

На правах рукописи



Солнцева-Чалей Мария Олеговна

**МЕТОДЫ АНАЛИЗА МНОГОМЕРНЫХ ДАННЫХ
ПОСАДОЧНЫХ ТРАЕКТОРИЙ САМОЛЁТОВ**

Специальность 05.13.17

Теоретические основы информатики

Автореферат

диссертации на соискание ученой степени

кандидата технических наук

Москва

2018 год

Работа выполнена федеральном государственном автономном образовательном учреждении высшего образования «Московский физико-технический институт (государственный университет)»

Научные руководители:

Кухаренко Борис Георгиевич,

Кандидат физико-математических наук, доцент

Востоков Вадим Юрьевич,

Кандидат технических наук, доцент

Официальные оппоненты:

Соловьёв Владимир Игоревич,

доктор экономических наук, руководитель департамента анализа данных, принятия решений и финансовых технологий федерального государственного образовательного учреждения высшего образования «Финансовый университет при правительстве Российской Федерации»

Игнатов Дмитрий Игоревич,

кандидат технических наук, доцент департамента анализа данных и искусственного интеллекта факультета компьютерных наук НИУ ВШЭ, научный сотрудник научно-учебной лаборатории интеллектуальных систем и структурного анализа НИУ ВШЭ

Ведущая организация:

Филиал «НИИ Аэронавигации» ФГУП ГосНИИ ГА

Защита состоится «31» мая 2018 года в 13:00 на заседании диссертационного совета

Д 002.073.05 при Федеральном Исследовательском Центре российской Академии наук (ФИЦ ИУ РАН) по адресу: 119333, г. Москва, ул. Вавилова, 40, конференц-зал.

С диссертацией можно ознакомиться в библиотеке и на сайте ФИЦ ИУ РАН <http://www.frccsc.ru/>.

Автореферат разослан «30» апреля 2018 г.

Ученый секретарь
диссертационного совета,
Д 002.073.05, д. ф.-м.н., профессор

Рязанов В.В.

Общая характеристика работы

Актуальность темы. Усложнение технических средств, рост интенсивности их использования и повышение требований эксплуатационной безопасности делают задачу обработки больших массивов данных (мониторинга технических средств) в режиме реального времени крайне актуальной. Одно из перспективных направлений решения указанной задачи связано с разработкой методов и алгоритмов обработки данных, учитывающих их характерные (специфические) особенности. Данные мониторинга технических средств могут быть представлены в виде многомерных временных рядов – последовательности векторов, компонентами которых являются значения различных показателей реализуемого процесса, зафиксированных в один момент времени. Типичным примером подобного массива данных являются посадочные траектории самолетов, фиксируемые радаром. К числу характерных (специфических) особенностей посадочных траекторий следует отнести:

- их кривизну и закрученность,
- возможность образования асимптотически сходящихся пучков,
- наличие в рамках пучка множественности пространственных пересечений.

Посадка (согласно статистике Росавиации и ИКАО) является наиболее критическим (с точки зрения безопасности) этапом полёта, в связи с чем в условиях постоянного увеличения воздушного трафика и загруженности аэропортов особую важность приобретают задачи оптимизации загруженности взлётно-посадочных полос и совершенствования существующих систем управления воздушным движением, решение которых без анализа посадочных траекторий не может быть корректно.

В настоящее время принят и реализуется ряд государственных и межгосударственных программ, направленных, в том числе, на поддержание высокого уровня безопасности полетов в современных условиях загруженности. В первую очередь, следует выделить следующие программы:

- федеральную целевую программу «Модернизация Единой системы организации воздушного движения Российской Федерации (2009 - 2020 годы)», утвержденную Постановлением Правительства Российской Федерации от 1 сентября 2008 г. № 652;
- научно-исследовательскую программу Single European Sky Air Traffic Management Research (SESAR) Program, инициированную Еврокомиссией в 2005 году;
- программу развития системы воздушного транспорта «Next Generation Air Transport System», для реализации которой Конгресс США в 2003 году учредил Joint Planning and Development Office.

В рамках указанных программ проводятся исследования, в которых методы анализа данных (data mining) применяются при обработке данных посадочных траекторий самолётов, регистрируемых радаром в зоне аэропорта, с целью:

- выделить при рассмотрении набора траекторий установившиеся потоки движения – пучки посадочных траекторий;
- определить характерную (типичную) траекторию посадки на данную ВПП,
- определить в наборе успешных посадок на заданную ВПП нетипичную траекторию.

(Следует отметить, что решение трех сформулированных задач крайне важно в ситуациях, когда одна диспетчерская служба (радар) обслуживает несколько независимых между собой аэропортов, например, как в районе залива Сан-Франциско, где расположено более 25 различных аэропортов, при этом более 10 аэропортов не имеют собственных диспетчерских служб (см. https://en.wikipedia.org/wiki/List_of_airports_in_the_San_Francisco_Bay_Area)).

Актуальность задачи выделения установившихся потоков движения обусловлена необходимостью разделения воздушного пространства на зоны ответственности между диспетчерскими службами и/или их сотрудниками. Также важно заметить, что траектории посадки самолетов представляют собой данные высокой размерности (high dimensional data), поэтому в связи со сложностью задачи и размерностью данных анализ траекторий самолетов осуществляется в двумерном пространстве (проекция на плоскость посадки). Такое сокращение размерности данных зачастую приводит к существенной потере информации и искажению результата. Для решения задачи выделения воздушных потоков в плоском случае в настоящее время используются методы, связанные с разбиением пространства на ячейки в стиле диаграмм Вороного и/или разбиением набора векторов, представляющих траектории движения, с помощью различных алгоритмов кластеризации (например, K-means или K-medoids). Данные подходы имеют ряд недостатков, например, таких как:

- разбиение пространства на ячейки не связано с характером существующих потоков движения,
- конечный результат в существенной мере зависит от параметров инициализации используемых алгоритмов кластеризации, что на практике не может обеспечить получение корректного результата.

Разработка алгоритмов выделения пучков траекторий, то есть групп траекторий движения в трёхмерном пространстве, имеющих сходные характеристики и близких по конечным координатам (как, например, в случае посадок с одного направления подлета на заданную взлетно-посадочную полосу), находится на начальной стадии. В литературе при выделении

пучков траекторий в плоском случае активно применяются такие методы как: анализ главных компонент (PCA), непараметрические Байесовские методы, спектральная кластеризация и т.п. Однако, при использовании вероятностных методов возможно получение неточного или нестабильного результата. Кроме того, существенным недостатком большинства известных методов является использование евклидовой меры расстояния в качестве меры близости траекторий движения, что не позволяет правильно разделять пересекающиеся траектории различной геометрической формы.

Особую актуальность решение задачи определения характерной (опорной) траектории посадки, также как и предыдущей, приобретает при посадке на взлетно-посадочные полосы без сопровождения диспетчерской службы и/или сложном рельефе местности. Определение характерной (опорной) траектории посадки возможно только после предварительного решения задачи выделения пучков траекторий. В условиях отсутствия такого решения делаются попытки определить характерную (опорную) траекторию в двумерном случае: Существует ряд алгоритмов (например, DBSCAN или методы спектральной кластеризации), которые позволяют выделить центроиды (центр масс траекторий в кластере или поточечное среднее) в определенных в плоском случае характерных потоках движения. Однако, такие методы вычислительно затратны, а для рассмотрения их результатов в качестве опорных траекторий нет никаких оснований.

Решение задач прогнозирования дальнейшего движения самолётов на основании данных об успешных реализациях в условиях постоянного увеличения траффика также приобретает всё большее значение. При решении этой задачи могут быть использованы различные методы машинного обучения, однако для уменьшения ошибки прогнозирования необходимо, после выделения группы траекторий потока, удалить из анализируемой выборки аутлаеры. В случае траекторий самолётов появление аутлаеров может быть связано с неправильной регистрацией данных радаром или с нетипичным поведением самолёта (Hrastovec & Solina (2016)). Задача определения аутлаеров при анализе траекторных данных достаточно сложна и активно разрабатывается в настоящее время. В литературе для этих целей предлагается использование различных методов таких, как TRAOD (A Partition-and-Detect Framework), анализа главных компонент (functional PCA) и др. Однако, они не позволяют на реальных данных получить устойчивые результаты.

При решении поставленных задач предварительное исследование посадочных траекторий самолётов с помощью ряда методов data mining и их комбинаций позволяет не только оценить применимость таких методов на практике, но и дает возможность выявить тонкую структуру анализируемых данных и их характерные особенности, что является важным

для получения корректного результата.

В настоящее время активные исследования, связанные с разбиением воздушного пространства на сектора и определением характерных (опорных) траекторий посадки, ведутся специалистами ряда организаций, в том числе, Филиал «НИИ Аэронавигации» ФГУП ГосНИИ ГА, ИПУ РАН им. Трапезникова, ЦАГИ, NASA Ames, The Boeing Company, Honeywell International Inc. и The MITRE Corporation.

Подробное описание методов data mining, используемых в настоящее время, можно найти в работах Sam Roweis, Zoubin Ghahramani, Andrew Ng, Kevin P. Murphy, Padhraic Smyth, Shumway R., Stoffer D., Keinosuke Fukunaga и Marco Zulliani.

Цель работы. Настоящая работа посвящена разработке методов анализа данных (data mining), позволяющих при обработке набора многомерных временных рядов, представленных четырехмерными траекториями движения объектов к различным целям:

- автоматически выделять в трехмерном случае пучки многомерных временных рядов, асимптотически сходящиеся с заданным параметром порога, что соответствует выделению установившихся потоков движения;
- смоделировать в трехмерном случае типичный для выделенного пучка многомерный временной ряд – центроид, представляющий собой линию тренда или характерную траекторию потока;
- определять в выделенном пучке выброс (outlier) – многомерный временной ряд, наиболее отклоняющийся от смоделированного центроида.

При рассмотрении набора траекторий самолётов решение поставленных задач позволяет с учётом установившегося движения разбивать воздушное пространство на зоны ответственности диспетчерских служб, и способствует обеспечению безопасности полётов, в частности, при посадках:

- в аэропорту, находящемся в зоне сложного географического ландшафта (в горах или на побережье);
- без сопровождения диспетчерских служб («малые» аэропорты и/или несанкционированные приземления).

Объект исследования: многомерные временные ряды (посадочные траектории самолётов, регистрируемые радаром).

Предмет исследования: методы и алгоритмы анализа данных (посадочных траекторий самолётов).

Положения, выносимые на защиту:

1. **Способ устойчивого разбиения в трехмерном пространстве набора многомерных временных рядов, представленных четырехмерными траекториями движения объектов к различным целям, на пучки, асимптотически сходящиеся с заданным параметром порога, с использованием меры косинуса в качестве меры близости многомерных временных рядов .**

Обоснованность результата подтверждается применением комплексного анализа хорошо известных методов и алгоритмов анализа данных, в том числе метода сокращения размерности анализируемых данных, алгоритма RANSAC (Random Sample and Consensus) и процедуры перехода в пространство исходной размерности.

Новизна научного результата заключается в том, что предложен оригинальный алгоритм разбиения набора многомерных временных рядов, представленных посадочными траекториями самолётов, на пучки – посадки на заданные ВПП, основанный на использовании меры косинуса в качестве меры близости траекторий движения к выделенной в пространстве меньшей размерности геометрической асимптоты, касательной пучку траекторий в окрестности сходимости.

Полученный результат является **вкладом** в развитие методов анализа многомерных данных, поскольку позволяет выделять в трёхмерном пространстве группы траекторий движения, имеющих общие цели, и данный результат является устойчивым.

Достоверность научного результата подтверждается примерами выделения в трёхмерном пространстве пучков посадочных траекторий при обработке данных радара TRACON над заливом Сан-Франциско с помощью программной реализации алгоритма.

2. **Способ моделирования в трёхмерном пространстве центроида для выделенного пучка многомерных временных рядов, представленных четырёхмерными траекториями движения объектов, посредством решения оптимизационной задачи с ограничениями и использовании меры косинуса в качестве меры близости многомерных временных рядов .**

Обоснованность результата подтверждается комплексным анализом решения стандартной оптимизационной задачи с ограничениями по определению центроида – траектории, сумма квадратов расстояний от точек которой до соответствующих точек всех траекторий в пучке минимальна по выбранной мере близости, и которая удовлетворяет ограничению параметра порога (при анализе посадочных траекторий самолё-

тов – конечная точка такой траектории должна находиться на взлётно-посадочной полосе).

Новизна научного результата заключается в том, что при решении оптимизационной задачи в качестве меры близости многомерных временных рядов, представленных посадочными траекториями самолётов, используется мера косинуса, которая учитывает характерные особенности трехмерных траекторий движения – кривизну, кручение и множественные пересечения.

Полученный результат является **вкладом** в развитие методов анализа многомерных данных, поскольку позволяет при обработке четырёхмерных данных траекторий движения объектов получить устойчивый результат.

Достоверность научного результата подтверждается примерами определения многомерного временного ряда, характерного для выделенной группы, – центроида для выделенных пучков посадочных траекторий при обработке данных радара TRACON над заливом Сан-Франциско с помощью программной реализации алгоритма.

3. Способ определения выброса (outlier) в выделенном пучке многомерных временных рядов, представленных четырёхмерными траекториями движения объектов, основанный на оценке отклонения траекторий пучка от его центроида по мере косинуса.

Обоснованность результата подтверждается комплексным анализом решения стандартной оптимизационной задачи с ограничениями по определению аутлаера – траектории, сумма квадратов расстояний от точек которой до соответствующих точек центроида максимальна по выбранной мере близости

Новизна научного результата заключается в том, что при решении оптимизационной задачи в качестве меры близости многомерных временных рядов, представленных посадочными траекториями самолётов, используется мера косинуса, учитывающая все характерные особенности анализируемых данных.

Полученный результат является **вкладом** в развитие методов анализа данных траекторий движения, поскольку позволяет получить точный и устойчивый результат.

Достоверность научного результата подтверждается примерами определения аутлаеров в выделенных пучках траекторий при обработке данных радара TRACON над заливом Сан-Франциско с помощью программной реализации алгоритма.

Апробация результатов: результаты работы докладывались и обсуждались на:

- научном семинаре ИППИ РАН (Москва, 17 марта 2017 г.);
- International conference on Big Data and its application (Москва, Deworkacy, 16 сентября 2016 года);
- научном семинаре им. братьев Белоцерковских (Москва, 16 июня 2016 г.).
- заседании кафедры информатики и вычислительной математики МФТИ (Долгопрудный, 16 июня 2015 г.);
- научном семинаре ВЦ РАН (Москва, 24 июня 2015 г.);
- 55-58 научных конференциях МФТИ (Долгопрудный, 2012-2015 г.);

Публикации: по теме диссертации опубликовано 20 работ, в том числе 12 в журналах, включенных в «Перечень российских рецензируемых научных журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней доктора и кандидата наук» Высшей аттестационной комиссии при Министерстве образования и науки Российской Федерации. По материалам диссертационного исследования было подготовлено 4 патента на изобретение RU № 2616106, RU № 2616107, RU № 2017101343 и US № 15/258,736.

Личный вклад автора: все представленные в работе результаты получены лично автором.

Структура и объем диссертации: диссертационное исследование состоит из введения, трех глав и заключения. Работа изложена на 131 странице машинописного текста, содержит 36 рисунков и список литературы из 194 наименований.

Краткое содержание работы

Во введении формулируется цель работы, перечисляются методы исследования и приводятся основные результаты.

В первой главе на основе комплексного анализа проблемы обеспечения безопасности полётов в расширенной зоне аэропорта обосновывается актуальность темы диссертационного исследования и выбор посадочных траекторий самолётов в качестве первичных данных для решения поставленных задач. Описываются основные подходы, применяемые для решения задачи выделения установившихся потоков движения в расширенной области воздушного пространства зоны аэропорта. Демонстрируется, что методы анализа данных, используемые в настоящее время для решения проблем обеспечения безопасности полётов, не учитывают особенностей пространственной геометрии многомерных посадочных траекторий самолётов в трёхмерном пространстве и не могут быть использованы при решении практических задач.

В второй главе на основе комплексного анализа широко известных методов data mining предложен способ выделения пучков посадочных траекторий в плоском случае, новизна которого состоит в применении спектрального метода с использованием полярной кривизны (LSCC, Linear Spectral Curvature Clustering). При выделении пучка траектории анализируемая выборка траекторий (многомерных векторов) моделируется объединением d -мерных аффинных (линейных) подпространств. При этом учитывается не бинарное, а d -арное подобие, и принадлежность рассматриваемых траекторий (многомерных векторов) к подвыборке оценивается (многомерным) тензором близости порядка $(d + 2)$, который разворачивается в матрицу близости (подобия) векторов, обеспечивая эффективное использование спектрального метода. Пучки, выделяемые с помощью алгоритма LSCC, соответствуют траекториям посадки на заданные взлётно-посадочные полосы с одного направления. Данный метод может быть использован при выделении пучков траекторий в двумерном случае, если число пучков, содержащихся в выборке, известно заранее.

После определения типичной траектории выделенного пучка с помощью Марковской модели непрерывного скрытого профиля (CPM, Continuous Profile Model), предлагается способ определения аутлаера – траектории, наиболее отклоняющейся от типичной, новизна которого состоит в использовании меры косинуса в качестве меры близости траекторий движения. Результаты применения способа демонстрируются на примере обработки данных посадочных траекторий, зарегистрированных радиолокационной системой TRACON (terminal radar approach control) 1 января 2006 г. над заливом Сан-Франциско и находящихся в свободном доступе.

В третьей главе, посвященной разработке математического аппарата обработки многомерных посадочных траекторий самолётов в трехмерном случае, предлагается оригинальный способ устойчивого разбиения набора многомерных посадочных траекторий самолётов на пучки траекторий, соответствующих посадкам на заданные взлётно-посадочные полосы (ВПП), в трёхмерном случае (метод без обучения, не требующий предварительного знания о числе пучков, содержащихся в выборке). Рассматриваемые пучки траекторий являются асимптотически сходящимися с заданным параметром порога, который определяется шириной ВПП. Точный результат разбиения достигается благодаря использованию алгоритма RANSAC при сокращении размерности данных и использовании меры косинуса в качестве меры близости траекторий при переходе в пространство исходной размерности. Предлагаемый способ позволяет точно решать задачу выделения установившихся потоков движения при обработке данных посадочных траекторий самолётов в трёхмерном случае. Результаты демонстрируются на примере выделения пучков траекторий в выборке траекторий, зарегистрированных радаром TRACON.

Помимо этого в настоящей главе предлагается способ определения типичной траектории выделенного потока – центра для выделенного пучка многомерных посадочных траекторий. Результат, полученный при решении оптимизационной задачи с ограничениями, является устойчивым и достигается благодаря использованию в качестве меры близости пространственных траекторий меры косинуса, которая учитывает особенности пространственной геометрии многомерных траекторий движения – кривизны, кручения и множественных пересечений. Результат демонстрируется на примере анализа реальных данных и может быть использован при решении практических задач.

Авторский метод, предлагаемый в настоящей главе, позволяет получать устойчивое разбиение анализируемых наборов посадочных траекторий на асимптотически сходящиеся пучки. Пучки траекторий выделяются на основании близости по мере косинуса траекторий рассматриваемой выборки к выделенным асимптотам траекторий. Оценка асимптот траекторий происходит при сокращении размерности данных – рассмотрении двумерной ортогональной проекции рассеянных данных. Оценка асимптот траекторий происходит при определении наиболее правдоподобной ортогональной линейной регрессии «рассеянных» данных алгоритмом RANSAC. Результаты, представленные в настоящей главе, опубликованы на русском и английском языке и запатентованы.

Геометрические особенности многомерных посадочных траекторий

Если исходный набор посадочных траекторий самолётов содержит несколько групп траекторий, соответствующих посадкам на заданные ВПП, – пучков траекторий, то их

возможно выделить.

Для векторов $\{\mathbf{X}_n \in \mathbb{R}^{3 \times L}, \mathbf{X}_n \in \mathbb{N}_k, n = \overline{1, N_k}\}$, ($L \gg 1$ – максимальная длина траектории), представляющих пучок траекторий $\mathbb{N}_k, k = \overline{1, K_0}$ (K_0 – эмпирический параметр), выполняется условие (асимптотического) схождения пучка

$$\forall (\mathbf{X}_n, \mathbf{X}_m) \in \mathbb{N}_k : \|X_n[L_n] - X_m[L_m]\|_2 < \varepsilon, \quad (1)$$

где $n, m = \overline{1, N_k}, n \neq m$, и $X_n[L_n], X_m[L_m]$ – конечные координаты точек траекторий, которые почти совпадают, $\|\dots\|_2$ – евклидова мера расстояния в пространстве \mathbb{R}^3 , ε – параметр порога (порядка ширины ВПП). Параметры L_n, L_m подлежат определению.

Пучки траекторий похожи на пучки функций и пучки решений дифференциальных уравнений. Также следует отметить, что траектории в сходящихся пучках (в рассматриваемом случае с порогом ε (1)) имеют типичный профиль и геометрическую асимптоту – линию в \mathbb{R}^3 , удовлетворяющую условию (1) и касательную в окрестности конечных точек $X_n[L_n]$ ($\mathbf{X}_n \in \mathbb{N}_k, n = \overline{1, N_k}$) всех траекторий пучка.

В выделенном пучке траекторий может быть определена центральная траектория – центроид. В таком случае, центроиды пучков $\{\mathbf{C}_k \in \mathbb{R}^{3 \times L}, k = \overline{1, K_0}\}$ также должны удовлетворять условию типа (1) в виде

$$\forall k = \overline{1, K_0}, \forall \mathbf{X}_n \in \mathbb{N}_k : \|\mathbf{C}_k[L_k] - X_n[L_n]\|_2 < \varepsilon, \quad (2)$$

где параметры L_n и L_k подлежат определению.

Многомерные посадочные траектории самолётов имеют множественные пересечения в пространстве, кривизну и кручение. Учитывать эти особенности можно, используя неевклидову меру расстояний для векторов траекторий в пространстве состояний. Например, меру косинуса

$$\cos \angle(\mathbf{X}_n, \mathbf{X}_m) = (\mathbf{X}_n \cdot \mathbf{X}_m) / \left(\sqrt{(\mathbf{X}_n \cdot \mathbf{X}_n)} \sqrt{(\mathbf{X}_m \cdot \mathbf{X}_m)} \right), n, m = \overline{1, N}, \quad (3.1)$$

которая наиболее адекватно отражает близость векторов $\{\mathbf{X}_n \in \mathbb{R}^{3 \times L}, n = \overline{1, N}\}$, $L \gg 1$ в пространстве состояний, представляющих пучки траекторий определённого профиля [3].

Далее для меры косинуса используется следующее обозначение:

$$\rho_{\cosine}(\mathbf{X}_n, \mathbf{X}_m) = \cos \angle(\mathbf{X}_n, \mathbf{X}_m). \quad (3.2)$$

Выделение пучков траекторий в трехмерном пространстве возможно при решении оптимизационной задачи, заменив скалярное произведение на общую меру расстояния. Целевая функция алгоритма K-means обобщается введением общей меры расстояния $\rho(\mathbf{X}_n, \mathbf{X}_m)$

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

$$\mathfrak{J} = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \rho^2(\mathbf{X}_n, \mathbf{C}_k), \quad (4)$$

(K-medoids algorithm). При заданных центроидах $\{\mathbf{C}_k, k = \overline{1, K}\}$ шаг оценки $\{r_{nk}, n = \overline{1, N}, k = \overline{1, K}\}$ (как в стандартном алгоритме K-means) включает назначение каждого вектора $\mathbf{X}_n, n = \overline{1, N}$ выборке $\mathbf{C}_k, k = \overline{1, K}$, для которого расстояние $\rho(\mathbf{X}_n, \mathbf{C}_k)$ с соответствующим центроидом минимально

$$\langle r_{nk} \rangle = \begin{cases} 1, & \text{если } k = \arg \min_m \rho(\mathbf{X}_n, \mathbf{C}_m), \\ 0, & \text{в противном случае} \end{cases} \quad (5)$$

с оценкой вычислительной сложности $O(K \cdot N)$.

Определение характерной геометрической асимптоты пучка многомерных траекторий движения с заданным параметром порога

Особенностью рассматриваемых пучков траекторий является то, что все они имеют характерную геометрическую асимптоту в области сходимости траекторий (1). Поскольку дискретные точки траекторий пучка плотно лежат в окрестности асимптоты, основа предлагаемого метода для оценки асимптот пучков траекторий состоит в том, что набор векторов выборки многомерных траекторий $\{\mathbf{X}_n \in \mathbb{R}^{3 \times L}, n = \overline{1, N_k}\}$ рассеивается во множество точек этих траекторий

$$\{\mathbf{X}_n \in \mathbb{R}^{3 \times L}, n = \overline{1, N_k}\} \Rightarrow \{(x_n[i], y_n[i], z_n[i]) \in \mathbb{R}^3, i = \overline{1, L}, n = \overline{1, N_k}\}. \quad (6)$$

Множество точек (6) 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 определённому профилю. После этого для «рассеянных» в результате такого упорядочения трёхмерных данных $\{Z_j = (x_j, y_j, z_j), j = \overline{1, L \cdot N_k}\}$ (6) с помощью алгоритма RANSAC (Random Sample and Consensus – случайная выборка и консенсус) (Fischler, 1981) анализируются модели ортогональной линейной регрессии

$$\mathfrak{M}(\boldsymbol{\theta}) = \left\{ (x, y, z) \in \mathbb{R}^3 \mid (a_1^*x + b_1^*y + c_1^*z = d_1^*) \wedge (a_2^*x + b_2^*y + c_2^*z = d_2^*) \right\}, \quad (7)$$

\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 евклидова расстояния $\rho_{\perp}(Z_j, \mathfrak{M}(\boldsymbol{\theta}))$, вычисляемого по ортогональной

проекции точки $Z_j = (x_j, y_j, z_j)$ из (6) на линию $\mathfrak{M}(\boldsymbol{\theta})$. Таким образом модель (7) симметрична относительно координат x, y, z . Для выдвижения гипотезы относительно модели ортогональной линейной регрессии (7) достаточно любой пары точек из (6). Окончательная модель (7) подтверждается наибольшим относительным количеством (процентом) «рассеянных» данных $\{Z_j = (x_j, y_j, z_j), j = \overline{1, L \cdot N_k}\}$ (6). В настоящей работе используется алгоритм MLESAC (Maximum Likelihood Estimation Sample Consensus – консенсус выборок с оценкой по максимуму правдоподобия) (Торг, 2000) – вероятностная версия алгоритма RANSAC. Этот алгоритм оценивает правдоподобие модели (7), представляя распределение расстояния «рассеянных» данных $\mathbb{Z} = \{Z_j = (x_j, y_j, z_j), j = \overline{1, L \cdot N_k}\}$ от модели $\mathfrak{M}(\boldsymbol{\theta})$ (7), как смесь распределения данных, подтверждающих модель (7) (inliers), и распределения данных, отклоняющих эту модель (outliers). Считая, что «рассеянные» данные \mathbb{Z} (6) независимые, получаем выражение для логарифма правдоподобия в виде

$$\begin{aligned} \mathcal{L}(\rho_{\perp}(\mathbb{Z}, \mathfrak{M}(\boldsymbol{\theta})) \mid \boldsymbol{\theta}) &= \\ &= \sum_{j=1}^{L \cdot N_k} \log \left(\gamma p(\rho_{\perp}(Z_j, \mathfrak{M}(\boldsymbol{\theta})) \mid Z_j \text{ is inlier}) + (1 - \gamma) p(\rho_{\perp}(Z_j, \mathfrak{M}(\boldsymbol{\theta})) \mid Z_j \text{ is outlier}) \right), \end{aligned} \quad (8)$$

где γ – параметр смешивания. Распределение расстояний до данных, подтверждающих модель (7), представляется гауссовым распределением

$$p(\rho_{\perp}(Z_j, \mathfrak{M}(\boldsymbol{\theta})) \mid Z_j \text{ is inlier}) \propto \exp \left(-\frac{(\rho_{\perp}(Z_j, \mathfrak{M}(\boldsymbol{\theta})))^2}{2\sigma^2} \right), \quad (9)$$

где σ – стандартное отклонение. Распределение расстояния до данных, отклоняющих модель (7), описывается равномерным распределением

$$p(\rho_{\perp}(Z_j, \mathfrak{M}(\boldsymbol{\theta})) \mid Z_j \text{ is outlier}) = \begin{cases} (2\rho_{max})^{-1}, & \rho_{\perp}(Z_j, \mathfrak{M}(\boldsymbol{\theta})) < \rho_{max}, \\ 0, & \rho_{\perp}(Z_j, \mathfrak{M}(\boldsymbol{\theta})) \geq \rho_{max}, \end{cases} \quad (10)$$

где ρ_{max} – наибольшее расстояние до данных (определяется контекстом). Минимизация логарифма правдоподобия (8) позволяет оценить вектор параметров $\boldsymbol{\theta}$ и параметр смешивания γ . Это обеспечивается итерациями EM-алгоритма.

Определённая таким образом наиболее правдоподобная линейная регрессия «рассеянных» данных выборки траекторий определяет геометрическую асимптоту $\mathfrak{M}(\boldsymbol{\theta})_k$, $k = \overline{1, K}$ (7) одного из пучков при условии (1) (см. Рис.1).

На Рис. 1 (а) показаны «рассеянные» данные (6) двумерных проекций траекторий рассматриваемой выборки и результат их линейной регрессии с использованием алгоритма MLESAC (Рис. 1), определяющий асимптоту первого пучка (голубая линия) (см. Рис. 1).

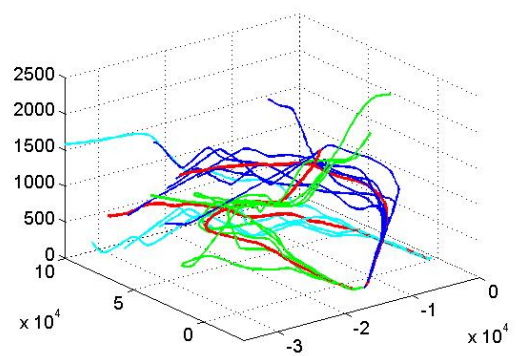
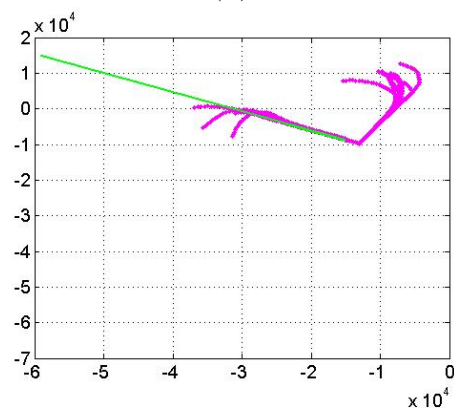
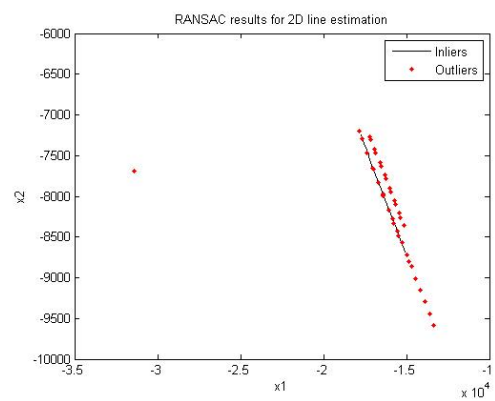
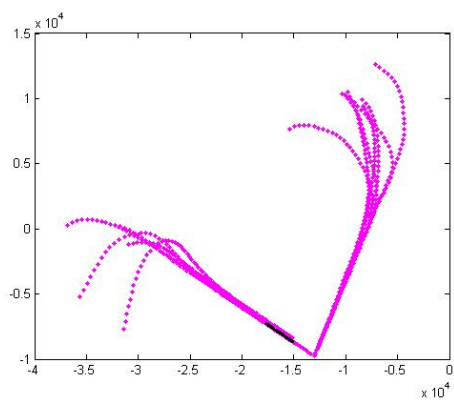
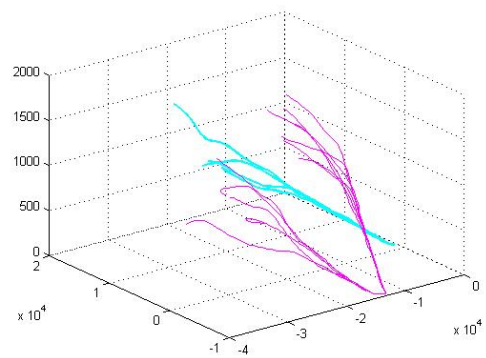
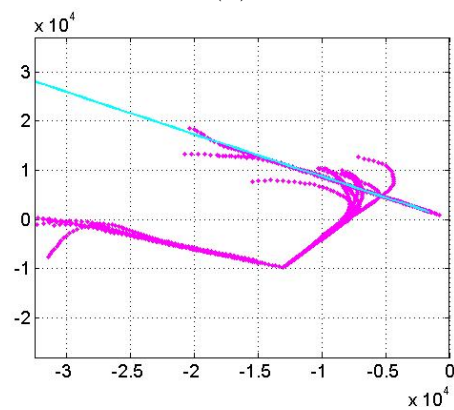
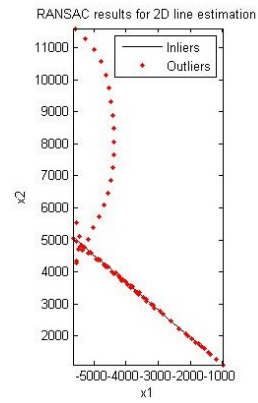
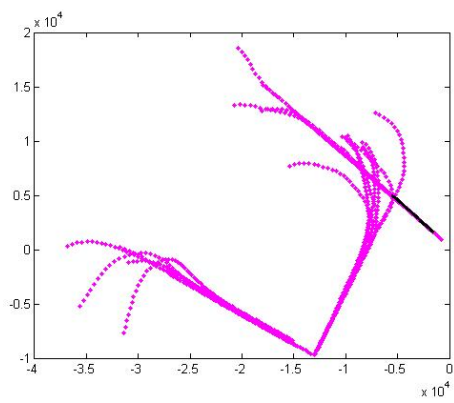


Рис. 1: Иллюстрация этапов определения геометрической асимптоты пучка траекторий (а-в, д-ж), которая определяется наиболее правдоподобной линейной регрессией данных

После выделения асимптоты касательный ей пучок траекторий определяется на основе однократного применения формулы (5) с мерой косинуса (3.1).

После удаления из «рассеянных» данных (6) тех точек, которые представляют траектории выделенного пучка, повторяется процедура определения геометрической асимптоты и выделяется следующий пучок траекторий. Поскольку определение модели (7) должно быть симметричным относительно координат x, y, z , при формировании «рассеянных» данных оставшихся траекторий в (6) производится сортировка по очередной пространственной координате отличной от использованной в (6) при определении предыдущей асимптоты (7). Возможная зависимость результата (7) от направления координат устраняется изменением направления сортировки в (6) с возрастания на убывание или наоборот. Анализ траекторий выборки завершается определением всех пучков в выборке (см. Рис. 2).

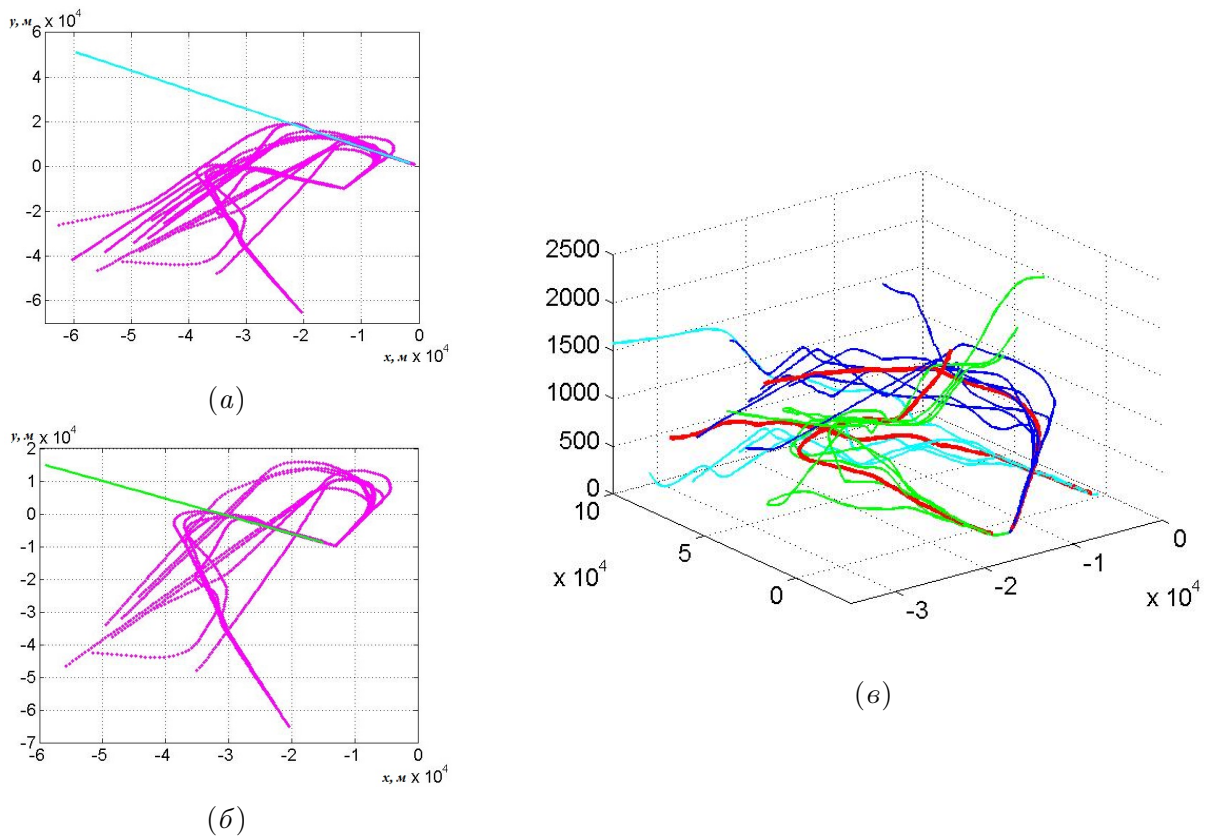


Рис. 2: Траектории анализируемой выборки.
 а,б – «рассеянные» данные траекторий и их линейные регрессии;
 в – три выделенных пучка траекторий

Множественные пересечения и наличие почти линейных участков в хвостах пучков траекторий вдали от фокуса может препятствовать непосредственному определению асимптот сходящихся пучков по методу, описанному выше. В таком случае при анализе траекторий рассматриваемой выборки используются «рассеянные» данные сокращённых траекторий с частью точек $quatum \approx 0.4$, считая от фокусов пучков.

Алгоритм устойчивого разбиения набора траекторий на пучки траекторий, асимптотически сходящиеся с заданным параметром порога

Представленные результаты показывают, что поставленная задача выделения в трёхмерном случае установившихся потоков движения — пучков траекторий, представляющих собой посадки самолётов на заданную ВПП с одного направления, может быть решена с помощью алгоритма, состоящего из семи шагов (см. Рис. 3).

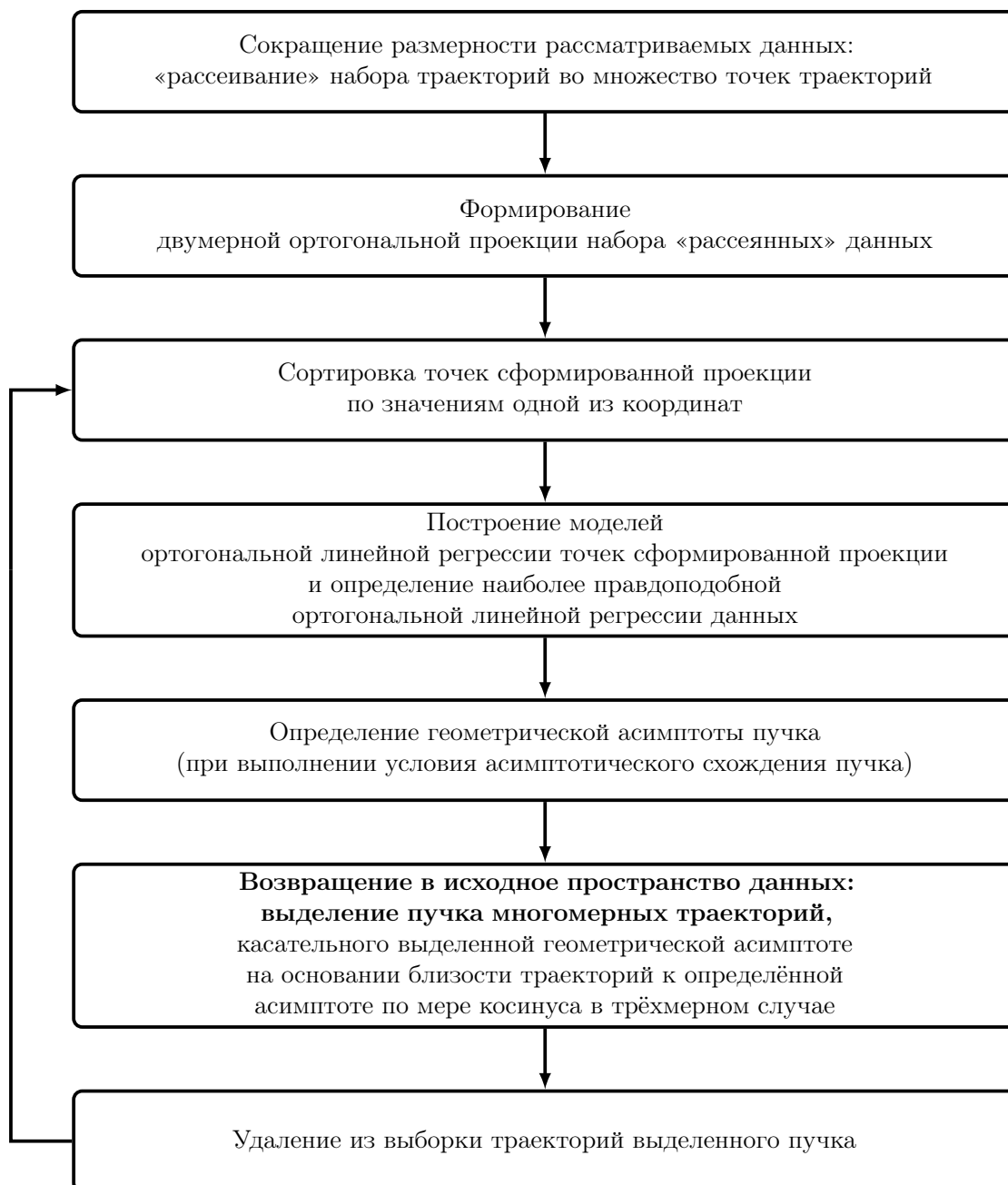


Рис. 3: Алгоритм устойчивого разбиения выборки посадочных траекторий на пучки

Данный алгоритм позволяет последовательно выделить в рассматриваемой выборке все асимптотически сходящиеся с параметром порога пучки многомерных траекторий,

соответствующие посадкам на заданные взлётно-посадочные полосы (ВПП). Параметр порога определяется шириной ВПП. Получаемое разбиение является устойчивым и не требует никакого предварительного знания о числе пучков, содержащихся в выборке.

Оптимизационная задача с ограничениями по определению центраида для выделенного пучка многомерных траекторий движения

После того, как точно выделены пучки траекторий \mathbb{C}_k , их центроиды $\{\mathbf{c}_k, k = \overline{1, K}\}$ определяются по схеме $\mathbb{C}_k \Rightarrow \mathbf{c}_k, k = \overline{1, K}$ на основе однократного применения формулы

$$\{\langle \mathbf{c}_k \rangle, k = \overline{1, K}\} = \arg \min_{\{\mathbf{c}_k, k = \overline{1, K}\}} \sum_{n=1}^N \sum_{k=1}^K r_{nk} \rho_{\cosine}^2(\mathbf{X}_n, \mathbf{c}_k) \quad (11)$$

при условии параметра порога с использованием квадрата меры косинуса (3.1).

Оценка (11) эффективна при представлении векторов $\{\mathbf{X}_n, n = \overline{1, N}\}$ и $\{\mathbf{c}_k, k = \overline{1, K}\}$ в исходном пространстве состояний. Центроиды, определяемые таким образом, выделены красным представлены на Рис. 4.

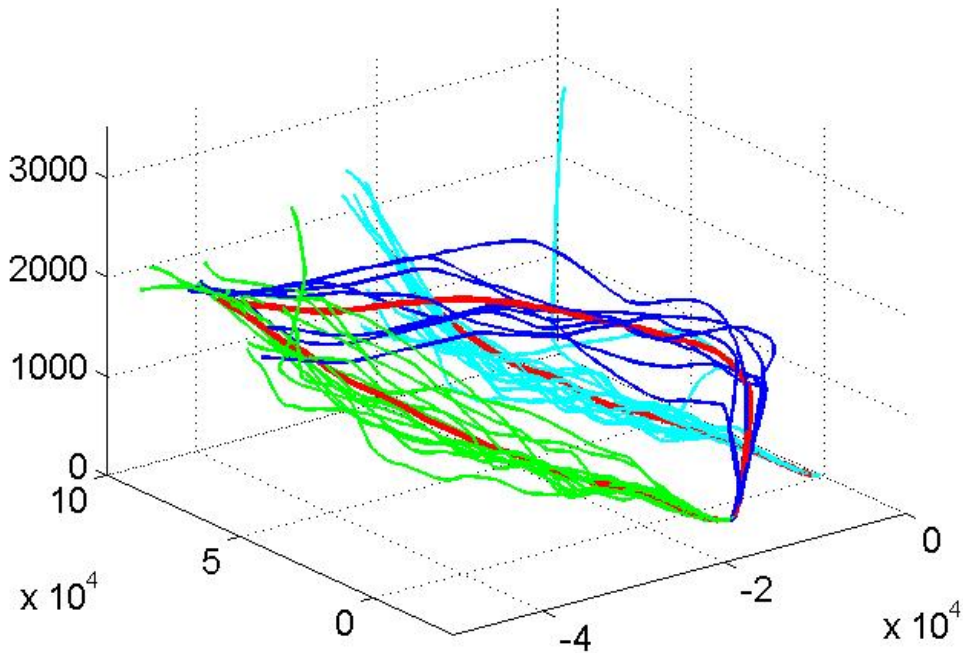


Рис. 4: Центроиды выделенных пучков траекторий анализируемой выборки (выделены красным)

Алгоритм определения характерной траектории потока – центроида для выделенного пучка многомерных траекторий движения

Представленные результаты показывают, что поставленная задача определения характерной траектории потока на примере анализа данных посадочных траекторий самолетов, может быть решена с помощью алгоритма, состоящего из трёх шагов (см. Рис. 5).

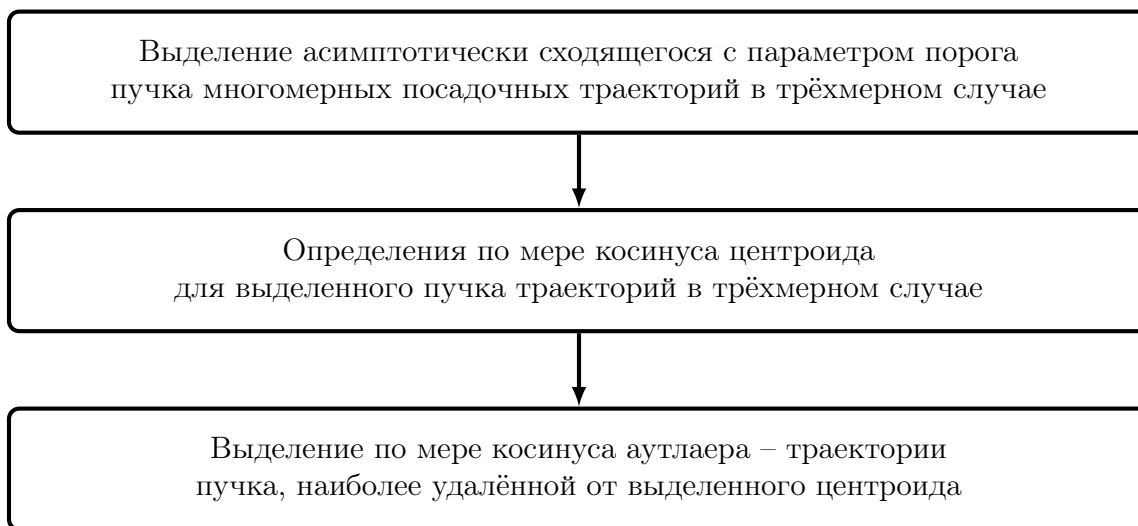


Рис. 5: Алгоритм определения центроида – характерной траектории потока движения

Шаг 1. На первом шаге выделяется пучок траекторий, соответствующий посадкам самолетов на выделенную взлётно-посадочную полосу, траектории в котором асимптотически сходятся с параметром порога.

Шаг 2. На втором шаге в выделенном пучке траекторий при решении оптимизационной задачи определяется центральная траектория, называемая в контексте работы центроидом, при этом суммарное среднеквадратичное отклонение всех точек траекторий в пучке от соответствующих точек центроида минимально.

Шаг 3. На третьем шаге по мере косинуса определяется аутлаер – траектория пучка, суммарное среднеквадратичное отклонение всех точек которой от соответствующих точек центроида максимально по мере косинуса. При решении задач безопасности можно говорить, что выделенная траектория представляет движение в зоне риска.

В **Заключении** приведены основные результаты работы.

Публикации автора по теме диссертации

Публикации в журналах, входящих в список ВАК:

1. Кухаренко Б.Г., Солнцева-Чалей М.О. Использование методов выделения ярких областей при сегментировании телеметрических изображений для идентификации групп объектов // *Информационные технологии*. — 2016. — Т. 22, № 7. — С. 504–509.
2. Кухаренко Б.Г., Солнцева-Чалей М.О. Моделирование центроидов для пучков многомерных траекторий // *Информационные технологии*. — 2016. — Т. 22, № 2. — С. 83–89.
3. Кухаренко Б.Г., Солнцева-Чалей М.О. Спектральный метод с использованием полярной кривизны для анализа результатов кластеризации многомерных траекторий // *Информационные технологии*. — 2015. — Т. 21, № 12. — С. 901–905.
4. Кухаренко Б.Г., Солнцева М.О. Анализ результатов кластеризации многомерных траекторий посредством моделей линейных динамических систем // *Информационные технологии*. — 2015. — № 2. — С. 104–109.
5. Кухаренко Б.Г., Солнцева М.О. Применение моделей нелинейных динамических систем для анализа результатов кластеризации многомерных траекторий // *Информационные технологии*. — 2015. — Т. 21, № 5. — С. 341–345.
6. Кухаренко Б.Г., Солнцева М.О. Применение моделей непрерывного скрытого профиля для анализа результатов кластеризации многомерных траекторий // *Информационные технологии*. — 2015. — Т. 21, № 8. — С. 585–590.
7. Кухаренко Б.Г., Солнцева М.О. Использование методов сокращения фона при сегментировании телеметрических изображений для идентификации групп объектов // *Информационные технологии*. — 2014. — № 2. — С. 3–8.
8. Кухаренко Б.Г., Солнцева М.О. Итеративная кластеризация траекторий управляемых объектов в многомерном пространстве характеристик // *Информационные технологии*. — 2014. — № 8. — С. 11–16.
9. Кухаренко Б.Г., Солнцева М.О. Кластеризация управляемых объектов на основе сходства их многомерных траекторий // *Информационные технологии*. — 2014. — № 5. — С. 3–7.
10. Кухаренко Б.Г., Солнцева М.О. Покомпонентный дискриминантный анализ результатов кластеризации многомерных траекторий // *Информационные технологии*. — 2014. — № 11. — С. 3–7.

11. *Кухаренко Б.Г., Солнцева М.О.* Использование методов сокращения фона при сегментировании телеметрических изображений для идентификации групп объектов // *Информационные технологии.* — 2014. — № 2. — С. 3–8.
12. *Кухаренко Б.Г., Солнцева М.О.* Принцип минимальной длины описания при анализе графов с разреженными матрицами смежности в задачах кластеризации их узлов // *Информационные технологии.* — 2013. — № 7. — С. 37–42.

Публикации в других журналах и тезисы конференций:

1. *Solntseva-Chaley, M.* New data mining technique for multidimensional aircraft trajectories analysis // International Conference on Big Data and its Applications (ICBDA 2016), ITM Web Conf. — 2016 — P. 1–7,01001, V. 8, <http://dx.doi.org/10.1051/itmconf/20160801001>
2. *Солнцева М.О., Кухаренко Б.Г.* Кластеризация управляемых объектов на основе сходства их траекторий и скоростных режимов // Труды МФТИ, — 2014. — Т. 6, № 3. — С. 65–72.
3. *Солнцева М.О., Кухаренко Б.Г.* Кластеризация многомерных траекторий управляемых объектов // Труды 57-й научной конференции МФТИ — 2014 г. — Т.2 — С.123-124.
4. *Солнцева М.О., Кухаренко Б.Г.* Кластеризация многомерных посадочных траекторий самолётов // Труды 57-й научной конференции МФТИ — 2014 г. — С.111-113.
5. *Солнцева М.О., Кухаренко Б.Г.,* Сокращение неинформативной части панорамного кадра по методу швов при управлении группами объектов // Сборник научных трудов МФТИ. — 2013 г. — С.91–97.
6. *Солнцева М.О., Кухаренко Б.Г.* Применение методов кластеризации узлов на графах с разреженными матрицами смежности в задачах логистики // Труды МФТИ — 2013 г. — Т.5, №3. — С.75-83.
7. *Кухаренко, Б.Г. and Солнцева, М.* Сокращение времени сегментирования телеметрических изображений при идентификации групп объектов за счёт удаления пикселений фона. // Труды 56-й научной конференции МФТИ. — 2013 — Т. 2, С. 129–130.
8. *Кухаренко Б.Г., Солнцева М.О.* Оптимизация расположения логистических центров на графе транспортной сети с помощью алгоритмов кластеризации // Труды 55-й научной конференции МФТИ — 2012 г. — Т.2 — С.128-129.

Патенты:

1. *Солнцева-Чалей М.О.* Способ последовательного определения усреднённых траекторий движения материальных объектов в трёхмерном пространстве. №2017101343 – 2018 г.
2. *Kukhareenko B.G, Solntseva-Chaley M.O.* Stable partition of trajectories set into asymptotically converged beams, <http://www.freepatentsonline.com/y2017/0084180.html> US Patent 20170084180A1 – 2017 г.
3. *Кухаренко Б.Г., Солнцева-Чалей М.О.* Способ определения посадочной траектории летательного аппарата на основании данных о зарегистрированных траекториях с применением в качестве меры сходства траекторий меры косинуса (варианты). №2616107 – 2017 г.
4. *Кухаренко, Б.Г. and Солнцева-Чалей, М.О.,* Способ определения посадочных траекторий летательных аппаратов, соответствующих одной взлётно-посадочной полосе. № 2616106 – 2017 г.