Санкт-Петербургский государственный университет

Кафедра компьютерных технологий и систем

Шавонина Ксения Сергеевна

Выпускная квалификационная работа бакалавра

Непрерывная модель в задаче выбора последовательности измерений

Направление 010400

Прикладная математика и информатика

Научный руководитель,

кандидат физ.-мат. наук,

доцент

Чашникова В. В.

Санкт-Петербург

2016

Содержание

Введение

Задача определения и прогноза движения объекта по результатам наблюдений стоит перед человечеством с древних времен. Еще в Древней Греции античные философы и ученые наблюдали за движением Солнца, Луны и других объектов на небе. Долгое время задача определения движения небесных тел оставалась одной из первостепенных задач науки. В XV-XVIII веках совместными усилиями Коперника, Кеплера и Ньютона была получена математическая модель Солнечной системы. Однако основные параметры этой системы (массы небесных тел, их координаты и составляющие скоростей) не измерялись напрямую. Для их определения использовались угловые координаты небесных объектов, которые получались в процессе астрономических наблюдений. Так появилась задача определения параметров сложной реальной системы по результатам наблюдений за ее движением.

В настоящее время эта задача находит применение в различных сферах научной и повседневной деятельности. С помощью вычисления задержки сигнала от спутников определяются координаты объектов на земной поверхности. Задачи рассматриваемого типа ставятся в геодезии, картографии, физике, биологии и социологии.

Для эффективного решения таких задач в 1794 году Гауссом был изобретен специальный математический аппарат – метод наименьших квадратов, который достаточно хорошо себя зарекомендовал. С появлением мощной вычислительной техники, а также появившейся потребностью определять параметры состояния системы в процессе управления движением, стали развиваться статистические методы обработки информации. Однако вследствие недостаточной надежности результатов, полученных таким способом, все чаще стал применяться минимаксный подход.

Следует заметить, что при применении любого метода получается не точное состояние системы, а лишь некоторая оценка. При использовании метода наименьших квадратов любое дополнительное измерение улучшает (или по крайней мере не ухудшает) точность получаемой оценки при условии некоррелированности получаемых измерений. Однако чаще всего имеется ограничение на количество возможных измерений, таким образом, возникает задача правильного распределения этих измерений между имеющимися источниками наблюдения с целью увеличения точности получаемой оценки состояния системы.

В статическом случае, при известном измерительном базисе, решение задачи распределения измерений между объектами наблюдения дает метод Эльвинга. Однако на практике возникают задачи, когда и сами наблюдаемые объекты изменяют свое положение во времени. В этом случае задача становится в разы сложнее, так как оптимальный измерительный базис меняется с течением времени, и возникает потребность определить не только количество измерений с каждого источника, но также и последовательность снятия измерений. Исследованию этого вопроса и посвящена данная работа.

В первой главе приведены два основных подхода к решению задачи определения параметров системы по результатам измерений. Во второй главе рассмотрена комбинация этих методов, а также переход к непрерывной модели для динамического случая с одним переключением между наблюдаемыми объектами. Также в третьем параграфе второй главы предложен алгоритм нахождения оптимального количества переключений при ограничении их числа и вычисления моментов времени смены объекта наблюдения. В третьей главе алгоритм проиллюстрирован на примере задачи спутниковой навигации.

Постановка задачи

Пусть имеется некоторая динамическая система, состояние которой необходимо определить. Предположим, что состояние этой системы в любой момент времени может быть описано конечным m-мерным вектором параметров q – вектором состояния данной системы. Для того чтобы определить вектор q, производят измерения некоторых величин, которые зависят от состояния самой системы. Пусть d – n-мерный вектор измерений, состоящий из всех используемых измерений (n – число измерений) и связанный с вектором состояния соотношением:

$$d = F(q). \tag{1}$$

В общем случае зависимость $F(q)$ нелинейная и $n > m$.

При разложении функции $F(q)$ в ряд Тейлора получается:

$$F(q) = F(q_0) + A(q - q_0) + [(q - q_0)^T B_i (q - q_0)]_{i=1,\dots,n},$$

где q_0 - некоторое приближенное опорное значение вектора q,

$A = \left[\frac{\partial f_i}{\partial q_j}\right]_{i=1,\dots,n;\, j=1,\dots,m}$ – матрица частных производных, вычисленная в точке q_0,

$B_i = \begin{bmatrix} \frac{\partial^2 f_i}{\partial q_1^2} & \cdots & \frac{\partial^2 f_i}{\partial q_1 \partial q_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f_i}{\partial q_m \partial q_1} & \cdots & \frac{\partial^2 f_i}{\partial q_m^2} \end{bmatrix}$ – матрица Гессе, вычисленная в некоторой точке $\tilde{q}$ такой, что $||\tilde{q} - q_0|| < ||q - q_0||$.

Пусть $d_0 = F(q_0)$. Пусть также нормы матриц B_i настолько малы, чтобы определить линейную систему

$$A(q - q_0) = d - d_0.$$

В общем случае элементы матрицы A зависят от времени: $A = A(t)$.

Пусть имеется M объектов наблюдения, с которых можно снимать измерения, и пусть количество возможных измерений ограничено числом N. Предполагается, что ошибки измерений некоррелированны. Задача состоит в том, чтобы определить, с каких источников и в какой последовательности необходимо снимать измерения при условии:

$$\sum_{i=1}^{M} m_i = N,$$

где m_i - количество измерений, отвечающее источнику i, чтобы получить оценку вектора состояния максимально близкую к реальному состоянию системы.

Можно рассмотреть линейную функцию оценки вектора состояния

$$l = cq, \tag{2}$$

где c – заданный вектор с постоянными коэффициентами, и в качестве критерия оптимальности рассматривать дисперсию данной функции.

Таким образом задача сводится к минимизации максимальной дисперсии функции l.

Обзор литературы

К настоящему моменту представлено не так много литературы, посвященной задачам рассматриваемого типа. В большинстве существующих работ рассматриваются статистические методы решения задачи нахождения параметров по результатам измерений. В работе [1] широко освящен метод максимального правдоподобия, а также рассмотрены вопросы состоятельности, условий несмещенности и эффективности оценок, полученных по этому методу. Основные критерии оценивания, а также ряд статистических методов может быть найден в работе [2]. Метод наименьших квадратов подробно описан в книгах [3], [4] и [5], а его прикладная сторона и некоторые примеры можно найти в работе [6]. Также в [5] отдельная глава посвящена минимаксному подходу получения оценок.

В [7] Эльвингом описан метод, комбинирующий статистический и минимаксный подходы. В трудах [8] и [9] изложены методы оптимального планирования эксперимента, которые основаны на указанных выше методах. Однако вся рассмотренная выше литература представлена для статических случаев. Переход к непрерывной модели для динамического случая впервые был предложен в статье [10] и лег в основу данной работы.

Глава 1 Два подхода к определению параметров системы по результатам измерений

Точное определение состояния системы, как уже говорилось выше, невозможно в силу многих факторов. Компоненты вектора измерения получаются в результате наблюдения за реальной системой, однако принятая математическая модель может быть неточной и описывать систему с некоторой погрешностью. Поэтому имеют место ошибки метода:

$$\eta = d_T - F(q_T),$$

где d_T – истинное значение вектора измерения, q_T – истинное значение вектора состояния системы.

Также непосредственно при самом измерении получаются неточные значения параметров вектора измерений, таким образом, появляется вектор ошибок измерений:

$$\xi = \tilde{d} - d_T,$$

где $\tilde{d}$ – полученное в результате измерений значение вектора d.

В рассмотрение вводится система условных уравнений:

$$F(q) = \tilde{d}.$$

Алгоритм получения оценки вектора состояния $\hat{q}$ по вектору измерений $\tilde{d}$ носит название алгоритма фильтрации:

$$\hat{q} = \Phi\left(\tilde{d}\right). \qquad (3)$$

Его задача – уменьшение влияния указанных выше ошибок.

Любой алгоритм фильтрации должен обладать следующими свойствами:

1. Зависимость (3) должна быть однозначной.
2. Должно выполняться условие несмещенности:

$$q = \Phi\big(F(q)\big),$$

т. е. если ошибки измерения или ошибки модели отсутствуют, то он должен давать истинное значение вектора состояния.

Точность полученного результата характеризует вектор погрешности:

$$\nu = \hat{q} - q_T.$$

Оценка $\hat{q}$ называется несмещенной, если справедливо:

$$E(\nu) = 0 \Leftrightarrow E(\hat{q}) = q_T$$

при некоторых условиях, наложенных на методические и измерительные ошибки.

1.1. Метод наименьших квадратов

Пусть теперь ξ - это суммарная ошибка, объединяющая измерительные и методические погрешности. Тогда получается равенство:

$$\tilde{d} = F(q_T) + \xi.$$

Имеются следующие предположения:

1. Определена математическая модель (1).
2. Известно математическое ожидание $E(\xi)$.
3. Матрица ковариации $D(\xi)$ задана с точностью до некоторого произвольного множителя и положительно определена.

Пусть имеется линейная математическая модель:

$$Aq = d. \quad (4)$$

В общем случае система является несовместной, так как число неизвестных меньше числа уравнений. В таком случае нужно отыскать алгоритм, который бы минимизировал норму вектора невязки:

$$\delta = A\hat{q} - \tilde{d},$$

где $\hat{q}$ - оценка вектора состояния.

Получается система:

$$A\hat{q} = \tilde{d}.$$

В [11] показывается, что норма вектора невязки минимизируется теми и только теми векторами $\hat{q}$, которые удовлетворяют системе:

$$A^T A\hat{q} = A^T \tilde{d}. \quad (5)$$

Если $rank(A) = m$, то система (5) имеет единственное псевдорешение:

$$\hat{q} = A^{+}\tilde{d}, \tag{6}$$

где A^{+} - псевдообратная матрица к матрице A, которая находится как:

$$A^{+} = (A^{T}A)^{-1}A^{T}.$$

Если $rank(A) < m$, то тогда:

$$A^{+} = SQA^{T}T^{-1},$$

где S – матрица координат векторов сингулярного базиса в пространстве состояний, T – матрица координат векторов сингулярного базиса в пространстве измерений, Q – матрица вида:

$$Q = \begin{pmatrix} \frac{1}{\rho_1^2} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\rho_r^2} & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 1 \end{pmatrix},$$

ρ_i – сингулярные числа матрицы $A,\ r = rank(A)$.

Таким образом, полученная псевдообратная матрица A^{+} – это и есть искомый алгоритм фильтрации. Если обозначить $\Phi = A^{+}$, то из формул (2) и (6) получается следующий результат:

$$\hat{l} = c\hat{q} = c\Phi\tilde{d} = X\tilde{d},$$

где

$$X = c\Phi. \tag{7}$$

Ковариационная матрица для оценки функции $\hat{l}$ принимает вид дисперсии случайной величины:

$$D(\hat{l}) = XD(\xi)X^{T}.$$

Согласно теореме Гаусса-Маркова, метод наименьших квадратов обеспечивает минимум дисперсии оценки произвольной линейной функции вектора q на множестве всех линейных несмещенных оценок при использовании линейной математической модели (4) и принятых допущениях 1-3, описанных выше. Важным следствием теоремы Гаусса-Маркова и очевидным плюсом метода наименьших квадратов является его универсальность, которая заключается в одновременном получении оптимальной оценки всех однородных линейных функций вектора состояния, то есть $\Phi = \Phi(A)$ и, следовательно, Ф никаким образом не зависит от выбора вектора c. Еще одно следствие заключается в том, что оптимально использовать все составляющие вектора измерений d [5].

К минусам же данного метода можно отнести то, что при некоторых условиях он дает завышенно оптимистичные оценки точности получаемых результатов.

Метод наименьших квадратов относится к статистическим методам и является частным случаем метода максимального правдоподобия. Более подробно статистические методы обработки измерительной информации описаны в работах [1] и [2].

1.2. Гарантирующий подход

Наряду с рассмотренным выше классическим подходом – методом наименьших квадратов – широкое распространение получил неклассический подход к решению данной задачи, который называется гарантирующим. Оценки состояния системы, получаемые с помощью данного метода, носят название минимаксных.

Множество M возможных ошибок ξ и критерий оптимальности задаются следующим способом:

1) Про составляющие ξ_i (i = 1, 2, …, n) вектора ошибок измерений ξ ничего не известно кроме того, что они ограничены по модулю некоторой положительной постоянной α:

$$|\xi_i| \leq \alpha, \qquad i = 1, 2, \ldots, n.$$

Если ввести обозначение:

$$\chi = |\hat{l} - l|,$$

то в качестве гарантированной характеристики точности используется значение:

$$\chi_{max} = \max_{\xi \in M} \chi.$$

2) Нахождение $\hat{l}$, которая обеспечит минимум величины χ_{max}, является целью алгоритма фильтрации при гарантирующем подходе. Также, не нарушая общности, все составляющие ξ_i, $i = 1, \dots, n$, вектора ошибок измерения ξ считаются равноточными.

Пусть рассматриваемая система описывается линеаризованной моделью (4) и имеется линейная функция l, определяемая формулой (2). Также задано измеренное значение вектора $\tilde{d}$ и ошибка ξ удовлетворяет указанным выше положениям 1-2. Согласно теореме, изложенной в [5], среди всех линейных оценок вида

$$\hat{q} = Y\tilde{d}, \qquad \hat{l} = c\hat{q} = X\tilde{d}, \qquad X = cY,$$

удовлетворяющих условию несмещенности

$$YA = I, \qquad XA = c,$$

может быть найдена по крайней мере одна оптимальная по указанному выше критерию оценка $\hat{q} = (\hat{q}_1, \dots, \hat{q}_m)^T$, получаемая в результате решения системы m линейных уравнений

$$\tilde{d}_{(m)} = A_{(m)} \hat{q}_{(m)},$$

где $\tilde{d}_{(m)} = \left(\tilde{d}_1, \dots , \tilde{d}_m\right)^T$ – некоторая достаточная совокупность измерений, выбираемая оптимальным образом из составляющих вектора $\tilde{d}$, а $A_{(m)}$ – квадратная матрица, составленная из m строк, соответствующих в матрице A выбранным измерениям $\tilde{d}_1, \dots, \tilde{d}_m$.

Данная теорема решает вопрос как о выборе оптимального алгоритма фильтрации, так и о выборе оптимального измерительного базиса, так как все составляющие вектора $\tilde{d}$, не входящие в состав вектора $\tilde{d}_{(m)}$, могут быть исключены из дальнейшего рассмотрения без ухудшения точности получаемой оценки [5].

Компоненты вектора $X = (x_1, \dots, x_n)$ – это не что иное, как коэффициенты разложения вектора c по строкам матрицы A, т.е.:

$$c = x_1 A_1 + x_2 A_2 + \dots + x_n A_n.$$

Таким образом, задача сводится к отысканию вектор-строки X такой, чтобы она обеспечивала решение следующей задачи абсолютной минимизации:

$$\sum_{i=1}^{n} |x_i| \rightarrow min,$$

при условии несмещенности $XA = c$.

В [12] показано, что решение этой задачи эквивалентно решению некоторой задачи линейного программирования. В [13] был предложен следующий способ сведения к задаче линейного программирования.

Можно заменить искомый вектор X вектором U длины $2n$ и на его компоненты ввести ограничение:

$$u_i \geq 0, \qquad i = 1, \dots, 2n.$$

Матрица A заменяется матрицей B, состоящей из $2n$ строк и получаемой из матрицы A приписыванием снизу матрицы $-A$. Условие несмещенности принимает вид $UB = c$, и решается следующая задача линейного программирования:

$$\sum_{i=1}^{2n} u_i \rightarrow min.$$

Ненулевые компоненты вектора U будут соответствовать выбранным базисным строкам матрицы A либо $-A$.

Способы решения данной задачи можно найти, например, в [14].

Также определить оптимальный измерительный базис можно с помощью решения, основанного на геометрической интерпретации. Особенно наглядной является геометрическая интерпретация для случаев, когда вектор состояния является двумерным или трехмерным. В двумерном случае строки матриц A и $-A$ рассматриваются как векторы на плоскости. По этим векторам строится выпуклая оболочка, и оставляются лишь те векторы, на которые эта выпуклая оболочка натянута. Вектор-строка c также изображается как вектор, находится пересечение этого вектора с границей выпуклой оболочки, и в качестве оптимального измерительного базиса берутся те вектора, на которые эта граница натянута. В трехмерном случае все делается аналогичным образом, лишь с тем отличием, что векторы изображаются в трехмерном пространстве, а в качестве границы выпуклой оболочки выступает поверхность многогранника. Математическое обоснование такого решения и различные случаи подробно рассмотрены в [5].

Например, для матрицы

$$A = \begin{pmatrix} 3 & 4 \\ -2 & 1 \\ 5 & 1 \\ -1 & 4 \end{pmatrix}$$

и вектора $c = (2, 2)$ геометрическая интерпретация показана на рисунке 1.

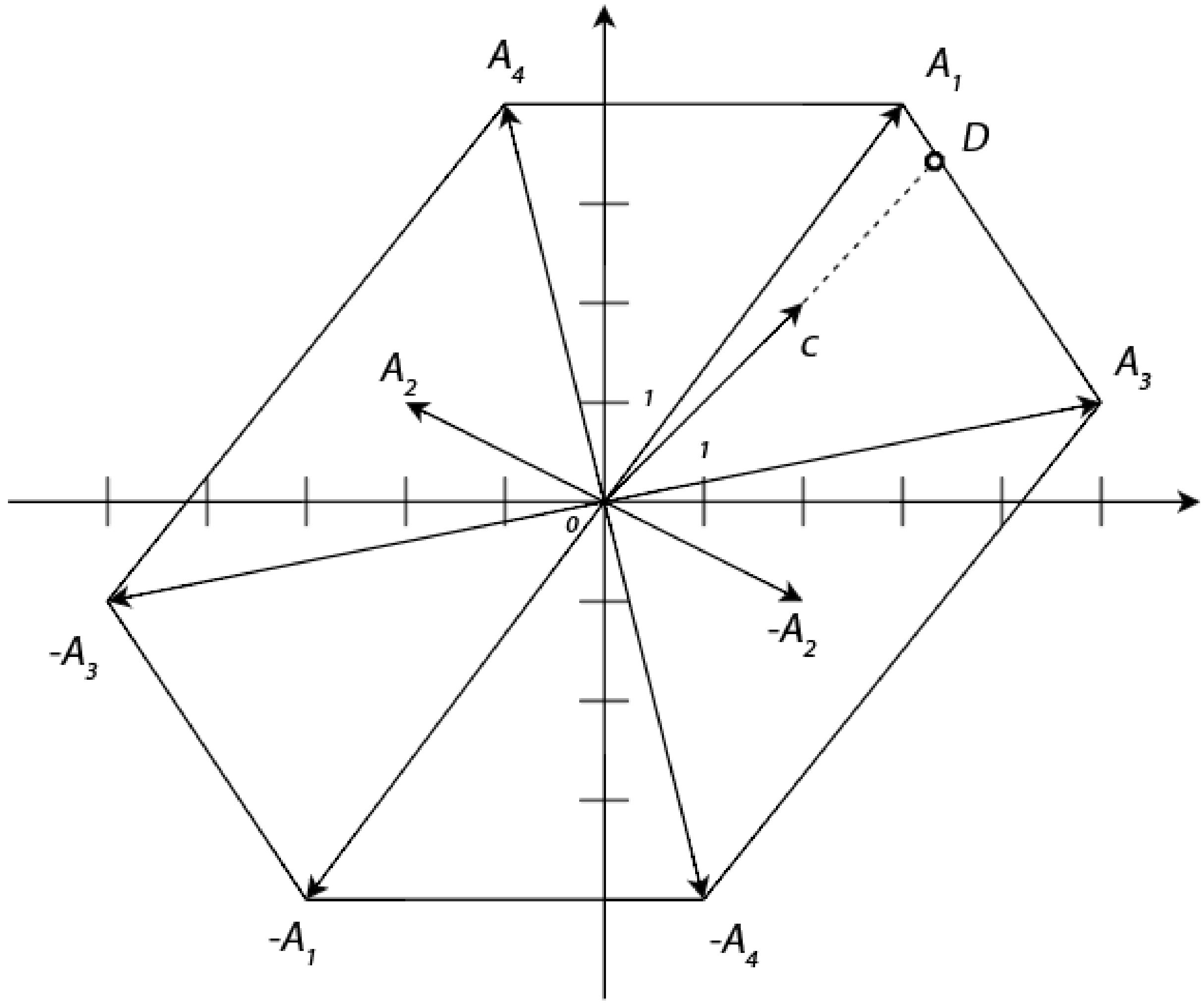

Рисунок 1

A_1, A_2, A_3, A_4 – строки матрицы A, а $-A_1, -A_2, -A_3, -A_4$ – это строки матрицы $-A$. Как видно из рисунка 1, вектора A_2 и $-A_2$ не достигают границы линейной оболочки. Вектор c пересекает границу линейной оболочки в точке D, которая принадлежит отрезку, соединяющему концы векторов A_1 и A_3. Таким образом, оптимальным будет выбор первой и третьей строки матрицы A.

Рассмотренный в этом параграфе метод всегда имеет решение, и размерность оптимального измерительного базиса не превышеет m [5]. В некоторых ситуациях этот метод дает несколько оптимальных решений, которые являются эквивалентными. Однако чаще всего в прикладных задачах достаточно найти какое-то одно оптимальное решение и ограничиться им.

Плюсом гарантирующего подхода является то, что он не требует знания вероятностных характеристик вектора ошибок, а также обеспечивает гарантированную оценку точности получаемых результатов. То есть по сравнению с методом наименьших квадратов данный метод является надежным, хоть и оценки состояния системы получаются очень грубыми. К существенным минусам этого метода можно отнести тот факт, что минимаксные оценки в общем случае не являются универсальными и зависят от выбора оцениваемого параметра (то есть выбора вектора c).

Глава 2 Оптимальный состав измерений

В предыдущей главе были описаны два различных подхода к решению поставленной задачи, каждый из которых имеет свои сильные и слабые стороны. В настоящей главе осуществляется переход к методам, которые получены комбинированием рассмотренных выше подходов с целью уменьшения влияния недостатков обоих. Получаемые результаты имеют привязку к уже конкретным источникам наблюдения, а также учитывается ограничение на число возможных измерений.

Имеется линейная или линеаризованная модель (4). Пусть с применением гарантирующего подхода найден некоторый оптимальный измерительный базис, и он состоит из M базисных строк матрицы A, каждой из которых соответствует свой источник наблюдения. Количество источников M, как говорилось выше, не превышает числа m размерности вектора состояния. Пусть число возможных измерений ограничено числом N. Нужно распределить N возможных измерений между имеющимися M источниками оптимальным образом.

2.1. Статический случай – задача Эльвинга

В [7] рассматривается следующий метод для случая, когда объекты наблюдения статичны. Пусть матрица A не меняется с течением времени. И пусть m_i – это количество измерений, которое нужно снять с источника i, $i = 1, \ldots, M$. Таким образом, выполняется равенство:

$$\sum_{i=1}^{M} m_i = N.$$

Вводится следующее обозначение:

$$p_i = \frac{m_i}{N}, \qquad i = 1, \ldots, M.$$

Очевидно, что выполняются следующие два условия:

$$0 \le p_i \le 1, \qquad \sum_{i=1}^{M} p_i = 1. \tag{8}$$

Оценка $\hat{q}$ вектора состояния системы определяется по методу наименьших квадратов, а оценка параметра l из формулы (2) находится:

$$\hat{l} = c\hat{q}.$$

Дисперсия оценки $\hat{l}$ может быть найдена по формуле:

$$D(\hat{l}) = \frac{\sigma^2}{N} \sum_{i=1}^{M} \frac{x_i^2}{p_i},$$

где x_i - это компоненты вектор-строки X, которая находится с помощью формулы (7) из предыдущей главы. Также считается, что все измерения независимые и некоррелированные.

Критерием оптимальности является достижение минимума величины $D(\hat{l})$, поэтому нужно решить следующую задачу:

$$\min_p \sum_{i=1}^{M} \frac{x_i^2}{p_i}.$$

Можно воспользоваться методом Лагранжа:

$$f(p) = \sum_{i=1}^{M} \frac{x_i^2}{p_i} + \lambda\left(\sum_{i=1}^{M} p_i - 1\right),$$

где λ - неопределенный множитель Лагранжа. Следовательно, необходимые условия минимума:

$$\frac{\partial f}{\partial p_i} = -\frac{x_i^2}{p_i^2} + \lambda = 0, \qquad i = 1, \dots, M.$$

Учитывая первое условие из (8):

$$p_i = \frac{|x_i|}{\sqrt{\lambda}}, \qquad i = 1, \dots, M. \tag{9}$$

При подстановке этих выражений во второе условие (8) получается выражение для знаменателя формулы (9):

$$\sqrt{\lambda} = \sum_{i=1}^{M} |x_i|$$

Таким образом, окончательно формула (9) принимает вид:

$$p_i = \frac{|x_i|}{\sum_{i=1}^{M} |x_i|}, \qquad i = 1, \dots, M.$$

При заданной строке X это решение единственное. Так как оно найдено из необходимых условий минимума, то достаточность такого решения следует из существования отыскиваемого минимума, которое доказано в [5]. Если теперь каждое p_i, $i = 1, \dots, M$, умножить на N, то будут найдены искомые m_i.

2.2. Непрерывная модель. Одно переключение

Теперь можно перейти к непрерывной модели для динамического случая, которая описана в работе [10]. Пусть теперь $q \in \mathbb{R}^2$ и компоненты матрицы A меняются с течением времени. Имеется несколько объектов, с которых можно снимать измерения, и получаемые измерения изменяют свои значения на отрезке $t \in [0, 1]$. Пусть два из них являются оптимальными для снятия измерений на всем рассматриваемом интервале, назовем эти объекты A и B. Это обозначает, что выполняется следующее условие:

$$\prod_{k=1}^{n} x_k(t) \prod_{i=m+1}^{n} \left(1 - \sum_{j=1}^{m} h_{ij}(t)\right) \neq 0, \qquad t \in [0,1],$$

где $x_k(t)$ – это k-ая компонента вектора c в базисе выбранных строк матрицы A, а h_{ij} – это -ая компонента i-ой строки базиса [15].

Объекту A соответствует строка $[a_1(t)\; a_2(t)]$, а объекту B строка $[b_1(t)\; b_2(t)]$. Пусть на интервале $[0, 1]$ можно провести N измерений. Необходимо определить целое число $m \in \{1, \dots, N\}$, отвечающее моменту времени $t = \frac{m}{N}$, в который нужно переключиться с наблюдения объекта A на наблюдение объекта B. В качестве критерия оптимальности снова рассматривается минимум дисперсии оценки величины l, задаваемой формулой (2).

Матрица S, отвечающая линеаризованной системе, определяется как:

$$S = \begin{bmatrix} a_1\left(\frac{1}{N}\right) & a_2\left(\frac{1}{N}\right) \\ \vdots & \vdots \\ a_1\left(\frac{m}{N}\right) & a_2\left(\frac{m}{N}\right) \\ b_1\left(\frac{m+1}{N}\right) & b_2\left(\frac{m+1}{N}\right) \\ \vdots & \vdots \\ b_1(1) & b_2(1) \end{bmatrix}.$$

Таким образом, получается связь $d = Sq$, где d – N-мерный вектор наблюдений. Пусть $D(d) = \sigma^2 I$ – ковариационная матрица. Вектор состояния q может быть получен с помощью псевдообратной матрицы S^+:

$$q = S^+ d, \qquad \text{где } S^+ = (S^T S)^{-1} S^T.$$

Тогда ковариационная матрица ошибки измерения принимает вид:

$$D(q) = S^+ D(d)(S^+)^T = \sigma^2 (S^T S)^{-1} =$$

$$= \frac{\sigma^2}{S_1^T S_1 S_2^T S_2 - (S_1^T S_2)^2} \begin{bmatrix} [S_2^T S_2] & [-S_1^T S_2] \\ [-S_1^T S_2] & [S_1^T S_1] \end{bmatrix},$$

где S_1 и S_2 - это столбцы матрицы S.

Пусть c = [0, 1], тогда $l = q_2$. Дисперсия:

$$D(l) = D(q_2) = cD(q)c^T = \frac{\sigma^2 S_1^T S_1}{S_1^T S_1 S_2^T S_2 - (S_1^T S_2)^2} =$$

$$= \frac{\sigma^2}{N} \left(\frac{\frac{1}{N} S_1^T S_1}{\frac{1}{N} S_1^T S_1 \frac{1}{N} S_2^T S_2 - (\frac{1}{N} S_1^T S_2)^2} \right).$$

Справедливы соотношения:

$$S_1^T S_1 = \sum_{k=1}^{m} a_1^2\left(\frac{k}{N}\right) + \sum_{k=m+1}^{N} b_1^2\left(\frac{k}{N}\right),$$

$$S_2^T S_2 = \sum_{k=1}^{m} a_2^2\left(\frac{k}{N}\right) + \sum_{k=m+1}^{N} b_2^2\left(\frac{k}{N}\right),$$

$$S_1^T S_2 = \sum_{k=1}^{m} a_1\left(\frac{k}{N}\right) a_2\left(\frac{k}{N}\right) + \sum_{k=m+1}^{N} b_1\left(\frac{k}{N}\right) b_2\left(\frac{k}{N}\right).$$

Если функции $a_1(t)$, $a_2(t)$, $b_1(t)$, $b_2(t)$ непрерывные и монотонно неубывающие, то выполняются неравенства:

$$\left|\frac{1}{N} S_1^T S_1 - \int_0^{\frac{m}{N}} a_1^2(t)dt - \int_{\frac{m}{N}}^{1} b_1^2(t)dt\right| \leq$$

$$\leq \frac{1}{N}\left|a_1^2\left(\frac{m}{N}\right) + b_1^2(1) - a_1^2(0) - b_1^2\left(\frac{m+1}{N}\right)\right|,$$

$$\left|\frac{1}{N} S_2^T S_2 - \int_0^{\frac{m}{N}} a_2^2(t)dt - \int_{\frac{m}{N}}^{1} b_2^2(t)dt\right| \leq$$

$$\leq \frac{1}{N}\left|a_2^2\left(\frac{m}{N}\right) + b_2^2(1) - a_2^2(0) - b_2^2\left(\frac{m+1}{N}\right)\right|,$$

$$\left|\frac{1}{N} S_1^T S_2 - \int_0^{\frac{m}{N}} a_1(t)a_2(t)dt - \int_{\frac{m}{N}}^{1} b_1(t)b_2(t)dt\right| \leq$$

$$\leq \frac{1}{N}\left|a_1\left(\frac{m}{N}\right) a_2\left(\frac{m}{N}\right) + b_1(1)b_2(1) - a_1(0)a_2(0)\right.$$

$$\left. - b_1\left(\frac{m+1}{N}\right) b_2\left(\frac{m+1}{N}\right)\right|,$$

Пусть

$$J_1(p) = \int_0^{p} a_1^2(t)dt + \int_p^{1} b_1^2(t)dt,$$

$$J_2(p) = \int_0^{p} a_2^2(t)dt + \int_p^{1} b_2^2(t)dt,$$

$$J_3(p) = \int_0^p a_1(t)a_2(t)dt + \int_p^1 b_1(t)b_2(t)dt.$$

Если N достаточно большое, то можно рассматривать непрерывную модель и минимизировать следующую функцию:

$$f(p) = \frac{\sigma^2}{N}\left(\frac{J_1(p)}{J_1(p)J_2(p) - J_3^2(p)}\right).$$

Если $p^* = \operatorname{argmin} f(p)$ на интервале $[0,1]$, то ищется дробь $\frac{m}{N}$, ближайшая к p^*. Найденное значение m будет искомым.

Пусть, например, имеются следующие данные:

$$a_1(t) = t,$$

$$a_2(t) = t^2,$$

$$b_1(t) = \frac{1}{2}t,$$

$$b_2(t) = \frac{1}{2}t + 1,$$

$$N = 1000.$$

Оцениваться будет вторая координата, т.е. $c = [0,1]$.

Функция $f(p)$ выглядит следующим образом:

$$f(p) =$$

$$= -\frac{\frac{p^3}{4} + \frac{1}{12}}{\left(\frac{1}{12}\left((p-1)(p+2)^2\right) - \frac{p^4}{4}\right)^2 + \left(\frac{p^3}{4} + \frac{1}{12}\right)\left(\frac{1}{12}\left((p-1)(p^2+7p+19)\right) - \frac{p^5}{5}\right)}$$

и на промежутке от нуля до единицы имеет график, представленный на рисунке 2.

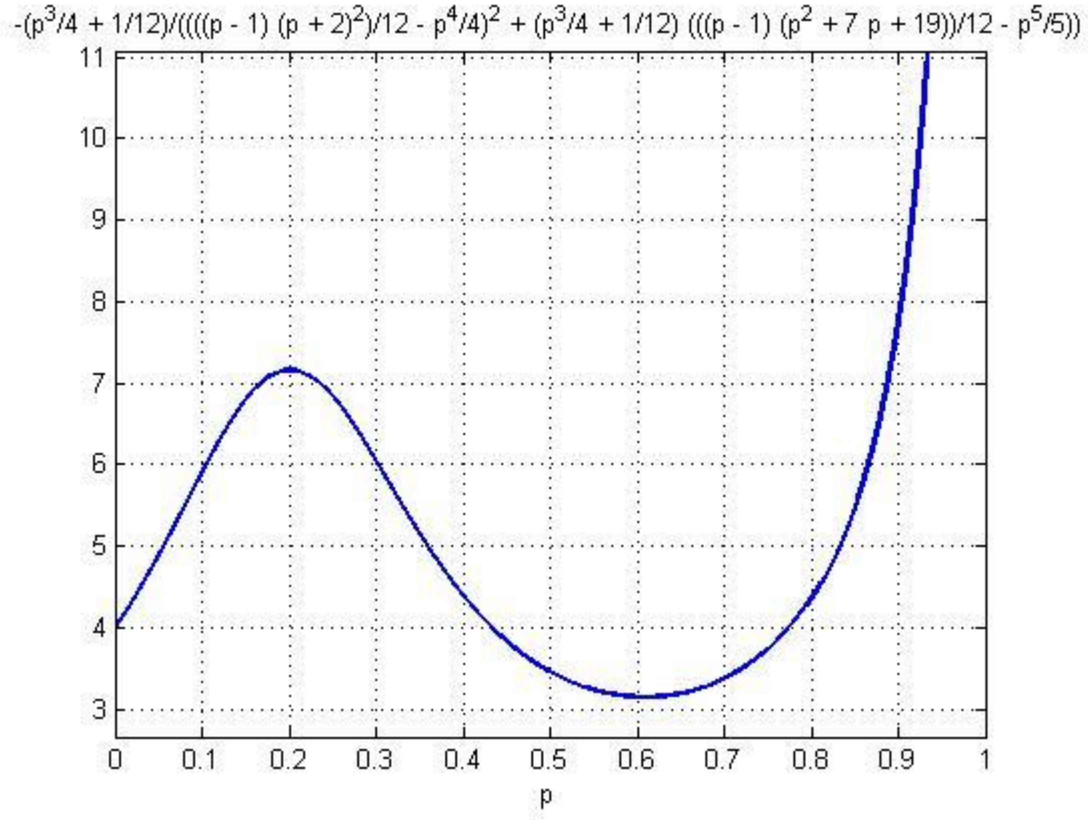

Рисунок 2

Эта функция имеет минимум в точке $p^* = 0.60858$, и, следовательно, переключение с первого объекта на второй необходимо произвести после произведения 609 измерения. В этом случае дисперсия по второй координате будет равна $D(q_2) = 0.003149$. Для сравнения, если первые 500 измерений произвести с первого объекта, а вторые – со второго, то дисперсия составит $D(q_2) = 0.003455$, то есть будет больше.

Пусть теперь:

$$a_1(t) = t^4 - 9,$$

$$a_2(t) = t - 98,$$

$$b_1(t) = \sin t + 1,$$

$$b_2(t) = t^3 + t,$$

$$N = 1000.$$

Функция $f(p)$ имеет очень сложный вид. Ее график представлен на рисунке 3.

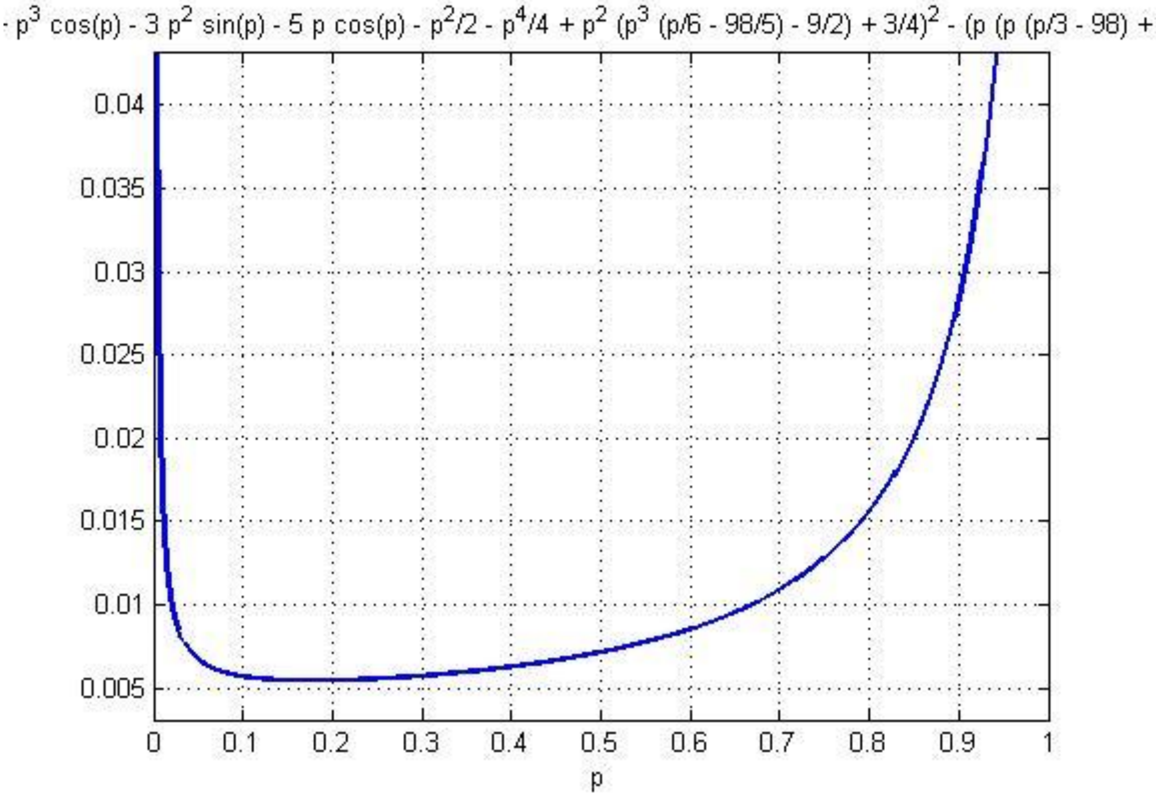

Рисунок 3

Минимум достигается в точке $p^* = 0.17697$, переключение нужно производить после 177 измерения. Дисперсия по второй координате составляет $D(q_2) = 5.44 * 10^{-6}$.

2.3. Непрерывная модель. Несколько переключений

Пусть теперь возможно более одного переключения между наблюдаемыми объектами. Так как в матрице S имеется N строк, то максимальное количество переключений необходимо в том случае, когда $S^i = [a_1(t)\ a_2(t)]$, когда i – нечетное, и $S^i = [b_1(t)\ b_2(t)]$, когда i – четное, где $i = 1, \dots, N$ (S^i - строки матрицы S). Тогда максимально возможное количество переключений равняется $N - 1$.

Пусть, для определенности, N – четное. Тогда функции J_1, J_2, J_3 можно ввести следующим образом:

$$J_1(p) = \int_0^{p_1} a_1^2(t)dt + \int_{p_1}^{p_2} b_1^2(t)dt + \dots + \int_{p_{N-2}}^{p_{N-1}} a_1^2(t)dt + \int_{p_{N-1}}^{1} b_1^2(t)dt,$$

$$J_2(p) = \int_0^{p_1} a_2^2(t)dt + \int_{p_1}^{p_2} b_2^2(t)dt + \dots + \int_{p_{N-2}}^{p_{N-1}} a_2^2(t)dt + \int_{p_{N-1}}^{1} b_2^2(t)dt,$$

$$J_3(p) = \int_0^{p_1} a_1(t)a_2(t)dt + \int_{p_1}^{p_2} b_1(t)b_2(t)dt + \dots$$

$$+ \int_{p_{N-2}}^{p_{N-1}} a_1(t)a_2(t)dt + \int_{p_{N-1}}^{1} b_1(t)b_2(t)dt,$$

где $p = (p_1, p_2, \dots, p_{N-1})$.

То есть функция $f(p)$ теперь зависит от $N-1$ переменных $p_1, \dots, p_{N-1}$.

Пусть $p^* = \operatorname{argmin} f(p) = (p_1^*, p_2^*, \dots, p_{N-1}^*)$ при условии, что $p_1 \le p_2 \le \dots \le p_{N-1}$ и $p_i \in [0,1]$, $i = 1, \dots, N-1$. Для каждого значения p_i^* существует ближайшая дробь $\frac{m_i}{N}$, $i = 1, \dots, N-1$. При нахождении всех m_i получается вектор $m = (m_1, m_2, \dots, m_{N-1})$. Если все его компоненты различны, то это искомый вектор.

Пусть теперь в векторе m имеются повторяющиеся элементы. Пусть начиная с некоторого индекса k_1 и по индекс k_2, $k_1 < k_2$, элементы в векторе m повторяются. Можно расписать функцию J_1 следующим образом (J_2 и J_3 аналогично):

$$J_1(p) = \int_0^{\frac{1}{N}} a_1^2(t)dt + \int_{\frac{2}{N}}^{\frac{3}{N}} b_1^2(t)dt + \dots + \int_{\frac{m_{k1-1}}{N}}^{\frac{m_{k_1}}{N}} b_1^2(t)dt + \int_{\frac{m_{k1}}{N}}^{\frac{m_{k_{1+1}}}{N}} a_1^2(t)dt$$
$$+ \int_{\frac{m_{k_{1+1}}}{N}}^{\frac{m_{k_1+2}}{N}} b_1^2(t)dt + \int_{\frac{m_{k_1+2}}{N}}^{\frac{m_{k_1+3}}{N}} a_1^2(t)dt + \dots + \int_{\frac{m_{k_2-1}}{N}}^{\frac{m_{k_2}}{N}} b_1^2(t)dt$$
$$+ \int_{\frac{m_{k2}}{N}}^{\frac{m_{k_{2+1}}}{N}} a_1^2(t)dt + \dots + \int_{\frac{N-2}{N}}^{\frac{N-1}{N}} a_1^2(t)dt + \int_{\frac{N-1}{N}}^{1} b_1^2(t)dt.$$

Очевидно, что интегралы вида $\int_{\frac{m_i}{N}}^{\frac{m_j}{N}} a_1^2(t)dt$ или $\int_{\frac{m_i}{N}}^{\frac{m_j}{N}} b_1^2(t)dt$ равны нулю, так как в данном случае $m_i = m_j$ при всех $i, j = k_1, \dots, k_2$. Таким образом, их можно просто убрать.

Теперь пусть количество повторяющихся элементов четное. Тогда переключения не происходит. Действительно, пусть количество повторяющихся элементов равно 2, тогда $k_2 = k_1 + 1$. Можно рассмотреть несколько слагаемых из функции $J_1(p)$:

$$\int_{\frac{m_{k1-1}}{N}}^{\frac{m_{k_1}}{N}} b_1^2(t)dt + \int_{\frac{m_{k1}}{N}}^{\frac{m_{k_{1+1}}}{N}} a_1^2(t)dt + \int_{\frac{m_{k_{1+1}}}{N}}^{\frac{m_{k_1+2}}{N}} b_1^2(t)dt.$$

Так как средний интеграл равен нулю, то его можно убрать. Но подынтегральные функции первого и третьего интегралов одинаковые, то всю эту сумму можно заменить одним интегралом

$$\int_{\frac{m_{k1-1}}{N}}^{\frac{m_{k_{1+2}}}{N}} b_1^2(t)dt.$$

Таким образом, переключения между объектами не произошло, то есть данные элементы вектора m можно убрать. Аналогичным образом доказывается для случая четного числа больше двух. Если же имеется нечетное количество повторяющихся элементов, то аналогично показывается, что можно убрать соответствующие интегралы, но уже переключение между наблюдаемыми объектами будет иметь место.

Вектор $\mu = (\mu_1, \dots, \mu_r)$, составленный из r элементов вектора m, которые повторяются нечетное число раз, будет искомым.

Можно взять задачу с начальными данными из примеров предыдущего параграфа и рассмотреть случай двух переключений. Сначала описанным выше способом будут получены моменты переключения с одного объекта на другой, а затем будет произведено снова 1000 измерений, которые произведутся в найденной последовательности и соотношении. По полученной матрице S будет найдена дисперсия.

Для первого примера минимум функции $f(p_1, p_2)$ достигается при:

$$p_1^* = 0, \qquad p_2^* = 0.745,$$

здесь 0 обозначает тот факт, что оптимально будет начать измерения не с первого объекта, а сразу со второго, и лишь после 745 измерения переключиться на первый объект. Дисперсия в таком случае будет:

$$D(q_2) = 0.00145,$$

что уже значительно меньше, чем в предыдущем случае.

Для второго примера

$$p_1^* = 0.1137, \qquad p_2^* = 0.9118.$$

Значит нужно провести 114 измерений с первого объекта, затем переключиться на второй, провести еще 798 измерений, а затем, суммарно проведя 912 измерений, снова переключиться на первый объект и провести еще 88 измерений. Дисперсия составит:

$$D(q_2) = 5.34 * 10^{-6},$$

то есть меньше, чем в случае одного переключения.

Глава 3 Выбор оптимального состава измерений в задачах спутниковой навигации

3.1. Некоторые теоретические сведения

Прежде чем перейти к рассмотрению конкретных примеров, нужно привести основные системы координат, используемые в спутниковой навигации, а также способы перехода от одной системы координат к другой [16].

I. Геоцентрическая экваториальная прямоугольная.

Начало координат – центр масс Земли. Ось X направлена в точку весеннего равноденствия, ось Z направлена на север и совпадает с осью вращения Земли, ось Y дополняет систему до правой тройки.

II. Гринвичская прямоугольная.

Начало координат – центр масс Земли. Ось x направлена в точку пересечения экватора и гринвичского меридиана, ось z направлена на север и совпадает с осью вращения Земли, ось y дополняет систему до правой тройки.

III. Гринвичская сферическая.

Начало координат – центр масс Земли.

r – модуль радиус-вектора,

ψ – угол между плоскостью экватора и радиус-вектором (широта),

λ – угол между плоскостью гринвичского меридиана и радиус-вектором (долгота).

Следующая система координат применяется для искусственных спутников Земли.

IV. Оскулирующая система координат.

i – угол между плоскостью орбиты и плоскостью экватора (наклонение орбиты),

Ω – долгота восходящего узла,

ω – аргумент перигея,

a – большая полуось орбиты,

e – эксцентриситет орбиты,

τ – время прохождения спутника через перицентр,

ω – угловая скорость движения спутника по орбите.

Если известны координаты в какой-то одной из описанных выше систем координат, то получить координаты в любой другой системе можно с помощью следующих формул перехода.

От I к II:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = A^{-1} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix},$$

где

$$A = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\gamma = S_0 + \widetilde{\omega}(t - t_0),$$

где S_0 – звездное время в среднюю гринвичскую полночь для данной даты, $\widetilde{\omega}$ – угловая скорость вращения Земли, t и t_0 – среднее солнечное время.

От II к III:

$$r = \sqrt{x^2 + y^2 + z^2},$$

$$\psi = \arcsin\left(\frac{z}{r}\right),$$

$$\lambda = arctg\frac{y}{x}.$$

От III к II:

$$x = r\cos\,\psi\cos\lambda,$$

$$y = r\cos\,\psi\sin\lambda,$$

$$z = r\sin\,\psi.$$

От IV к I:

$$X = R(\cos \Omega \cos u - \sin \Omega \sin u \cos i),$$
$$Y = R(\sin \Omega \cos u - \cos \Omega \sin u \cos i),$$
$$Z = R \sin u \sin i,$$

где $u = \omega + \text{w}(t - \tau)$, t – текущий момент времени.

Одним из методов, используемых для определения координаты точки на Земле при помощи спутников, является дальномерный метод, основанный на следующей навигационной формуле:

$$\rho = \sqrt{(x - x_s)^2 + (y - y_s)^2 + (z - z_s)^2}, \tag{10}$$

где (x_s, y_s, z_s) – это координаты спутника в гринвичской прямоугольной системе координат, а (x, y, z) – координаты некоторой приближенной точки на Земле, называемой счислимым местом.

В этом случае компонентами строк матрицы S будут являться частные производные данной функции по координатам ψ и λ, вычисляемые по следующим формулам:

$$\frac{\partial \rho}{\partial \psi} = \frac{r}{\rho}(x_s \cos \lambda \sin \psi + \ y_s \sin \lambda \sin \psi - z_s \cos \psi\ ), \tag{11}$$

$$\frac{\partial \rho}{\partial \lambda} = \frac{r}{\rho} \cos \psi\ (x_s \sin \lambda - y_s \cos \lambda\ ). \tag{12}$$

Однако для того, чтобы была возможность измерять расстояние до спутника, он должен находиться в зоне видимости. Это гарантируется выполнением условия:

$$(\vec{r}_s - \vec{r}, \vec{r}) > 0,$$

где $\vec{r}_s$ – радиус-вектор спутника, $\vec{r}$ – радиус-вектор счислимого места.

Абсолютное время находится по формуле:

$$t = \left[\sum_{i=1}^{month} p_i + date - 81 \right]_{mod365} * 1440 + (hour - t_0) * 60 + min,$$

где date – это день, month – месяц, hour – час, min – минута, t_0 – сдвиг в часах относительного гринвичского времени, p_i – коэффициенты вектора:

$$p = [0, 31, 28, 31, 30, 31, 30, 31, 31, 30, 31, 30].$$

3.2. Анализ результатов на примере системы спутников

В качестве примера можно рассмотреть систему из шести спутников с координатами в оскулирующей системе координат. Все эти спутники имеют орбиты радиусом $R = 15000$ км, эти орбиты являются круговыми, поэтому эксцентриситет $e = 0$, наклон орбит составляет $i = 60$ градусов, аргумент перигея $\omega = 0$, угловая скорость $\text{ɯ} = 2$. Долгота восходящего узла и время прохождения каждого спутника через перицентр представлены в таблице 1.

№	Ω (градусы)	τ (минуты)
1	0	30
2	45	100
3	90	70
4	135	0
5	210	150
6	300	120

Таблица 1

В качестве счислимого места будут выступать координаты Санкт-Петербурга в гринвичской сферической системе координат: $R = 6300$ – радиус Земли, $\psi = 60°$ – широта, $\lambda = 30°$ – долгота. 25 мая 2016 года в 10:00 в зоне видимости будут находиться первый и четвертый спутники. Первый спутник будет наблюдаем в течение 1 часа и 5 минут с рассматриваемого момента, четвертый – в течение 4 минут.

Координаты спутников меняются в зависимости от времени t. После преобразования координат спутников сначала в экваториальную прямоугольную, затем в гринвичскую прямоугольную систему координат, а также преобразования сферических координат счислимого места в прямоугольные, можно определить вид навигационной функции (10), а также найти ее производные по формулам (11) и (12). Таким образом, будут найдены строки $[a_1(t)\ a_2(t)]$ и $[b_1(t)\ b_2(t)]$, отвечающие спутникам A (первый спутник) и B (четвертый спутник) соответственно. Они будут иметь следующий вид:

$a_1(t) =$ (1.7716*10^7*cos(u1) - 1.02274*10^7*sin(u1) - 4.09233*10^7*sin(u2) + 5.31574*10^7*cos(u3) + 3.06877*10^7*sin(u3))/((3749.5*cos(u1) + 11250.5*cos(u3) - 2728.41)^2 + (3749.5*sin(u1) - 11250.5*sin(u3) + 1575.11)^2 + (12989.8*sin(u2) - 5455.72)^2)^(1/2),

$a_2(t) =$ -(1.0*(1.14097*10^7*cos(u1) - 3058400.0*sin(u1) - 2.50611*10^7*cos(u3) + 2.5064*10^7*sin(u3)))/((3749.5*cos(u1) + 11250.5*cos(u3) - 2728.41)^2 + (3749.5*sin(u1) - 11250.5*sin(u3) + 1575.11)^2 + (12989.8*sin(u2) - 5455.72)^2)^(1/2),

$b_1(t) =$-(1.0*(1.77136*10^7*cos(u1) + 4.09233*10^7*sin(u2) - 1.02315*10^7*sin(u1) + 5.31503*10^7*cos(u3) + 3.07*10^7*sin(u3)))/(2.04576*10^7*cos(u1) - 0.00000253962*cos(u4) - 1.41737*10^8*sin(u2) + 0.00000101877*sin(u4) - 1.18164*10^7*sin(u1) + 6.13836*10^7*cos(u3) + 3.54556*10^7*sin(u3) + 2.6469*10^8)^(1/2),

$b_2(t) =$-(6.00895*10^(-26)*(9.83237*10^31*cos(u1) + 1.70226*10^32*sin(u1) + 2.95024*10^32*cos(u3) - 5.10768*10^32*sin(u3)))/(2.04576*10^7*cos(u1) - 0.00000253962*cos(u4) - 1.41737*10^8*sin(u2) + 0.00000101877*sin(u4) - 1.18164*10^7*sin(u1) + 6.13836*10^7*cos(u3) + 3.54556*10^7*sin(u3) + 2.6469*10^8)^(1/2),

u1 = 0.0399519*t - 1.83246,
u2 = 0.034904*t - 1.04712,
u3 = 0.0298561*t - 0.26178,
u4 = 0.069808*t.

Времени 10:00 25 мая соответствует абсолютное время $t = 92580$ минут. Пусть производится измерение расстояния до спутников раз в три секунды на протяжении трех минут, то есть всего будет произведено 60 измерений. Пусть возможно два переключения между наблюдаемыми объектами. Функции $f(p_1, p_2)$ достигает своего минимума при значениях:

$$p_1 = 0.6787, \qquad p_2 = 1.$$

Это говорит о том, что необходимости во втором переключении нет, так как момент времени p_2 соответствует 60 измерению, после которого наблюдение спутников прекращается. Таким образом получается, что нужно снять 41 измерения с первого спутника, а затем переключиться на второй. В этом случае дисперсия по второй координате, то есть по долготе, составит $D(\lambda) = 2.9911 * 10^{-9}$.

Выводы

1. Произведено исследование различных подходов к решению задачи определения параметров системы по результатам наблюдений, их комбинации, а также перехода к непрерывному случаю.
2. Предложен метод нахождения нескольких моментов переключения между наблюдаемыми объектами, а также проиллюстрировано его преимущество перед методом с одним переключением.
3. Непрерывная модель для динамического случая проиллюстрирована на примере определения координат точки на земной поверхности с помощью системы спутников.

Заключение

Таким образом, несмотря на существующий к настоящему времени мощный математический аппарат, а также увеличивающиеся с каждым годом технические возможности, проблема определения параметров системы по данным, получаемым в результате наблюдений, остается не до конца изученной. Результаты, полученные в ходе выполнения данной работы, могут послужить основой для дальнейших исследований. В частности, остается открытым вопрос о связи оптимального количества переключений с видом функций, которые соответствуют наблюдаемым объектам.

Список литературы

1. Кендалл М., Стюарт А. Статистические выводы и связи. Наука, 1973.
2. Рао С. Р. Линейные статистические методы и их применение. Наука, 1968.
3. Ющенко А. П. Способ наименьших квадратов. Изд. Морской транспорт, 1956.
4. Чеботарев А. С. Способ наименьших квадратов с основами теории вероятностей. ОНТИ, 1936.
5. Эльясберг П. Е. Определение движения по результатам измерений. Наука, 1976.
6. Линник Ю. В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. Государственное издательство физико-математической литературы, 1958.
7. Зуховицкий С. И., Авдеева Л. И. Линейное и выпуклое программирование. Наука, 1964.
8. Федоров В. В. Теория оптимального эксперимента. Наука, 1971.
9. Kiefer J. Omtimum experimental designs // Actes du congrés international des mathématiciens, Nice, France, T. 3, Septembre 1970.
10. V. Chashnikova. Choosing a measuring basis to estimate parameters of a dynamic system // International conference ICNAAM 2015, Rhodos, Greece, 23-30 September 2015.
11. Воеводин В. В. Линейная алгебра. Наука, 1980.
12. Лидов М., Л. К априорным оценкам точности определения параметров по методу наименьших квадратов // Космические исследования, Т. 2, Вып. 5, 1964.
13. Эльясберг П. Е., Бахшиян Б. Ц. Определение траектории полета космического аппарата при отсутствии сведений о законе распределения ошибок измерений // Космические исследования, Т. 7,

Вып. 1, 1969.

14. Юдин Д. Б., Гольдштейн Е. Г. Линейное программирование. Наука, 1969.
15. V. Tchashnikova. Optimal Choice of Observed Values for C2-class Functions // International Conference of Informatics and Control, St. Peterburg, 1997.
16. Чуров Е. П., Суворов Е. Ф. Космические средства судовождения. Транспорт, 1979.