}[j]) \cdot (\mathbf{x}[n] - \boldsymbol{\mu}[j])) \\ 0, \text{ в противном случае} \end{cases} \quad (5)$$

и оценки $\{\boldsymbol{\mu}[k], k = \overline{1, K}\}$ при фиксированных $\{r[n; k], n = \overline{1, N}, k = \overline{1, K}\}$

$$\{\langle \boldsymbol{\mu}[k] \rangle, k = \overline{1, K}\} = \\ = \arg\min_{\{\boldsymbol{\mu}[k],\ k = \overline{1, K}\}} \sum_{n=1}^{N} \sum_{k=1}^{K} r[n; k]((\mathbf{x}[n] - \boldsymbol{\mu}[j]) \cdot (\mathbf{x}[n] - \boldsymbol{\mu}[j]))$$

в замкнутой форме

$$\langle \boldsymbol{\mu}[k] \rangle = \frac{\sum_{n=1}^{N} r[n; k] \mathbf{x}[n]}{\sum_{n=1}^{N} r[n; k]} \quad (6)$$

до достижения сходимости. Поскольку каждый шаг уменьшает целевую функцию J (3), сходимость алгоритма K-средних гарантируется. Однако вследствие неудачной инициализации $\{\boldsymbol{\mu}[k], k = \overline{1, K}\}$ он может сходиться к локальному, а не глобальному минимуму J при условии (4), гарантирующем, что кластеры $\mathbb{C}_k$, $k = \overline{1, K}$, представляют пучки траекторий $\mathbb{N}_k$, $k = \overline{1, K}$.

Непосредственная реализация алгоритма K-средних относительно медленная, поскольку на каждом шаге определения $r[i; k]$, $i = \overline{1, N}$, $k = \overline{1, K}$ (5) вычисляется Евклидово расстояние между каждым вектором $\boldsymbol{\mu}[k]$, $k = \overline{1, K}$ и каждым вектором $\mathbf{x}[i]$, $i = \overline{1, N}$. Ускорение алгоритма K-средних обеспечивается предварительным построением дерева, в котором ближайшие вектора $\{\mathbf{x}[i], i = \overline{1, N}\}$ находятся в одном поддереве [7, 8]. Используя неравенство треугольников для расстояний, также сокращают число вычислений расстояний [9, 10]. Использование в исходном пространстве состояний $\mathbb{R}^{3 \times L}$ Евклидовой меры расстояния и представление о геометрическом разбиении пространства векторов $\{\mathbf{x}[i] \in \mathbb{R}^{3 \times L}, i = \overline{1, N}\}$, $L \gg 1$, на ячейки в стиле диаграмм Воронова не отражает характер пучков траекторий (посадки самолетов). Кроме того, определение центроидов оказывается неустойчивым к случайным отклонениям (*outliers*).

Первый способ адаптации условной задачи оптимизации (3)—(4) к анализу пучков пространственных траекторий состоит в использовании для центроидов $\{\boldsymbol{\mu}[k], k = \overline{1, K}\}$ модели полиномиальной регрессии, которая решает задачу выравнивания $\forall i \in \mathbb{N}_k$, $\mathbf{x}[L_i; i]$ — координаты точек траекторий пучка, которые почти совпадают [11, 12].

Второй способ состоит в использовании представлений векторов $\{\mathbf{x}[i] \in \mathbb{R}^{3 \times L}, i = \overline{1, N}\}$, $L \gg 1$ в пространствах абстрактных характеристик $\{\mathbf{y}[i] \in \mathbb{R}^M, i = \overline{1, N}\}$ с Евклидовой мерой расстояния [13—15]. Вектора, представляющие пучки траекторий, которые геометрически неразделимы в исходном пространстве состояний $\mathbb{R}^{3 \times L}$, в пространствах абстрактных характеристик становятся разделимыми, поэтому в этих пространствах используется Евклидова мера расстояния [16, 17]. При отображении в исходное пространство состояний эта метрика становится неевклидовой. Существуют примеры неевклидовой меры расстояния для векторов траекторий в пространстве состояний. Например, мера косинуса

$$\boldsymbol{\rho}_{\text{cosin}}(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}') / (\sqrt{(\mathbf{x} \cdot \mathbf{x})} \sqrt{(\mathbf{x}' \cdot \mathbf{x}')}) \quad (7)$$

наиболее адекватно отражает близость векторов $\{\mathbf{x}[i] \in \mathbb{R}^{3 \times L}, i = \overline{1, N}\}$, $L \gg 1$ в пространстве состояний, представляющих пучки траекторий определенного профиля [17]. Поэтому можно модифицировать задачу оптимизации (3)—(4), заменив в выражении (3) скалярное произведение на общую меру расстояния.

1.2. Использование неевклидовой меры расстояния

Целевая функция (3) алгоритма K-средних обобщается введением общей меры расстояния $\boldsymbol{\rho}(\mathbf{x}, \mathbf{x}')$

между двумя векторами $\mathbf{x}$, $\mathbf{x}' \in \mathbb{R}^{3\times L}$ и минимизацией целевой функции

$$J = \sum_{i=1}^{N}\sum_{k=1}^{K} r[i;k]\rho^2(\mathbf{x}[i], \boldsymbol{\mu}[k]) \tag{8}$$

(*K-medoids algorithm*) [6]. При заданных центроидах $\{\boldsymbol{\mu}[k],\ k = \overline{1,K}\}$ шаг оценки $\{r[i;k],\ i = \overline{1,N},\ k = \overline{1,K}\}$ (как в стандартном алгоритме K-средних) включает назначение каждого вектора $\mathbf{x}[i]$, $i = \overline{1,N}$ кластеру $\mathbb{C}_k$, $k = \overline{1,K}$, для которого расстояние $\rho(\mathbf{x}[i], \boldsymbol{\mu}[k])$ с соответствующим центроидом минимально:

$$\langle r[i;k]\rangle = \begin{cases} 1, \text{ если } k = \arg\min\limits_{j} \rho(\mathbf{x}[i], \boldsymbol{\mu}[k]) \\ 0, \text{ в противном случае} \end{cases} \tag{9}$$

с оценкой вычислительной сложности $O(K \cdot N)$. Однако шаг оценки $\{\boldsymbol{\mu}[k],\ k = \overline{1,K}\}$ является потенциально более сложным. При стандартном ограничении, что каждый центроид является одним из векторов, назначенных соответствующему кластеру, условие (4) выполняется автоматически. Это позволяет реализовать алгоритм для любого выбора меры расстояния $\rho(\mathbf{x}[i], \boldsymbol{\mu}[k])$, которая непосредственно вычисляется. Шаг определения $\{\boldsymbol{\mu}[k],\ k = \overline{1,K}\}$ включает дискретный поиск по всем N_k векторам, назначенным этому кластеру $\mathbb{C}_k$, $k = \overline{1,K}$, и требует $O(N_k^2)$ оценок меры расстояния $\rho(\mathbf{x}[i], \boldsymbol{\mu}[k])$.

Отметим, что взаимосвязь центроидов и кластеров

$$\boldsymbol{\mu}[k] \Leftrightarrow \mathbb{C}_k,\ k = \overline{1,K},$$

навязана логикой алгоритма K-средних и его обобщениями и, в принципе, центроиды $\{\boldsymbol{\mu}[k],\ k = \overline{1,K}\}$ и кластеры $\mathbb{C}_k$, $k = \overline{1,K}$, представляющие пучки траекторий $\mathbb{N}_k$, $k = \overline{1,K}$, можно определять независимо.

1.3. Оценки центроидов и определение пучков траекторий

В качестве независимых методов оценки центроидов $\{\boldsymbol{\mu}[k],\ k = \overline{1,K}\}$ используются скрытые компоненты линейных и нелинейных динамических моделей (см. [18, 19]), и скрытые последовательности марковских моделей (см. [20, 21]) при условии (4). После этого пучки траекторий определяются по схеме $\boldsymbol{\mu}[k] \Rightarrow \mathbb{C}_k$, $k = \overline{1,K}$, на основе однократного применения формулы (9) с мерой косинуса (7).

1.4. Оценка пучков траекторий и определение центроидов

Для оценки пучков траекторий можно использовать геометрические методы триангуляции [5, 22]. После этого центроиды $\{\boldsymbol{\mu}[k],\ k = \overline{1,K}\}$ определяются по схеме $\mathbb{C}_k \Rightarrow \boldsymbol{\mu}[k]$, $k = \overline{1,K}$ на основе однократного применения формулы

$$\{\langle\boldsymbol{\mu}[k]\rangle,\ k = \overline{1,K}\} = \\ = \arg\min_{\{\boldsymbol{\mu}[k],\ k=\overline{1,K}\}} \sum_{i=1}^{N}\sum_{k=1}^{K} r[i;k]\rho^2_{\text{cosin}}(\mathbf{x}[i], \boldsymbol{\mu}[k]) \tag{10}$$

при условии (4) с использованием квадрата меры косинуса (7). Оценка (10) эффективна при представлении векторов $\{\mathbf{x}[i],\ i = \overline{1,N}\}$ и $\{\boldsymbol{\mu}[k],\ k = \overline{1,K}\}$ в исходном пространстве состояний.

2. Линейная регрессия рассеянных данных алгоритмом RANSAC для оценки асимптот пучков траекторий

Как отмечается во введении, особенностью рассматриваемых пучков траекторий является то, что все они имеют характерную геометрическую асимптоту в области сходимости траекторий (1) [4]. Поскольку дискретные точки траекторий пучка плотно лежат в окрестности асимптоты, основа предлагаемого метода для оценки асимптот пучков траекторий состоит в том, что набор векторов кластера многомерных траекторий $\{\mathbf{x}[i] \in \mathbb{R}^{3\times L},\ i = \overline{1,N}\}$ рассеивается во множество точек этих траекторий

$$\{\mathbf{x}[i] \in \mathbb{R}^{3\times L},\ i = \overline{1,N}\} \Rightarrow \\ \Rightarrow \{(x[j,i], y[j,i], z[j,i]) \in \mathbb{R}^3,\ j = \overline{1,L},\ i = \overline{1,N}\}. \tag{11}$$

Множество точек (11) должно быть упорядочено по значениям одной из координат (в направлении возрастания — *ascend* или убывания — *descend*). При этом происходит упорядочение по остальным координатам всех точек, представляющих сходящийся пучок траекторий движения по определенному профилю. После этого для рассеянных трехмерных данных $\{\mathbf{z}_i = (x_i, y_i, z_i),\ i = \overline{1, L\cdot N}\}$ (11) с помощью алгоритма RANSAC (Random Sample and Consensus — случайная выборка и консенсус) анализируются модели ортогональной линейной регрессии

$$\mathfrak{M}(\boldsymbol{\theta}) = \{(x, y, z) \in \mathbb{R}^3 | (a_1x + b_1y + c_1z = d_1) \wedge \\ \wedge (a_2x + b_2y + c_2z = d_2)\}, \tag{12}$$

где $\wedge$ — конъюнкция; $\boldsymbol{\theta} = \{a_1, b_1, c_1, d_1, a_2, b_2, c_2, d_2\}$ — вектор параметров этих моделей при заданном пороге Евклидового расстояния $\rho_\perp(\mathbf{z}, \mathfrak{M}(\boldsymbol{\theta}))$, вычисляемого по ортогональной проекции точки $\mathbf{z} = (x, y, z)$ из множества (11) на линию $\mathfrak{M}(\boldsymbol{\theta})$. Таким образом, модель (12) симметрична относительно координат x, y, z. Для выдвижения гипотезы относительно модели ортогональной линейной регрессии (12) достаточно любой пары точек из множества (11). Окончательная модель (12) подтверждается наибольшим относительным количеством (процентом) рассеянных данных $\{\mathbf{z}_i = (x_i, y_i, z_i),\ i = \overline{1, L\cdot N}\}$ множества (11).

В настоящей работе используется алгоритм MLESAC (Maximum Likelihood Estimation Sample Consensus — консенсус выборок с оценкой по максимуму правдоподобия) — вероятностная версия алгоритма RANSAC [23, 24]. Этот алгоритм оценивает правдоподобие модели (12), представляя распределение расстояния рассеянных данных $\{\mathbf{z}_i = (x_i, y_i, z_i), i = \overline{1, L \cdot N}\}$ от модели $\mathfrak{M}(\boldsymbol{\theta})$ (12), как смесь распределения данных, подтверждающих модель (12) (*inliers*), и распределения данных, отклоняющих эту модель (*outliers*). Считая, что рассеянные данные $\mathbb{Z}$ множества (11) независимые, получаем выражение для логарифма правдоподобия в виде

$$L(\rho_{\perp}(\mathbb{Z}, \mathfrak{M}(\boldsymbol{\theta}))|\boldsymbol{\theta}) = \sum_{i=1}^{L \cdot N} \log(\gamma p(\rho_{\perp}(\mathbf{z}_i, \mathfrak{M}(\boldsymbol{\theta}))|\mathbf{z}_i \textit{ is inlier}) + (1-\gamma) p(\rho_{\perp}(\mathbf{z}_i, \mathfrak{M}(\boldsymbol{\theta}))|\mathbf{z}_i \textit{ is outlier})), \quad (13)$$

где γ — параметр смешивания. Распределение расстояний до данных, подтверждающих модель (12), представляется гауссовым распределением

$$p(\rho_{\perp}(\mathbf{z}_i, \mathfrak{M}(\boldsymbol{\theta}))|\mathbf{z}_i \textit{ is inlier}) \propto \exp\left(-\frac{(\rho_{\perp}(\mathbf{z}_i, \mathfrak{M}(\boldsymbol{\theta})))^2}{2\sigma^2}\right), \quad (14)$$

где σ — стандартное отклонение. Распределения расстояний до данных, отклоняющих модель (12), описываются равномерным распределением

$$p(\rho_{\perp}(\mathbf{z}_i, \mathfrak{M}(\boldsymbol{\theta}))|\mathbf{z}_i \textit{ is outlier}) = \begin{cases} (2\rho_{\max})^{-1}, \ \rho_{\perp}(\mathbf{z}_i, \mathfrak{M}(\boldsymbol{\theta})) < \rho_{\max} \\ 0, \ \rho_{\perp}(z_i, \mathfrak{M}(\boldsymbol{\theta})) \geqslant \rho_{\max}, \end{cases} \quad (15)$$

где $\rho_{\max}$ — наибольшее расстояние до данных (определяется контекстом). Минимизация логарифма правдоподобия (13) позволяет оценить вектор параметров $\boldsymbol{\theta}$ и параметр смешивания γ. Это обеспечивается итерациями алгоритма ожидания-максимизации правдоподобия [25].

Наиболее правдоподобная линейная регрессия рассеянных данных кластера траекторий определяет геометрическую асимптоту $\mathfrak{M}(\boldsymbol{\theta})[k]$, $k = \overline{1, K}$, модели (12) одного из пучков кластера при условии (1). А эта асимптота играет ту же роль, что и центроид $\boldsymbol{\mu}[k]$, $k = \overline{1, K}$, при условии (2), для оценки касательного ей пучка многомерных траекторий на основе однократного применения формулы (9) с мерой косинуса (7) (подраздел 1.3). После удаления из рассеянных данных (11) тех точек, которые представляют траектории выделенного пучка, повторяется процедура определения геометрической асимптоты и выделяется следующий пучок траекторий. Поскольку определение модели (12) должно быть симметричным относительно координат x, y, z, при формировании рассеянных данных оставшихся траекторий в (11) проводится сортировка по очередной пространственной координате по сравнению с использованной в (11) при определении предыдущей асимптоты модели (12). Возможная зависимость результата (12) от направления координат устраняется изменением направления сортировки в (11) с возрастания на убывание или наоборот. Анализ траекторий кластера завершается определением всех пучков в кластере.

3. Численный эксперимент

Используются траектории 117 самолетов, идущих на посадку в международном аэропорту и зарегистрированных радаром TRACON 1 января 2006 г. (данные в открытом доступе на сайте https://c3.nasa.gov/dashlink/resources/132/). Начало координат совпадает с положением радара, интервал времени между точками регистрации составляет около 5 с. В работе учитываются только 160 последних точек каждой траектории, что исключает случайные маневры самолетов перед заходом на посадку. Эти траектории в трехмерном пространстве представлены в работе [12]. Пять кластеров траекторий самолетов (на рис. 1 в проекциях на оси x, y и z, см. вторую сторону обложки) выделяются в результате применения метода полиномиальных регрессий [11, 12]. Распределение траекторий по кластерам следующее: 16 траекторий — в розовом кластере; 13 — в зеленом; 3, 37 и 38 — в синем, черном и красном кластерах соответственно.

Кластеры на рис. 1 (см. вторую сторону обложки) состоят из нескольких пучков траекторий, соответствующих определенным профилям посадки.

Определим пучки траекторий, составляющих красный кластер (см. рис. 1). Для структуры `trajs` размерности `size(trajs, 1) x size(trajs, 2) x size(trajs, 3)`, представляющей 38 трехмерных траекторий красного кластера (см. рис. 1), рассеянные данные `Data` (11) являются результатом кода на `MATLAB` (рис. 2).

```
% рассеяние данных
Data = zeros(size(trajs,2),size(trajs,3));
for k = 1:size(trajs,1),
    Data = [Data; squeeze(trajs(k,:,:))];
end;
% сортировка рассеянных данных
[~,Isort] = sort(Data(:,1),1);
Data = Data(Isort,:,:);
```

Рис. 2. Код на MATLAB для получения рассеянных данных траекторий

На рис. 3, *а* (см. вторую сторону обложки) показаны рассеянные данные `Data (:, 1:2)` (11) двумерных проекций `trajs(:, :, 1:2)` траекторий красного кластера (см. рис. 1) и результат их линейной регрессии с использованием алгоритма MLESAC, определяющий асимптоту первого пучка (голубая линия).

Траектории первого (голубого) пучка удаляются из красного кластера (см. рис. 1) на основе близости траекторий к голубой асимптоте (рис. 3, *а*, см. вторую сторону обложки) по мере косинуса (7). На рис. 3, *б* (см. вторую сторону обложки) показаны рассеянные данные `Data(:, 1:2)` двумерных проекций `trajs(:, :, 1:2)` траекторий красного кластера (см. рис. 1) за вычетом траекторий первого (голубого) пучка и результат их линейной регрессии (зеленая линия) с использованием алгоритма MLESAC, определяющий асимптоту второго (зеленого) пучка. Траектории второго (зеленого) пучка удаляются из красного кластера (см. рис. 1) на основе близости траекторий к зеленой асимптоте (см. рис. 3, *б*) по мере косинуса (7). В результате определяется третий (синий) пучок траекторий (см. рис. 3, *в*). Для трех выделенных пучков на рис. 3, *в*: голубого, зеленого и синего, их центроиды $\{\boldsymbol{\mu}[k], k = \overline{1,K}\}$ (10) показаны тремя толстыми красными линиями. Результат определения пучков в красном кластере тот же, что и в работе [17]. Потенциально посторонние траектории движения самолетов в зоне риска (выбросы) специально не определяли, потому что используется мера косинуса, а это значит, что выбросы уже определены в работе [17].

Пучки траекторий розового кластера (см. рис. 1) существенно пересекаются. Другой особенностью этих траекторий является присутствие практически линейных участков в их хвостах вдали от фокуса. Это препятствует непосредственному определению асимптот сходящихся пучков по методу раздела 2. Поскольку все траектории в кластере имеют одинаковое направление времени, при анализе траекторий розового кластера используются рассеянные данные сокращенных траекторий с частью точек `quotum ≈ 0,4`, считая от фокусов пучков. На рис. 3, *а* (см. вторую сторону обложки) показаны рассеянные данные `Data (:, 1:2)` (11) двумерных проекций `trajs (:, :, 1:2)` полных траекторий розового кластера (см. рис. 1) и результат их линейной регрессии с использованием алгоритма MLESAC, определяющий асимптоту первого пучка (голубая линия), поскольку в этом случае результаты ортогональной линейной регрессии рассеянных данных сокращенных и полных траекторий совпадают.

Траектории первого (голубого) пучка удаляются из розового кластера (см. рис. 1) на основе близости траекторий к голубой асимптоте (рис. 4, *а*, см. вторую сторону обложки), по мере косинуса (7). На рис. 4, *б* (см. вторую сторону обложки) для наглядности показаны рассеянные данные `Data (:, 1:2)` двумерных проекций `trajs(:, :, 1:2)` полных траекторий розового кластера (см. рис. 1) за вычетом траекторий первого (голубого) пучка, но результат линейной регрессии (зеленая линия), определяющий асимптоту второго (зеленого) пучка, получен с использованием алгоритма MLESAC для рассеянных данных сокращенных траекторий. Траектории второго (зеленого) пучка удаляются из розового кластера (см. рис. 1) на основе близости траекторий к зеленой асимптоте (рис. 4, *б*) по мере косинуса (7). В результате определяется третий (синий) пучок траекторий (рис. 4, *в*, см. вторую сторону обложки). Для трех существенно пересекающихся пучков закрученных траекторий на рис. 4, *в*: голубого, зеленого и синего, их центроиды $\{\boldsymbol{\mu}[k], k = \overline{1,K}\}$ (10) показаны тремя толстыми красными линиями. Сравнение центроидов (рис. 4, *в*, см. вторую сторону обложки) с полиномиальной регрессией траекторий розового кластера (см. рис. 1) показывает, что результат алгоритма MLESAC на рис. 4, *в* является нетривиальным.

Заключение

Рассмотрены подходы к моделированию центроидов для сходящихся пучков многомерных траекторий. Жесткая взаимозависимость при определении центроидов и соответствующих им пучков траекторий является следствием идиомы алгоритма *K*-средних. Поэтому существуют двухшаговые методы: сначала независимое определение сходящихся пучков и затем моделирование центроидов для уже определенных пучков траекторий. Например, сходящиеся пучки траекторий могут быть идентифицированы посредством определения касательных им геометрических асимптот в точках их фокусов. Показано, как для последовательного определения асимптот, касательных к сходящимся пучкам траекторий, используются наиболее правдоподобные ортогональные линейные регрессии рассеянных данных траекторий, полученные с использованием алгоритма MLESAC. После определения геометрических асимптот соответствующие им пучки траектории отделяются на основе меры косинуса от траекторий до касательных им асимптот, и для этих пучков вычисляются центроиды. В качестве примера определены центроиды пучков траекторий посадки самолетов при заходе на полосы аэродрома.

Список литературы

1. **Johnson W. W.** On singular solutions of differential equations of the first order // Annals of Mathematics. April 1887. Vol. 3, N. 2. P. 33—38.
2. **Griffiths P., Harris J.** Principles of Algebraic Geometry, New York: John Wiley & Sons, 1994.
3. **Eisenbud D., Harris J.** 3264 and All That Intersection Theory: A Second Course in Algebraic Geometry. Cambridge, UK: Cambridge University Press, 2015.
4. **Погорелов А. И.** Дифференциальная геометрия. 6-е изд. М.: Наука, 1974.
5. **de Berg M., Cheong O., van Kreveld M., Overmars M.** Computational Geometry. Algorithms and Applications. Third Edition. Berlin, Heidelberg: Springer-Verlag, 2008.
6. **MacQueen J.** Some methods for classification and analysis of multivariate observations / LeCam L. M., Neyman J., eds. // Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. 1967. Vol. 1. P. 281—297.
7. **Ramasubramanian V., Paliwal K. K.** A generalized optimization of the k—d tree for fast nearest-neighbour search // Proceedings Fourth IEEE Region 10 International Conference (TENCON'89). 1990. P. 565—568.

8. **Moore A. W.** The anchors hierarch: using the triangle inequality to survive high dimensional data // Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence. 2000. P. 397—405.

9. **Hodgson M. E.** Reducing computational requirements of the minimum-distance classifier // Remote Sensing of Environments. 1998. Vol. 25. P. 117—128.

10. **Elkan C.** Using the triangle inequality to accelerate *k*-means / Proceedings of the Twelfth International Conference on Machine Learning. AAAI. 2003. P. 147—153.

11. **Gaffhey S., Smyth P.** Joint probabilistic curve clustering and alignment / Saul L., Weiss Y., Bottou L., eds. Proceedings of Neural Information Processing Systems (NIPS 2004). December 13—18, 2004, Vancouver, British Columbia, Canada. Advances in Neural Information Processing Systems. Vol. 17. Cambridge, MA: MIT Press. 2005. P. 473—480.

12. **Кухаренко Б. Г., Солнцева М. О.** Кластеризация управляемых объектов на основе сходства их многомерных траекторий // Информационные технологии. 2014. № 5. С. 3—7.

13. **Anjum N., Cavallaro A.** Trajectory clustering for scene context learning and outlier detection / Schonfeld D., Shan C., Tao D., Wang L,, eds. Video Search and Mining. Berlin. Studies in Computational Intelligence. Vol. 287. Heidelberg: Springer, 2010. P. 33—51.

14. **Кухаренко Б. Г., Солнцева М. О.** Покомпонентный дискриминантный анализ результатов кластеризации многомерных траекторий // Информационные технологии. 2014. № 11. С. 3—7.

15. **Кухаренко Б. Г., Солнцева М. О.** Итеративная кластеризация траекторий управляемых объектов в многомерном пространстве характеристик // Информационные технологии. 2014. № 8. С. 11—16.

16. **Chen G., Lerman G.** Spectral curvature clustering (SCC) // International Journal on Computer Vision. 2009. Vol. 81, N. 3. P. 317—330.

17. **Кухаренко Б. Г., Солнцева-Чалей М. О.** Спектральный метод с использованием полярной кривизны для анализа результатов кластеризации многомерных траекторий // Информационные технологии. 2015. Т. 21, N. 12. С. 901—905.

18. **Кухаренко Б. Г., Солнцева М. О.** Анализ результатов кластеризации многомерных траекторий посредством моделей линейных динамических систем // Информационные технологии. 2015. Т. 21, № 2. С. 104—109.

19. **Кухаренко Б. Г., Солнцева-Чалей М. О.** Применение моделей нелинейных динамических систем для анализа результатов кластеризации многомерных траекторий // Информационные технологии. 2015. Т. 21, № 5. С. 341—345.

20. **Listgarten J., Neal R. M., Roweis S. T., Emili A.** Multiple alignment of continuous time series / Saul L. K., Weiss Y., Bottou L., eds. Proceedings of Neural Information Processing Systems (NIPS 2004). December 13—18, 2004, Vancouver British Columbia, Canada. Advances in Neural Information Processing Systems. Vol. 17. Cambridge, MA: MIT Press, 2005. P. 5—13.

21. **Кухаренко Б. Г., Солнцева-Чалей М. О.** Применение моделей непрерывного профиля для анализа результатов кластеризации многомерных траекторий // Информационные технологии. 2015. Т. 21, № 8. С. 585—590.

22. **Галанин М. П., Щеглов И. А.** Разработка и реализация алгоритмов трехмерной триангуляции сложных пространственных областей: прямые методы. Препринт. М.: Институт прикладной математики им. М.В. Келдыша РАН, 2006.

23. **Torr P. H. S., Zisserman A.** MLESAC: A new robust estimator with application to estimating image geometry // Journal of Computer Vision and Image Understanding. 2000. Vol. 78, N. 1. P. 138—156.

24. **Кухаренко Б. Г.** Алгоритмы анализа изображений для определения локальных особенностей и распознавания объектов и панорам // Информационные технологии. 2011. № 7. Приложение. С. 1—32.

25. **Dempster A., Laird N. M., Rubin D. B.** Maximum likelihood from incomplete data via the EM algorithm // Journal of the Royal Statistical Society B. 1977. Vol. 39, N. 1. P. 1—38.

B. G. Kukharenko, Leading Research Scientist, e-mail: kukharenkobg@gmail.com,
Blagonravov Mechanical Engineering Research Institute of the RAS,
M. O. Solntseva-Chalei, Postgraduate Student, e-mail: solnceva.chalei@gmail.com,
Moscow Institute of Physics and Technology (SU)

Centroid Modeling for Multi Dimensional Trajectory Pencils

In this paper we describe how prevalent flight routes can be found automatically out of a comprehensive set of arrival flight trajectories. As example k-means clustering of the actual arrival routes is a fundamental step directing to the determination of the airport's prevalent arrival trajectory pencils. For each arrival trajectory pencils then a typical arrival route can be computed, describing the pencil specific mean arrival route (centriod). The polynomial regression model of centroid is useful in this case. As interrelation of centroid and respective trajectory pencils is a sequence of k-means idiom, independent centroid modeling methods for multi dimensional trajectory pencils are under study in present paper. With respect to independent centroid modeling methods, linear and nonlinear dynamical models and Markov models are mentioned. Next, to estimate converging trajectory pencil geometric asymptotes, maximum likelihood orthogonal linear regression models of the trajectory scattered data are in use. The models are obtained by MLESAC algorithm, which is a probabilistic version of RANSAC where the distance of scattered data points from the linear regression model is assumed as distributed according to a mixture of a Gaussian and a uniform distribution. After determining pencil geometric asymptotes sequentially, the trajectory pencils are selected based on trajectory cosine measure to respective asymptotes, and their centroids are defined. As example, trajectory pencil centroids are determined of specific airplane final descending on airport landing-strips of different directions.

Keywords: *modeling, optimization, trajectory pencil, centroids, geometric asymptotes, linear regressions, scattered data*

References

1. **Johnson W. W.** On singular solutions of differential equations of the first order, *Annals of Mathematics*, april 1887, vol. 3, no. 2, pp. 33—38.

2. **Griffiths P., Harris J.** *Principles of Algebraic Geometry*, New York: John Wiley & Sons, 1994.

3. **Eisenbud D., Harris J.** *3264 and All That Intersection Theory: A Second Course in Algebraic Geometry*, Cambridge, UK: Cambridge University Press, 2015.

4. **Pogorelov A. V.** *Differential Geometry*. Translated from the first Russian ed. by Boron L. F. Groningen: P. Noordhoff, N. V., 1959.

5. **de Berg M., Cheong O., van Kreveld M., Overmars M.** *Computational Geometry. Algorithms and Applications*. Third ed. Berlin, Heidelberg: Springer-Verlag, 2008.

6. **MacQueen J.** Some methods for classification and analysis of multivariate observations, LeCam L. M., Neyman J., eds. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, vol. 1, 1967. P. 281—297.

7. **Ramasubramanian V., Paliwal K. K.** A generalized optimization of the k-d tree for fast nearest — neighbour search, *Proceedings Fourth IEEE Region 10 International Conference (TENCON'89)*, 1990. P. 565—568.

8. **Moore A. W.** The anchors hierarch: using the triangle inequality to survive high dimensional data, *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, 2000, pp. 397—405.

9. **Hodgson M. E.** Reducing computational requirements of the minimum-distance classifier, *Remote Sensing of Environments*, 1998, vol. 25, pp. 117—128.

10. **Elkan C.** Using the triangle inequality to accelerate k-means, *Proceedings of the Twelfth International Conference on Machine Learning*, AAAI, 2003, pp. 147—153.

11. **Gaffney S., Smyth P.** Joint probabilistic curve clustering and alignment / Saul L., Weiss Y., Bottou L., eds. *Proceedings of Neural Information Processing Systems (NIPS 2004), December 13—18, 2004, Vancouver, British Columbia, Canada*. Advances in Neural Information Processing Systems. V. 17. Cambridge, MA: MIT Press, 2005, pp. 473—480.

12. **Kukharenko B. G., Solntseva M. O.** Klasterizacia upravlyaemyh objektov na osnove shodstva ih mnogomemyh trajektoryi, *Informacionnye tehnologii*, 2014, no. 5. P. 3—7.

13. **Anjum N., Cavallaro A.** Trajectory clustering for scene context learning and outlier detection, Schonfeld D., Shan C., Tao D., Wang L., eds., *Video Search and Mining*, Berlin, Studies in Computational Intelligence. V. 287. Heidelberg: Springer, 2010, pp. 33—51.

14. **Kukharenko B. G., Solntseva M. O.** Pokomponentnyi discriminantnyi analiz resultatov klasterizacii mnogomernyh trajektoryi, *Informacionnye tehnologii*, 2014, no. 11, pp. 3—7.

15. **Kukharenko B. G., Solntseva M. O.** Iterativnaja klasterizacia trajektoryi upravlyaemych objektov v mnogomernom prostranstve harakteristik, *Informacionnye tehnologii*, 2014, no. 8, pp. 11—16.

16. **Chen G., Lerman G.** Spectral curvature clustering (SCC), *International Journal on Computer Vision*, 2009, vol. 81, no. 3, pp. 317—330.

17. **Kukharenko B. G., Solntseva-Chalei M. O.** Spectralnyi metod s ispolsovaniem polyarnoi krivizny dlya analiza resultatov klasterizacii mnogomernyh trajektoryi, *Informacionnye tehnologii*, 2015, vol. 21, no. 12, pp. 901—905.

18. **Kukharenko B. G., Solntseva M. O.** Analiz resultatov klasterizacii mnogomernyh trajektoryi posredstvom modelei lineinyh dinamicheskih sistem, *Informacionnye tehnologii*, 2015, no. 2, pp. 104—109.

19. **Kukharenko B. G., Solntseva-Chalei M. O.** Primenenie modelei nelineinyh dinamicheskih sistem dlya analiza resultatov klasterizacii mnogomernyh trajektoryi, *Informacionnye tehnologii*, 2015, vol. 21, no. 5, pp. 341—345.

20. **Listgarten J., Neal R. M., Roweis S. T., Emili A.** Multiple alignment of continuous time series, Saul L. K., Weiss Y., Bottou L., eds. *Proceedings of Neural Information Processing Systems (NIPS 2004), December 13—18, 2004, Vancouver, British Columbia, Canada*. Advances in Neural Information Processing Systems. V. 17. Cambridge, MA: MIT Press. 2005. P. 5—13.

21. **Kukharenko B. G., Solntseva-Chalei M. O.** Primenenie modelei nepreryvnogo profilya dlya analiza resultatov klasterizacii mnogomernyh trajektoryi, *Informacionnye tehnologii*, 2015, vol. 21, no. 8, pp. 585—590.

22. **Galanin M. P., Scheglov I. A.** *Razrabotka i realizacia algoritmov trehmernoi triangulyacii slognyh prostranstvennyh oblastei: pryamye metody, Preprint*, Moscow M. V. Keldysh Institute of Applied Mathematics of the RAS, 2006.

23. **Torr P. H. S., Zisserman A.** MLESAC: A new robust estimator with application to estimating image geometry, *Journal of Computer Vision and Image Understanding*, 2000, vol. 78, no. 1, pp. 138—156.

24. **Kukharenko B. G.** Algoritmy analiza izobragenyi dlya opredelenija lokalnyh osobennostei i raspoznavaniya objectov i ponoram, *Informacionnye tehnologii,* 2011, no. 7, Prilogenie, pp. 1—32.

25. **Dempster A., Laird N. M., Rubin D. B.** Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society B,* 1977, vol. 39, no. 1, pp. 1—38.

УДК 330.43

М. Г. Матвеев, д-р техн. наук, проф., зав. каф., e-mail: mgmatveev@yandex.ru,
Воронежский государственный университет,
В. В. Михайлов, д-р техн. наук, проф., нач. факультета, e-mail: VladimirMihaiIov36@gmail.com,
ВУНЦ ВВС "Военно-воздушная академия им. профессора Н. Е. Жуковского и Ю. А. Гагарина",
Е. А. Сирота, канд. физ.-мат. наук, доц. каф. цифровых технологий, e-mail: atoris@list.ru,
Воронежский государственный университет

Комбинированная прогностическая модель нестационарного многомерного временного ряда для построения пространственного профиля атмосферной температуры

Предлагается обоснование, построение и анализ комбинированной прогностической модели нестационарного многомерного временного ряда в целях построения пространственного профиля атмосферной температуры. Результаты моделирования показывают, что использование комбинированной модели с выделением кластеров однородной статистики для построения прогноза профиля атмосферной температуры может обеспечить приемлемые характеристики прогноза, а также может служить для построения вертикальных температурных профилей и, в конечном итоге, для составления рекомендаций при выполнении полетов авиации.

Ключевые слова: *нестационарный многомерный временной ряд, прогностическая модель, метеорология, векторная авторегрессия, профиль атмосферной температуры, моделирование, классы однородной статистики*