

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ НАБЛЮДЕНИЙ

О. А. Титов

(Санкт-Петербургский Университет, кафедра астрономии)

Санкт-Петербург

2001

О. А. Титов

Данное пособие предназначено для студентов старших курсов астрономического отделения математико-механического факультета СПбГУ. В нем последовательно излагаются метод наименьших квадратов и его последующие модификации.

© Санкт-Петербургский Государственный Университет, 2001

Введение

Создание новых алгоритмов обработки высокоточных наблюдений в настоящее время является одной из основных проблем, возникающих перед исследователем, независимо от того, в какой области науки он работает. Развитие новых технических средств, применение новых наблюдательных методик, компьютеризация – все это способствует повышению точности измерений. Но для того, чтобы реализовать эту точность, методы обработки также должны постоянно совершенствоваться, “подстраиваясь” под решение той или иной проблемы.

Метод наименьших квадратов (МНК) был развит в начале XIX века в трудах Ж. Лежандра (1806 г.) и К.Ф. Гаусса (1809 г.) и в течение многих лет применялся практически в первозданном виде. Начиная примерно с 40-х годов XX века, появляются многочисленные модификации этого метода. При этом появление новых технических средств и наблюдательных методик заставило математиков разрабатывать довольно сложные алгоритмы обработки информации. Так, например, создание ракетной техники в годы Второй мировой войны повлекло за собой разработку алгоритмов последовательной обработки данных – фильтра Калмана-Бьюси и его модификаций. [Браммер, Зиффлинг, 1982; Jazwinski, 1970] Поэтому можно сказать, что развитие технических средств способствовало (и способствует) развитию математического аппарата, предназначенного для обработки данных, с помощью этих средств полученных.

Радиоинтерферометрия со сверхдлинными базами (РСДБ) как новая наблюдательная техника также требует применения нестандартных алгоритмов [Ma, Ryan, 1995]. На это есть несколько причин, одна из которых состоит в том, что наблюдения проводятся одновременно на нескольких инструментах (иногда число радиотелескопов, задействованных, в эксперименте превышает 10), в то время как классические астрометрические наблюдения, как правило, проводятся на одном инструменте. С точки зрения обработки данных это означает, что РСДБ наблюдения в каждый момент времени дают несколько разностей $O-C$, т.е. вектор, в то время как классические виды наблюдений дают в каждый момент времени одну разность $O-C$, т.е. скаляр. Поэтому при обработке РСДБ данных возникают такие новые понятия, как, например, корреляции между наблюдениями,

полученными одновременно. Кроме того, практически с самого начала оказалось, что разности хода водородных стандартов испытывают большие флуктуации (опять сказывается специфика РСДБ наблюдений). Это заставило разбивать 24-часовой интервал наблюдений на сегменты продолжительностью 1-2 часа и оценивать параметры, характеризующие подобные флуктуации, на каждом сегменте отдельно. Этот метод получил название сегментированного или последовательного МНК [Ma et al., 1990]. Подобную процедуру пришлось использовать и для оценивания флуктуаций тропосферной задержки в зените. В результате алгоритм оценивания в целом оказался достаточно громоздким. Кроме того, выяснилось, что нельзя делать сегмент слишком коротким (менее 1 часа) из-за того, что точность оценок начинает ухудшаться. Снижение точности связано с тем, что при уменьшении интервала разбиения внутри каждого сегмента попадает недостаточное количество наблюдений. А поскольку корреляции между параметрами в соседних сегментах не учитываются, то разброс оценок от сегмента к сегменту оказывается неправдоподобно большим. Поэтому в качестве альтернативы последовательному МНК начал применяться фильтр Калмана, учитывающий возможные изменения стохастических параметров с помощью выбранной динамической модели [Herring, Davis, Shapiro, 1990; Himwich et al; 1991]. Однако оказалось, что фильтр Калмана имеет свои недостатки, связанные с накоплением ошибок и появлением плохо обусловленных ковариационных матриц.

Все это привело к тому, что в настоящее время нет общепринятого алгоритма уравнивания данных РСДБ наблюдений. Различные исследовательские группы, создавая собственные пакеты программ, как правило, пишут заново программы оценивания параметров, выбирая алгоритм, который кажется наиболее подходящим, и внося в него при этом свои коррективы. В пакете GLORIA, например, вместо МНК был использован метод наименьших модулей, менее чувствительный к грубым ошибкам наблюдений. [Gontier, 1992; Charlot, 1993].

С середины 80-х годов возрос интерес исследователей к изучению внутрисуточных характеристик различных физических процессов [Brosche, 1982; Brosche et. al., 1989]. В геодинимике это, в первую очередь, относится к параметрам вращения Земли (ПВЗ), поскольку с их помощью можно уточнять модель строения Земли, характер взаимодействия ее внутренних слоев

(мантии, ядра), изучать различные процессы, происходящие в атмосфере и т.д. Особое внимание уделяется изучению суточных и полусуточных вариаций ПВЗ, вызванных влиянием приливов, хотя это не единственная причина появления высокочастотных колебаний. К сожалению, при обработке суточной серии РСДБ наблюдений обычно получают только одну оценку ПВЗ за 24 часа. Поэтому, несмотря на стремительное улучшение точности поступающей информации, высокочастотные характеристики, как правило, остаются неисследованными. В последнее время были сделаны попытки исправить положение, например, с помощью программы CALC/SOLVE были получены часовые оценки изменений координат полюса и UT1–UTC [Gipson, 1996]. Однако в целом проблема оценивания ПВЗ на внутрисуточном интервале времени остается нерешенной. Следует также отметить, что и другие параметры, которые определяются из РСДБ наблюдений, также могут претерпевать значительные флуктуации в течение 24-часового эксперимента. Например, совсем недавно появились доказательства влияния вариаций атмосферного давления на вертикальные деформации земной коры [vanDam, Herring, 1994; MacMillan, Gipson, 1994]. Очевидно, что быстрые изменения приповерхностного давления (при прохождении атмосферных фронтов) могут вызвать вертикальные смещения земной коры. Эти смещения (амплитуда до 1 см) достаточно заметны на современном уровне точности РСДБ наблюдений, и пренебрежение подобными эффектами может сказаться на точности оценок, которые получаются при уравнивании.

Метод среднеквадратической коллокации (МСКК) был разработан в 60-е годы, в основном, для представления гравитационного поля Земли и уравнивания глобальных геодезических сетей [Moritz, 1973; Мориц, 1980]. Однако оказалось, что его можно применять и как метод оценивания в тех случаях, когда некоторые определяемые параметры изменяются в течение интервала наблюдений, поскольку МСКК позволяет использовать более гибкую параметрическую модель данных наблюдений, чем МНК. "Среднеквадратическая коллокация есть обобщенный метод оценивания, который сочетает уравнивание, фильтрацию и прогноз. Кроме параметров рассматриваются два вида случайных переменных: ошибки измерений или шум, а также сигнал; эти случайные переменные связаны друг с другом через ковариации" – так определяется предложенный метод [Moritz, 1973]. Идея при-

менения МСКК для обработки РСДБ наблюдений оказалась очень перспективной, потому что, как уже было сказано выше, флуктуации шкал водородных стандартов и тропосферной задержки в зените не позволяют применять классическую схему МНК для решения данной задачи. Адаптация МСКК для обработки РСДБ наблюдений позволяет, несмотря на увеличение времени, получить несмещенные оценки постоянных параметров, исключая влияние стохастических. Кроме того, появилась принципиальная возможность проанализировать поведение на 24-часовом интервале времени тех параметров, которые традиционно считались постоянными, например, ПВЗ или координаты станций [Губанов, Титов, 1994; Титов, 1996].

Конечно, все модификации классической схемы МНК стали возможны лишь с развитием компьютеров. В докомпьютерную эпоху вычислительная техника была довольно примитивной, и обращение больших матриц было долгим и утомительным занятием. Вот как, например, это описывается в монографии Идельсона [1932]: *"Но прежде всего нужно коснуться самой техники вычисления. Составление и решение нормальных систем требует умножения и деления трехзначных и четырехзначных чисел, и эти действия следует производить либо по особым таблицам умножения, либо на арифмометре. Пользование логарифмами решительно не рекомендуется, поскольку ими загромаждается схема и требуется производство целого ряда действий "в уме". Логарифмическая линейка, вообще говоря, не обладает достаточной точностью и кроме того требует слишком напряженного внимания в отношении места запятой. Из численных таблиц наиболее подходящими являются известные таблицы Crelle, или Zimmermann'a, или О'Рурка. Но неоспоримое преимущество в этой работе принадлежит арифмометру, благодаря возможности производить на нем последовательно ряд арифметических действий, не останавливаясь на результате каждого из них в отдельности".* Из приведенного отрывка следует, что решение даже сравнительно простых задач требовало длительных и трудоемких вычислений, причем результат мог зависеть от того, какие численные таблицы использовались для работы. Лишь появление компьютеров позволило ускорить процесс вычислений, что, в свою очередь, способствовало разработке более сложных алгоритмов.

1. Формулы

Введем оператор математического ожидания $E[.]$, основываясь на стандартном определении [Гнеденко, 1961]. Рассмотрим непрерывно распределенную величину X . Математическое ожидание $E[X]$ определяется равенством

$$E[X] = \int_{-\infty}^{+\infty} xp(x)dx ,$$

где $p(x)$ – функция распределения X . $p(x)\Delta x$ – элементарная вероятность того, что $x < X < x+\Delta x$.

Если X – дискретная величина, имеющая N возможных значений, тогда

$$E[X] = \sum_{i=1}^N x_i p(x_i) .$$

Рассмотрим центрированную случайную величину w , распределенную по нормальному закону. Ее математическое ожидание будет равно нулю

$$E[w] = 0 . \quad (1.1)$$

Математическое ожидание постоянной величины x равно x

$$E[x] = x . \quad (1.2)$$

Введем оператор дисперсии $D[.]$

Дисперсия случайной величины w не равна нулю

$$D[w] = E[(w - E[w])^T (w - E[w])] = E[w^T w] = \sigma_w^2 \neq 0 . \quad (1.3)$$

Параметр σ_w называется среднеквадратическим отклонением случайной величины w .

Дисперсия постоянной величины x равна нулю.

$$D[x] = E[(x - E[x])^T (x - E[x])] = E[(x - x)^T (x - x)] = E[0] = 0 . \quad (1.4)$$

Рассмотрим некоторые формулы матричной алгебры, которые будут нужны в дальнейшем. Возьмем матрицу A размером $N \times n$, где $N > n$. N – число строк, n – число столбцов. Рассмотрим выражения

$$A(A^T A)^{-1} A^T = R \quad (1.5)$$

и

$$I - A(A^T A)^{-1} A^T = I - R, \quad (1.6)$$

где I – единичная матрица.

Матрицы R и $I - R$ размером $N \times N$ называются проекционными. Их свойства:

$$\text{rank } R = n \quad (1.7)$$

$$\text{rank } (I - R) = N - n. \quad (1.8)$$

$$R^2 = R \cdot R = R. \quad (1.9)$$

$$R \cdot (I - R) = (I - R) \cdot R = 0. \quad (1.10)$$

$$(I - R)^2 = I - R. \quad (1.11)$$

$$R \cdot R^T = R \cdot R = R. \quad (1.12)$$

$$(I - R) \cdot (I - R)^T = (I - R) \cdot (I - R) = (I - R)^T \cdot (I - R)^T = I - R. \quad (1.13)$$

Кроме того, выпишем некоторые формулы матричной алгебры

$$(A^T B A + C)^{-1} = C^{-1} - C^{-1} A^T (A C^{-1} A^T + B^{-1})^{-1} A C^{-1}. \quad (1.14)$$

$$(A^T B A + C)^{-1} A^T B = C^{-1} A^T (A C^{-1} A^T + B^{-1})^{-1}. \quad (1.15)$$

$$B A^T (A^T B A + C)^{-1} = (A C^{-1} A^T + B^{-1})^{-1} A C^{-1}. \quad (1.16)$$

$$(A + B)^{-1} = B^{-1} (A B^{-1} + I)^{-1}. \quad (1.17)$$

$$(A + B)^{-1} = A^{-1} - A^{-1}(A^{-1} + B^{-1})^{-1}A^{-1}. \quad (1.18)$$

$$(A^{-1} + B^{-1})^{-1} = A - A(A + B)^{-1}A. \quad (1.19)$$

2. Метод наименьших квадратов

Общепринятая параметрическая модель данных наблюдений имеет вид

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + w_1 &= h_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + w_2 &= h_2 \\ \dots & \\ a_{N1}x_1 + a_{N2}x_2 + \dots + a_{Nn}x_n + w_N &= h_N \end{aligned} \quad (2.1)$$

или, в матричном виде

$$Ax + w = h, \quad (2.2)$$

где A – матрица частных производных, имеющая размер $N \times n$ (N – число наблюдений, n – число неизвестных); x – вектор-столбец параметров, подлежащих оцениванию, размером $n \times 1$; w – вектор-столбец ошибок наблюдений размером $N \times 1$; вектор h – вектор-столбец разностей О–С размером $N \times 1$. Тогда, в соответствии с принципом наименьших квадратов, необходимо минимизировать функционал вида

$$S = \sum_{i=1}^N w_i^2, \quad (2.3)$$

т.е. найти такое решение, при котором сумма квадратов ошибок наблюдений окажется минимальной. В качестве альтернативы выражение (2.3) может быть также записано в матричном виде

$$S = w^T w. \quad (2.4)$$

Вектор случайных ошибок имеет математическое ожидание, равное нулю

$$E[w] = 0 \quad (2.5)$$

и дисперсию

$$D[w] = \sigma_w^2. \quad (2.6)$$

Очевидно, что матричная форма записи проще, хотя при этом название метода в какой-то мере теряет свой наглядный смысл. Тем не менее, для удобства далее будем пользоваться матричной формой записи всех выражений, если это не оговорено особо. Следует также отметить, что в соответствии с предпосылками МНК вектор x является детерминированной величиной, то есть,

$$E[x] = x, \quad (2.7)$$

и

$$D[x] = 0. \quad (2.8)$$

Вычислим частные производные $\frac{\partial S}{\partial x}$:

$$\begin{aligned} \frac{\partial S}{\partial x} &= \frac{\partial (w^T w)}{\partial x} = \frac{\partial ((h - Ax)^T (h - Ax))}{\partial x} = \\ &= \frac{\partial}{\partial x} (h^T h - x^T A^T h - h^T A x + x^T A^T A x) = 2A^T A x - 2A^T h \end{aligned} \quad (2.9)$$

Приравнявая к нулю частные производные $\frac{\partial S}{\partial x}$ (по необходимому условию существования экстремума), находим оценку вектора x

$$\hat{x} = (A^T A)^{-1} A^T h. \quad (2.10)$$

Вектор x , входящий в (2.2), можно оценить различными способами. Его оценка (2.10), записанная как $\hat{x}$ – всего лишь одна из множества возможных оценок, полученная методом наименьших квадратов. Следует понимать, что оценка вектора $\hat{x}$ (2.10) и сам вектор x в (2.2) – совершенно разные величины. Вектор x – это то, что есть на самом деле и что мы хотим оценить. Оценка $\hat{x}$ – "оценка" в прямом смысле этого слова, некоторая величина, близкая к вектору x , точное значение которого получить невозможно. Таким образом, как бы мы ни старались, но следующее выражение

$$\hat{x} \neq x \quad (2.11)$$

будет справедливо не только для оценки (2.10), но и для оценки, полученным любым другим способом. Подставив (2.2) в (2.10), получим

$$\hat{x} = (A^T A)^{-1} A^T h = (A^T A)^{-1} A^T (Ax + w) = x + (A^T A)^{-1} A^T w. \quad (2.12)$$

Второе слагаемое в (2.12) равно нулю только в том случае, если вектор случайных ошибок равен нулю, и точное значение вектора x можно найти лишь при полном отсутствии случайных ошибок. На практике случайные ошибки есть всегда (т.е. $w \neq 0$), поэтому (2.11) можно считать очевидным.

Тем не менее, оценка МНК (2.10) обладает тремя важными свойствами.

Свойство 1. Несмещенность

$$E[\hat{x}] = x. \quad (2.13)$$

Это свойство утверждает, что с исходным вектором x совпадает не оценка (2.6), а ее математическое ожидание. Действительно, используя (2.5), (2.7) и (2.12), имеем

$$\begin{aligned} E[\hat{x}] &= E[(A^T A)^{-1} A^T h] = E[(A^T A)^{-1} A^T (Ax + w)] = \\ &= E[x] + E[(A^T A)^{-1} A^T w] = x + (A^T A)^{-1} A^T E[w] = x. \end{aligned} \quad (2.14)$$

Свойство 2. Эффективность

Свойство эффективности заключается в том, что МНК позволяет получить оценку с минимальной дисперсией по сравнению с другими оценками. Чтобы провести доказательство, сперва получим формулу для дисперсионной матрицы оценки вектора (2.10)

$$\begin{aligned} \hat{Q}_x &= E[\hat{x}\hat{x}^T] = E[(A^T A)^{-1} A^T h h^T A (A^T A)^{-1}] = \\ &= (A^T A)^{-1} A^T E[hh^T] A (A^T A)^{-1} = \\ &= (A^T A)^{-1} A^T E[(Ax + w)(Ax + w)^T] A (A^T A)^{-1} = \\ &= (A^T A)^{-1} A^T E[ww^T] A (A^T A)^{-1}. \end{aligned} \quad (2.15)$$

В (2.4) мы неявно предполагали, что вектор ошибок имеет дисперсию равную единице, то есть, $E[ww^T] = Q_w = I$, где I – единичная матрица.

Тогда из (2.15) следует, что

$$\hat{Q}_x = (A^T A)^{-1} A^T E[w^T w] A (A^T A)^{-1}. \quad (2.16)$$

Диагональные элементы матрицы $\hat{Q}_x$ являются оценками дисперсии элементов вектора x , а недиагональные элементы матрицы $\hat{Q}_x$ характеризуют взаимосвязь между элементами вектора x – корреляции. Можно показать, что оценки дисперсии, полученные данным методом, окажутся минимальными, то есть диагональные элементы $\hat{Q}_x$, полученные по МНК, меньше диагональных элементов, полученных другим методом.

Свойство 3. Состоятельность.

$$\lim_{N \rightarrow \infty} \{\hat{x}_N\} = x. \quad (2.17)$$

Свойство состоятельности означает, что при увеличении числа наблюдений МНК-оценка приближается к исходному вектору x .

Кроме указанных трех основных свойств МНК-оценки обладают еще и другими свойствами, которые можно найти в монографиях, посвященных этой проблеме [Рао, 1968; Себер, 1980].

Эти свойства оценки МНК, а также простота приведенных формул, привели к тому, что этот алгоритм широко применяется в различных областях знаний для решения самых разнообразных задач.

Введем вектор остаточных невязок

$$\varepsilon = h - A\hat{x} = h - A(A^T A)^{-1} A^T h = (I - R)h, \quad (2.18)$$

где R – проекционная матрица вида (2.6). Остаточные невязки характеризуют разброс наблюдений относительно выбранной модели.

Рассмотрим дисперсию вектора остаточных невязок

$$\begin{aligned} D[\varepsilon] &= E[\varepsilon^T \varepsilon] = E[(h - A\hat{x})^T (h - A\hat{x})] = E[((I - R)h)^T ((I - R)h)] = \\ &= E[h^T (I - R)^T (I - R)h] = E[h^T (I - R)h] = E[w^T (I - R)w]. \end{aligned} \quad (2.19)$$

Квадратичная форма $w^T (I - R)w$ имеет χ^2 -распределение с числом степеней свободы равным $\text{rank}(I - R) = N - n$. Используя свойства χ^2 -распределения [Рао, 1968], получаем

$$D[\varepsilon] = D[w](N - n), \quad (2.20)$$

Формула (2.20) связывает дисперсию вектора случайных ошибок и дисперсию вектора невязок. Но обычно (2.20) нужна, чтобы ввести безразмерную величину, которую называют "нормированный хи-квадрат" [Рао, 1968]

$$\hat{\chi}^2 = \frac{\varepsilon^T \varepsilon}{N - n}. \quad (2.21)$$

Она часто используется в различных статистических критериях [Себер, 1980; Валеев, 1991].

Среднеквадратическая ошибка вектора x вычисляется по формуле

$$\hat{\sigma}_x = \hat{\chi} \cdot [\text{diag}(\hat{Q}_x)]^{1/2}, \quad (2.22)$$

Результаты оценивания, как правило, записываются в виде $\hat{x} \pm \hat{\sigma}_x$. Это означает, что с вероятностью 0.9973 истинное значение вектора x лежит в интервале

$$\hat{x} - 3\hat{\sigma}_x \leq x \leq \hat{x} + 3\hat{\sigma}_x.$$

Учитывая вероятностный характер оценки, существует ненулевая вероятность того, что оценка вектора будет отличаться от его истинного значения больше, чем на 3σ .

Может показаться, что параметр $\hat{\chi}^2$ в (2.21) имеет размерность и совпадает с $D[w]$ в (2.20). Это вызвано тем, что в классическом варианте МНК априорные дисперсии каждого отдельного наблюдения принимаются равными единице и не входят явно в выражение (2.21). По той же причине верно равенство $D[w] = 1$. На самом деле $\hat{\chi}^2$ – величина безразмерная. В следующем разделе будет показано, как он вычисляется для более общего случая.

Выше мы предполагали, что параметрическая модель (2.2) является полной, то есть все физические эффекты, образующие вектор наблюдений, известны и включены в эту модель. Теперь рассмотрим случай, когда некоторые параметры не вошли в модель из-за отсутствия информации. В этом случае оценка вектора x (2.10) может оказаться смещенной.

Пусть параметрическая модель имеет вид

$$Ax + Cz + w = h, \quad (2.23)$$

где слагаемое Cz представляет компоненту данных, о которой ничего неизвестно. В этом случае при оценивании используется выражение (2.2), и математическое ожидание оценки вектора x

$$\begin{aligned} E[\hat{x}] &= E[(A^T Q_w^{-1} A)^{-1} A^T Q_w^{-1} h] = \\ &= (A^T Q_w^{-1} A)^{-1} A^T Q_w^{-1} E[Ax + Cz + w] = \\ &= (A^T Q_w^{-1} A)^{-1} A^T Q_w^{-1} A E[x] + (A^T Q_w^{-1} A)^{-1} A^T Q_w^{-1} E[Cz + w] = \\ &= x + (A^T Q_w^{-1} A)^{-1} A^T Q_w^{-1} C E[z]. \end{aligned} \quad (2.24)$$

Поскольку z – детерминированный вектор, то $E[z] = z$. В результате имеем

$$E[\hat{x}] = x + (A^T Q_w^{-1} A)^{-1} A^T Q_w^{-1} Cz, \quad (2.25)$$

где второе слагаемое, в общем случае не равно нулю

$$(A^T Q_w^{-1} A)^{-1} A^T Q_w^{-1} Cz = Rz = \Delta x \neq 0, \quad (2.26)$$

и есть смещение оценки вектора x , вызванное неполнотой принятой параметрической модели. Из (2.25)

$$E[\hat{x}] = x + Rz = x + \Delta x \neq x, \quad (2.27)$$

то есть статистическая оценка вектора x не совпадает с вектором x . Поэтому при неполной параметрической модели велик риск получить неправильную оценку оцениваемой величины.

3. Взвешенный метод наименьших квадратов

Предполагая неравноточность наблюдений, вводят матрицу априорных дисперсий, исходя из априорных предположений о точности каждого отдельного наблюдения.

Рассмотрим параметрическую модель вида (2.2). Будем минимизировать функционал

$$S = w^T Q_w^{-1} w, \quad (3.1)$$

где матрица ковариаций определяется выражением

$$Q_w = E[ww^T]. \quad (3.2)$$

Здесь $E[.]$ – оператор математического ожидания. При условии, что ошибки наблюдений некоррелированы, матрица (3.2) оказывается диагональной.

Приравнявая к нулю частные производные вида $\frac{\partial S}{\partial x}$ (по необходимому условию существования экстремума), находим оценку вектора x

$$\hat{x} = (A^T Q_w^{-1} A)^{-1} A^T Q_w^{-1} h, \quad (3.3)$$

оценку матрицы ковариаций

$$\hat{Q}_x = D[\hat{x}] = E[\hat{x}\hat{x}^T] = (A^T Q_w^{-1} A)^{-1} \quad (3.4)$$

и вектор остаточных невязок

$$\varepsilon = h - A\hat{x} = (I - R)h. \quad (3.5)$$

Здесь $R = A(A^T Q_w^{-1} A)^{-1} A^T Q_w^{-1}$ - также проекционная матрица типа (1.6).

Среднеквадратическая ошибка вычисляется по формуле

$$\hat{\sigma}_x = \hat{\chi} \cdot [\text{diag}(\hat{Q}_x)]^{1/2}, \quad (3.6)$$

а "нормированный хи-квадрат"

$$\hat{\chi}^2 = \frac{\varepsilon^T Q_w^{-1} \varepsilon}{N - n}. \quad (3.7)$$

Теперь $\hat{\chi}^2$ явно выглядит величиной безразмерной. Для случая обычного МНК вместо матрицы Q_w^{-1} используется единичная матрица, и (3.7) сводится к (2.21).

Формулы (3.1) – (3.7) представляют взвешенный МНК ("weighted least squares method"). Часто в литературе используются формулы, в которых вместо обратной матрицы априорных ковариаций Q_w используется так называемая весовая матрица. На ее главной диагонали вместо априорных дисперсий стоят веса наблюдений, определяемые формулой

$$p_i = \frac{\sigma_w^2}{\sigma_{w(i)}^2}, \quad (3.8)$$

где i – порядковый номер элемента матрицы. В числителе стоит σ_w^2 – "усредненная" априорная дисперсия вектора наблюдений, в знаменателе – индивидуальные дисперсии каждого отдельного наблюдения.

Замена ковариационной матрицы на весовую иногда бывает удобной. В этом случае формулы (3.1), (3.3) – (3.7) сохраняют смысл при простой замене одной матрицы на другую за двумя исключениями: матрица ковариаций (3.4) теряет размерность, а нормированный хи-квадрат заменяется на его размерный аналог – "ошибка единицы веса".

Однако, назначение весов в любом случае не может быть сделано совершенно произвольно. Исходя из (3.8) и приближенного соотношения

$$\sum_{i=1}^N \sigma_{w(i)}^2 \approx \sigma_w^2 (N - n), \quad (3.9)$$

получаем выражение

$$\sum_{i=1}^N \frac{1}{p_i} = N - n, \quad (3.10)$$

которое выполняет роль условия нормировки при назначении индивидуальных весов.

Взвешенный МНК применяется, наверное, чаще других методов, ввиду простоты и малых вычислительных затрат. При его применении максимальный размер обрабатываемой матрицы равен $n \times n$, в то время как для обобщенного МНК, описанного в следующем разделе, нужно также обращать матрицу размера $N \times N$. Обращение матрицы больших размеров требует увеличения мощности процессора и оперативной памяти. При недостаточно мощном процессоре время вычислений становится слишком большим.

Рассмотрим еще величину, которая носит название взвешенное средне-квадратическое отклонение (СКО), по-английски, "weighted root-mean squares" (wrms). Она вычисляется по формуле

$$\text{wrms} = \left(\frac{\boldsymbol{\varepsilon}^T \mathbf{Q}_w^{-1} \boldsymbol{\varepsilon}}{\mathbf{e}^T \mathbf{Q}_w^{-1} \mathbf{e}} \right)^{1/2}, \quad (3.11)$$

где $\mathbf{e}$ – вектор, составленный из единиц. С учетом диагональности матрицы априорных ковариаций в скалярном виде (3.11) приобретает вид

$$wrms = \left(\frac{\sum_{i=1}^N \frac{\varepsilon_i^2}{\sigma_{W(i)}^2}}{\sum_{i=1}^N \frac{1}{\sigma_{W(i)}^2}} \right)^{1/2}, \quad (3.12)$$

Аналитически формулы (3.7) и (3.11) похожи, но они представляют разные характеристики полученного решения параметрической модели. Безразмерный нормированный хи-квадрат отражает (скорее, качественно) полноту выбранной модели и используется для вычисления среднеквадратических ошибок. Взвешенное СКО имеет размерность вектора наблюдений и количественно характеризует средний уровень случайных ошибок ("шума" наблюдений). Оно используется достаточно часто для того, чтобы соотнести внутреннюю и внешнюю ошибку выполненных наблюдений. Например, временные ряды оценок длин баз, полученные в результате обработки последовательных 24-часовых РСДБ серий, также выглядят зашумленными. Этот разброс вычисляется по формуле (3.12) и характеризует внешнюю ошибку. Но каждая индивидуальная точка из этого временного ряда имеет свою характеристику точности, полученную в результате оценивания индивидуальной 24-часовой серии. При обработке РСДБ наблюдений, как правило, оказывается, что внешняя оценка точности в полтора-два раза хуже внутренней, то есть реальная наблюдательная ошибка больше той, которая получается в результате оценивания. Такое может произойти, если ошибки наблюдений являются коррелированными. Взвешенный МНК как раз пренебрегает таким обстоятельством. В таких случаях, чтобы избежать несоответствия между внешней и внутренней точностью оценок, следует пользоваться обобщенным методом наименьших квадратов.

Если ошибки наблюдений по какой-то причине коррелированы, то матрица априорных ковариаций Q_w (3.2) оказывается полной. Эта разновидность МНК называется обобщенным МНК – "generalized least squares method". В этом случае все формулы (3.1) – (3.7) и (3.11) сохраняют вид.

При полной матрице ковариаций переход к матрице весов становится некорректным, поскольку обратная матрица будет недиагональной, и традиционное понятие веса теряет свой смысл. Но в связи с тем, что вычислить

априорные корреляции между наблюдениями, как правило, затруднительно, обобщенный МНК в чистом виде практически не применяется, хотя такие попытки недавно были сделаны [Schuh, Tesmer; 1999; Schuh, Tesmer, 2000].

4. Метод максимального правдоподобия

В этом разделе, упоминая метод наименьших квадратов, будем иметь ввиду взвешенный или обобщенный варианты МНК, ссылаясь на формулы раздела 3.

При изложении МНК авторы довольно часто излагают вывод рабочих формул, начиная прямо с минимизации функционала (3.1), не уточняя, почему было выбрано именно это выражение. В действительности же функционал (3.1) определяется выражением для условной плотности вероятности вектора наблюдений h относительно вектора неизвестных x , записываемой в виде $p(h|x)$. Обычно предполагается, что ошибки наблюдений являются случайными, распределенными по нормальному закону. В общем случае при ковариационной матрице ошибок наблюдений Q_w условная плотность вероятности имеет вид

$$p(h|x) = \frac{1}{(2\pi)^{N/2} [\det Q_w]^{1/2}} \exp \left[-\frac{1}{2} (h - Ax)^T Q_w^{-1} (h - Ax) \right]. \quad (4.1)$$

Для вычисления необходимого функционала можно применить метод максимального правдоподобия (ММП) [Сейдж, Мелса; 1976]. Этот метод основан на максимизации условной плотности вероятности наблюдений $p(h|x)$ относительно параметра x . По необходимому условию существования экстремума оценка определяется из уравнения

$$\frac{\partial p(h|x)}{\partial x} \Big|_{x=\hat{x}(h)} = 0, \quad (4.2)$$

при этом параметру x присваивается значение, при котором наиболее вероятно появление наблюдавшейся реализации вектора h . Никакой априорной информации о самом параметре x , подлежащем оцениванию, не требуется.

Соответствующая производная от плотности вероятности (4.1) имеет вид

$$\frac{\partial p(h|x)}{\partial x} \bigg|_{x=\hat{x}(h)} = \frac{1}{(2\pi)^{N/2} [\det Q_w]^{1/2}} \frac{\partial}{\partial x} \left[-\frac{1}{2} (h - Ax)^T Q_w^{-1} (h - Ax) \right] \times \\ \times \exp \left[-\frac{1}{2} (h - Ax)^T Q_w^{-1} (h - Ax) \right]. \quad (4.3)$$

Приравнявая (4.3) к нулю, получаем, что условие (4.2) выполняется, если

$$\frac{\partial}{\partial x} (h - Ax)^T Q_w^{-1} (h - Ax) = 0, \quad (4.4)$$

то есть условие для вычисления оценки ММП аналогично условию минимизации функционала (4.1) и выражение для оценки ММП

$$\hat{x}_{\text{ММП}} = (A^T Q_w^{-1} A)^{-1} A^T Q_w^{-1} h \quad (4.5)$$

совпадает с оценкой МНК (3.3).

Естественно, что при другом выражении для $p(h|x)$ оценка вектора неизвестных изменится. Это означает, что ММП является гораздо более гибким методом оценивания, чем МНК, поскольку позволяет учитывать реальную функцию распределения ошибок наблюдений.

Априорное предположение о нормальном законе распределения ошибок, конечно, выглядит не вполне обосновано. Однако поскольку никогда нельзя знать истинное распределение ошибок наблюдений, приходится делать некоторые предположения о возможном характере их распределения. При астрометрических наблюдениях имеются несколько источников ошибок (технического или атмосферного происхождения). Поэтому на результат измерений влияют различные случайные процессы с приблизительно одинаковыми дисперсиями. Это позволяет сделать предположение о том, что суммарные случайные ошибки (вектор w) будут распределены по нормальному закону. Применение метода максимального правдоподобия к отысканию неизвестных параметров в этих условиях, в свою очередь, приводит к определенной вычислительной схеме – МНК. Если реальные ошибки наблюдений имеют распределение, немного отличающееся от нормального закона, то и в этом случае МНК позволяет получить достаточно надежные оценки. Они обладают меньшей точностью, чем истинно правдоподобные оценки, зато они получены ценой не слишком больших усилий. [Мудров, Кушко; 1983].

Конечно, если распределение ошибок существенно отличается от нормального закона, то следует применять не МНК, а другие методы оценивания.

Теперь рассмотрим случай, когда вектор параметров является не детерминированной величиной, определяемой формулами (2.7), (2.8), а случайной величиной с математическим ожиданием

$$E[x] = 0 \quad (4.6)$$

и ковариационной матрицей

$$Q_x = E[xx^T]. \quad (4.7)$$

Что касается выражения для условной плотности вектора наблюдений h относительно вектора неизвестных x , то оно выглядит сложнее, чем (4.3). Это вызвано тем, что вектор x , как и вектор h , является случайным, поэтому при выводе необходимо использовать формулу Байеса [Гнеденко, 1961]

$$p(h|x) = \frac{p(h)p(x|h)}{p(x)}. \quad (4.8)$$

Если случайные вектора x и h распределены по нормальному закону, то (4.8) имеет следующий вид [Сейдж, Мелса, 1974]

$$p(h|x) = \frac{[\det(AQ_x A^T + Q_w)]^{1/2}}{(2\pi)^{N/2} [\det Q_w]^{1/2} [\det Q_x]^{1/2}} \times \\ \times \exp \left[-\frac{1}{2} (x - m(x))^T (A^T Q_w^{-1} A + Q_x^{-1}) (x - m(x)) \right], \quad (4.9)$$

где

$$m(x) = (A^T Q_w^{-1} A + Q_x^{-1})^{-1} A^T Q_w^{-1} h. \quad (4.10)$$

Можно показать, что для выполнения условия (4.2) из формул (4.9) и (4.10) следует равенство

$$(A^T Q_w^{-1} A + Q_x^{-1})x = A^T Q_w^{-1} h, \quad (4.11)$$

откуда получаем выражение для оценки вектора x

$$\hat{x}_{\text{ММП}} = (A^T Q_w^{-1} A + Q_x^{-1})^{-1} A^T Q_w^{-1} h. \quad (4.12)$$

Найдем дисперсию оценки ММП

$$\begin{aligned}\hat{Q}_x &= D[\hat{x}] = E[\hat{x}\hat{x}^T] = \\ &= (A^T Q_w^{-1} A + Q_x^{-1})^{-1} A^T Q_w^{-1} E[hh^T] Q_w^{-1} A (A^T Q_w^{-1} A + Q_x^{-1})^{-1}.\end{aligned}\quad (4.13)$$

Поскольку x – случайный вектор, то

$$E[hh^T] = A Q_x A^T + Q_w. \quad (4.14)$$

Из (4.13) и (4.14) с использованием (1.16)

$$\begin{aligned}\hat{Q}_x &= D[\hat{x}] = E[\hat{x}\hat{x}^T] = \\ &= (A^T Q_w^{-1} A + Q_x^{-1})^{-1} A^T Q_w^{-1} (A Q_x A^T + Q_w) (A Q_x A^T + Q_w)^{-1} A Q_x = \\ &= Q_x A^T (A Q_x A^T + Q_w)^{-1} A Q_x = Q_x - (A^T Q_w^{-1} A + Q_x^{-1})^{-1}\end{aligned}\quad (4.15)$$

Поскольку диагональные элементы матричного выражения $(A^T Q_w^{-1} A + Q_x^{-1})^{-1}$ из (4.15) неотрицательны, то всегда имеет место неравенство

$$\text{diag}(\hat{Q}_x) \leq \text{diag}(Q_x). \quad (4.16)$$

Невязки, СКО и $\hat{\chi}^2$ вычисляются по формулам (3.5) – (3.7).

В формуле для оценки ММП (4.12) появилась матрица ковариаций Q_x из (4.7). Свойство несмещенности этой оценки выполняется только при соблюдении условия (4.6) (математическое ожидание должно быть равным нулю). В противном случае, если $E[x] \neq 0$, то оценка ММП оказывается более сложной, а именно

$$\hat{x}_{\text{ММП}} = (A^T Q_w^{-1} A + Q_x^{-1})^{-1} (A^T Q_w^{-1} h + Q_x^{-1} E[x]). \quad (4.17)$$

Несмещенность оценки (4.17) легко проверяется.

Таким образом, метод максимального правдоподобия "работает" в более широком диапазоне возможных вариантов, чем МНК, например, при произвольной функции распределения случайных ошибок или при недетерминированном векторе определяемых параметров.

5. Метод среднеквадратической коллокации и его связь с другими методами оценивания.

В последнее время в ряде задач все чаще появляется необходимость учитывать корреляционные связи между наблюдениями. По-видимому, впервые такая проблема встала перед обработчиками геодезической информации. Например, Мориц, один из крупнейших современных геодезистов вводит так называемый “вектор сигнала” для описания стохастических параметров, которые не вошли в параметрическую модель МНК [Moritz, 1973; Мориц, 1980]. Этот сигнал описывается как случайная величина, то есть в терминах математического ожидания и ковариационной матрицы. Такой подход позволяет существенно расширить возможности исследователя при изучении различных природных явлений.

Рассмотрим параметрическую модель данных вида

$$Ax + By + w = h, \quad (5.1)$$

в которой по сравнению с моделью (3.1) добавлен член By . Он представляет те параметры, которые являются стохастическими и могут быть описаны в статистических терминах, то есть с помощью математического ожидания, дисперсии и ковариационной матрицы. Обычно предполагается, что математическое ожидание стохастических параметров входит в ту часть (5.1), в которой представлены детерминированные параметры, то есть в Ax . Иногда в детерминированную часть необходимо включать и линейный тренд, что при обработке РСДБ данных проделывается при исключении влияния рассинхронизации водородных стандартов времени. Подобный подход обеспечивает выполнение условия

$$E[y] = 0. \quad (5.2)$$

Поскольку таких параметров может быть несколько, то и матрица B , и вектор y в общем случае являются блочными структурами. Каждый элемент блочной матрицы B есть матрица размером $N \times N$, а каждый элемент блочного вектора y – вектор размером $N \times 1$. Раскладывая такой блок в сумму, получим выражение, в котором будет столько одноблочных слагаемых, сколько имеется стохастических параметров. Если число таких параметров m , то выражение (5.1) можно переписать в виде

$$\begin{aligned}
Ax + By + w &= Ax + \sum_{i=1}^m B_i y_i + w = \\
&= Ax + (B_1, B_2, \dots, B_m) \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_m \end{pmatrix} + w = h.
\end{aligned} \tag{5.3}$$

При использовании параметрической модели вида (5.1) нужно вычислять не только оценки постоянных параметров x , но и стохастических параметров y . Условная плотность вероятности вектора наблюдений h относительно вектора x будет напоминать (4.9)

$$p(h|x) = \frac{1}{(2\pi)^{N/2} [\det Q_w]^{1/2}} \times \exp \left[-\frac{1}{2} (h - Ax - By)^T Q_w^{-1} (h - Ax - By) \right]. \tag{5.4}$$

Что касается выражения для условной плотности вектора наблюдений h относительно вектора неизвестных y , то оно выглядит сложнее. Это вызвано тем, что вектор y , как и вектор h , является случайным, поэтому при выводе необходимо использовать формулу Байеса [Гнеденко, 1961]

$$p(h|y) = \frac{p(h)p(y|h)}{p(y)}. \tag{5.5}$$

Если случайные вектора y и h распределены по нормальному закону, то (5.5) с учетом условия (5.2) имеет следующий вид [Сейдж, Мелса, 1974]

$$\begin{aligned}
p(h|y) &= \frac{[\det(BQ_y B^T + Q_w)]^{1/2}}{(2\pi)^{N/2} [\det Q_w]^{1/2} [\det Q_y]^{1/2}} \times \\
&\times \exp \left[-\frac{1}{2} (y - m(y))^T (B^T Q_w^{-1} B + Q_y^{-1}) (y - m(y)) \right],
\end{aligned} \tag{5.6}$$

где

$$m(y) = (B^T Q_w^{-1} B + Q_y^{-1})^{-1} B^T Q_w^{-1} (h - Ax). \tag{5.7}$$

Для нахождения оценок параметров необходимо вычислить следующие частные производные: из (5.4)

$$\frac{\partial}{\partial x} (h - Ax - By)^T Q_w^{-1} (h - Ax - By) = 0 \quad (5.8)$$

и из (5.6)

$$\frac{\partial}{\partial y} [(y - m(y))^T (B^T Q_w^{-1} B + Q_y^{-1}) (y - m(y))] = 0. \quad (5.9)$$

После дифференцирования и достаточно громоздких преобразований получим выражения для оценок постоянных

$$\hat{x} = (A^T Q_w^{-1} A)^{-1} A^T Q_w^{-1} (h - By) \quad (5.10)$$

и стохастических параметров

$$\hat{y} = Q_y B^T (B Q_y B^T + Q_w)^{-1} (h - Ax). \quad (5.11)$$

Теперь покажем, как эти же оценки могут быть получены с применением подхода, использованного в разделе 2. Если имеется параметрическая модель (5.1), то для вычисления оценок неизвестных, нужно минимизировать функционал вида

$$S_1 = w^T Q_w^{-1} w + y^T Q_y^{-1} y, \quad (5.12)$$

Приравнивая к нулю производные $\frac{\partial S_1}{\partial x}$ и $\frac{\partial S_1}{\partial y}$, получим выражения

$$A^T Q_w^{-1} Ax = A^T Q_w^{-1} (h - By) \quad (5.13a)$$

и

$$(B^T Q_w^{-1} B + Q_y^{-1}) y = B^T Q_w^{-1} (h - Ax), \quad (5.13b)$$

из которых следуют формулы (5.10) и (5.11).

Оценка вектора x (5.10) зависит явно от вектора стохастических параметров y . Чтобы найти выражение для оценки x , не зависящее явно от y , подставим оценку (5.11) в правую часть выражения (5.10). После некоторых преобразований получим формулу

$$\hat{x} = [A^T (Q_w + BQ_y B^T)^{-1} A]^T A^T (Q_w + BQ_y B^T)^{-1} h. \quad (5.14)$$

Теперь имеется два эквивалентных выражения для оценки параметров вектора x : (5.10) и (5.14), при этом первая из них зависит явно от y , а вторая определяется только ковариационной матрицей Q_y . Разница заключается в том, что для использования (5.10) необходимо знать сам вектор y , а для использования (5.14) необходимо иметь только априорную информацию о векторе y , которая содержится в матрице Q_y . В этом заключается преимущество формулы (5.14) по отношению к (5.10). Появляется возможность очередного оценивания параметров: на первом шаге вычисляется оценка вектора x (5.14), на втором – оценка вектора y по формуле .

$$\hat{y} = Q_y B^T (BQ_y B^T + Q_w)^{-1} (h - A\hat{x}). \quad (5.15)$$

Формулы (5.14) и (5.15) являются основными рабочими формулами МСКК для оценивания параметров модели (5.1).

Теперь вернемся к функционалу (5.12). Используя представление (5.3), перепишем его в развернутом виде

$$\begin{aligned} S_1 &= w^T Q_w^{-1} w + \sum_{i=1}^m y_i^T Q_{y(i)}^{-1} y_i = \\ &= w^T Q_w^{-1} w + (y_1, y_2, \dots, y_m) \begin{pmatrix} Q_{y(1)}^{-1} & 0 & 0 & 0 \\ 0 & Q_{y(2)}^{-1} & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & Q_{y(m)}^{-1} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_m \end{pmatrix} \end{aligned} \quad (5.16)$$

Формулу (5.16) можно обобщить, объединяя вектор ошибок w с блочным вектором y . Тогда функционал можно записать так

$$S_1 = (w, y_1, \dots, y_m) \begin{pmatrix} Q_w^{-1} & 0 & 0 & 0 \\ 0 & Q_{y(1)}^{-1} & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & Q_{y(m)}^{-1} \end{pmatrix} \begin{pmatrix} w \\ y_1 \\ \dots \\ y_m \end{pmatrix} = r Q_r^{-1} r, \quad (5.17)$$

где матрица Q_w определяется формулой (3.2), матрицы ковариаций стохастических параметров – выражением

$$Q_{y(i)} = E[y_i y_i^T]. \quad (5.18)$$

В (5.17) вектор r – обобщенный блочный вектор, элементами которого являются случайные ошибки и стохастические параметры

$$r = \begin{pmatrix} w \\ y_1 \\ \dots \\ y_m \end{pmatrix}; \quad (5.19)$$

Q_r – обобщенная блочная матрица ковариаций вида

$$Q_r = \begin{pmatrix} Q_w & 0 & 0 & 0 \\ 0 & Q_{y(1)} & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & Q_{y(m)} \end{pmatrix}. \quad (5.20)$$

Очевидно, что переход от (5.17) к (5.12) справедлив только в том случае, если в матрице (5.20) действительно нет корреляций между компонентами вектора y , а также между ними и вектором w . Только тогда все недиагональные блоки в (5.20) оказываются равными нулю, и (5.12) будет упрощенным вариантом выражения (5.17). Таким образом, можно отметить, что имеется более общее выражение для функционала, подлежащего минимизации, по сравнению с (5.12), а именно (5.17), при этом даже не нужно накладывать никаких условий на недиагональные блоки матрицы Q_r , помня, однако, что в этом случае переходить к (5.12) нельзя.

Теперь запишем (5.14) в виде

$$\hat{x} = (A^T Q^{-1} A)^{-1} A^T Q^{-1} h, \quad (5.21)$$

где матрица Q есть сумма матриц

$$Q = Q_w + B Q_y B^T. \quad (5.22)$$

Сравнение формул (3.3) и (5.21) показывает, что выражения для оценок МНК и МСКК совпадают по своей структуре. Различие заключается лишь в том, что на месте диагональной матрицы Q_w стоит матрица $Q_w + B Q_y B^T$, которая будет полной. Соответственно, изменится и матрица ковариаций (3.4)

$$\hat{Q}_x = D[\hat{x}] = E[\hat{x}\hat{x}^T] = (A^T Q^{-1} A)^{-1}, \quad (5.23)$$

где Q взято из (5.22).

Очевидно, что МСКК может быть интерпретирован как обобщенный МНК с полной матрицей априорных ковариаций ошибок наблюдений. При этом (5.22) дает аналитическое выражение для вычислений всех элементов этой матрицы.

Теперь вычислим невязки в соответствии с (3.5)

$$\varepsilon = h - A\hat{x} = h - A(A^T Q^{-1} A)^{-1} A^T Q^{-1} h = (I - R)h. \quad (5.24)$$

Несмотря на то, что в (3.5) и (5.24) использовались различные ковариационные матрицы, матрица R является, по-прежнему, проекционной, а значит, сохраняет свои свойства. Это означает, что для вычисления χ^2 можно пользоваться аналогом формулы (3.7). При этом, однако, следует отметить, что невязки (5.24) не будут некоррелированными, поскольку в них еще присутствуют стохастические параметры W . Тем не менее, эти невязки также можно использовать для вычисления среднеквадратических ошибок и построения статистических критериев. Подробнее этот вопрос рассматривается ниже.

Точность оценивания случайного вектора y (5.11), характеризуется отклонением полученной оценки от истинного значения $\Delta y = y - \hat{y}$. Тогда дисперсия будет равна

$$\begin{aligned} \hat{Q}_y &= E[\Delta y (\Delta y)^T] = E[(y - \hat{y})(y - \hat{y})^T] = \\ &= E[yy^T] - E[y\hat{y}^T] - E[\hat{y}y^T] + E[\hat{y}\hat{y}^T]. \end{aligned} \quad (5.25)$$

По определению ковариационной матрицы

$$E[yy^T] = Q_y. \quad (5.26)$$

Кроме того, используя формулу (5.11), будем иметь следующие выражения для слагаемых из (5.25)

$$\begin{aligned} E[\hat{y}y^T] &= E[(\hat{y}\hat{y}^T)^T] = E[Q_y B(BQ_y B^T + Q_w)^{-1} (h - A\hat{x})y^T] = \\ &= Q_y B(BQ_y B^T + Q_w)^{-1} (I - R)E[hy^T]. \end{aligned} \quad (5.27)$$

и

$$\begin{aligned}
E[\hat{y}\hat{y}^T] &= E[Q_y B(BQ_y B^T + Q_w)^{-1} (h - A\hat{x}) \times \\
&\times (h - Ax)^T (BQ_y B^T + Q_w)^{-1} B^T Q_y] = \\
&= Q_y B(BQ_y B^T + Q_w)^{-1} (I - R) E[hh^T] (I - R)^T (BQ_y B^T + Q_w)^{-1} B^T Q_y.
\end{aligned} \tag{5.28}$$

В (5.27) и (5.28) при вычислении математического ожидания предполагаем, что случайные вектора h и y некоррелированы. Тогда

$$\begin{aligned}
E[\hat{y}y^T] &= Q_y B(BQ_y B^T + Q_w)^{-1} (I - R) E[(By + w)y^T] = \\
&= Q_y B(BQ_y B^T + Q_w)^{-1} (I - R) BQ_y
\end{aligned} \tag{5.29}$$

и

$$\begin{aligned}
E[\hat{y}\hat{y}^T] &= Q_y B(BQ_y B^T + Q_w)^{-1} (I - R) E[(By + w)(By + w)^T] \times \\
&\times (I - R)^T (BQ_y B^T + Q_w)^{-1} B^T Q_y = Q_y B(BQ_y B^T + Q_w)^{-1} (I - R) \times \\
&\times (BQ_y B^T + Q_w) (I - R)^T (BQ_y B^T + Q_w)^{-1} B^T Q_y.
\end{aligned} \tag{5.30}$$

Чтобы упростить (5.30), рассмотрим отдельно следующее выражение:
 $(BQ_y B^T + Q_w)^{-1} (I - R) (BQ_y B^T + Q_w)$.

$$\begin{aligned}
&(BQ_y B^T + Q_w)^{-1} (I - R) (BQ_y B^T + Q_w) = \\
&= Q^{-1} (I - A(A^T Q^{-1} A)^{-1} A^T Q^{-1}) Q = \\
&= I - Q^{-1} A(A^T Q^{-1} A)^{-1} A^T = (I - R)^T.
\end{aligned} \tag{5.31}$$

Используя (5.31), упростим (5.30):

$$E[\hat{y}\hat{y}^T] = Q_y B(I - R)^T (I - R)^T (BQ_y B^T + Q_w)^{-1} B^T Q_y. \tag{5.32}$$

Пользуясь свойством проекционной матрицы (1.13) из (5.32) получим

$$E[\hat{y}\hat{y}^T] = Q_y B^T (I - R)^T (BQ_y B^T + Q_w)^{-1} BQ_y. \tag{5.33}$$

Теперь, собрав (5.26), (5.29) и (5.33), напишем выражение для дисперсии y

$$\begin{aligned}
\hat{Q}_y &= Q_y - Q_y B^T (I - R)^T (BQ_y B^T + Q_w)^{-1} BQ_y - \\
&- Q_y B^T (BQ_y B^T + Q_w)^{-1} (I - R) BQ_y + \\
&+ Q_y B(I - R)^T (BQ_y B^T + Q_w)^{-1} BQ_y.
\end{aligned} \tag{5.34}$$

Уместно заметить, что

$$(I - R)^T (BQ_y B^T + Q_w)^{-1} = (BQ_y B^T + Q_w)^{-1} (I - R), \quad (5.35)$$

поэтому окончательный вариант можно записать двумя способами. Чтобы избежать лишнего знака транспонирования, выберем выражение из правой части (5.35). Тогда

$$\hat{Q}_y = Q_y - Q_y B^T (BQ_y B^T + Q_w)^{-1} (I - R) BQ_y. \quad (5.36)$$

Теперь рассмотрим невязки. Если в (5.24) приведено выражение для невязок только после удаления оценок x , то теперь приведем выражение, удалив также и оценку y

$$\begin{aligned} \varepsilon_1 &= h - Ax - By = h - A\hat{x} - BQ_y B^T (BQ_y B^T + Q_w)^{-1} (h - A\hat{x}) = \\ &= (I - R)h - BQ_y B^T (BQ_y B^T + Q_w)^{-1} (I - R)h = \\ &= (I - RQ_y B^T (BQ_y B^T + Q_w)^{-1}) (I - R)h = \\ &= (I - BQ_y B^T (BQ_y B^T + Q_w)^{-1}) \varepsilon. \end{aligned} \quad (5.37)$$

Теперь нужно найти выражение, связывающее $D[w]$ и $D[\varepsilon]$, то есть нужно найти матрицу Z , такую что

$$D[w] = \hat{\sigma}_w^2 = \frac{E[\varepsilon_1^T Z^{-1} \varepsilon_1]}{N - n}. \quad (5.38)$$

Можно показать, что такая матрица существует и вычисляется по формуле

$$Z = Q_w (BQ_y B^T + Q_w)^{-1} Q_w \quad (5.39)$$

Кроме того, невязки (5.24) и (5.37) связаны друг с другом следующим выражением.

$$\varepsilon^T Q^{-1} \varepsilon = \varepsilon_1^T Z^{-1} \varepsilon_1 \quad (5.40)$$

или

$$\varepsilon^T (BQ_y B^T + Q_w)^{-1} \varepsilon = \varepsilon_1^T Q_w^{-1} (BQ_y B^T + Q_w) Q_w^{-1} \varepsilon_1. \quad (5.41)$$

Тождество (5.41) можно проверить прямой подстановкой соответствующих выражений для невязок. Важность его состоит в том, что, как оказывается, для того, чтобы вычислить χ^2 , совсем не обязательно вводить в алгоритм

матрицу Z^{-1} и проводить оценивание вектора y (он может и не представлять интереса). Использование вектора ε , который вычисляется по (5.24) сразу после оценивания вектора x , позволяет получить χ^2 без дополнительных вычислительных затрат по формуле

$$\hat{\chi}^2 = \frac{\varepsilon^T Q^{-1} \varepsilon}{N - n} = \frac{\varepsilon^T (BQ_y B^T + Q_w)^{-1} \varepsilon}{N - n}. \quad (5.42)$$

6. Общие замечания.

В формуле (5.36) раскроем скобки при матрице $(I - R)$ и, пользуясь (5.22), перепишем ее, сделав некоторые преобразования

$$\begin{aligned} \hat{Q}_y &= Q_y - Q_y B^T (BQ_y B^T + Q_w)^{-1} BQ_y + \\ &+ Q_y B^T Q^{-1} A (A^T Q^{-1} A)^{-1} A^T Q^{-1} BQ_y. \end{aligned} \quad (6.1)$$

Пользуясь тождеством (1.14) и подставляя (5.23), преобразуем выражение (6.1) к виду

$$\hat{Q}_y = (B^T Q_w^{-1} B + Q_y^{-1})^{-1} + H \hat{Q}_x H^T. \quad (6.2)$$

Здесь для краткости введено обозначение

$$H = Q_y B^T Q^{-1} A. \quad (6.3)$$

Формула (6.2) показывает, из чего складывается дисперсия вектора y . В нее входят как формально принятые дисперсии Q_w и Q_y , так и вычисленная дисперсия вектора постоянных параметров. Если в модели (4.1) принять $Ax=0$, то последнее слагаемое в (6.2) исчезнет.

Применение МСКК полезно в тех случаях, когда есть подозрение, что на наблюдения влияет некоторый систематический фактор, который, тем не менее, нельзя параметризовать, то есть нельзя представить вектор наблюдений в виде параметрической модели (3.1). В некоторых случаях этот систематический фактор просто не поддается подобному описанию.

Если известны статистические характеристики поведения параметров, "подозрительных" на переменность в течение 24-часовой серии наблюдений, то предоставляется возможность использовать эту информацию при оценивании, для того чтобы получить более надежные результаты. Введение в оценку вектора x ковариационной матрицы Q_y равносильно устранению систематических компонент, вызывающих смещение этой оценки.

Условие (3.1), записанное в классической форме, имеет вид

$$S = \sum_{i=1}^N \frac{w_i^2}{\sigma_i^2}, \quad (6.4)$$

в соответствии с которым условие минимума данного функционала есть условие минимума суммы квадратов случайных ошибок (нормированных к априорным дисперсиям). Отсюда и возникло название – метод наименьших квадратов. Использование матрично-векторной формы записи, как гораздо более компактной, исключило из формул квадраты, но название метода и его суть от этого не изменились.

Очень часто возникает вопрос: почему в (6.4) используются именно квадраты случайных ошибок? Ведь можно использовать что-то другое, например, абсолютные величины случайных ошибок. Использовать, конечно, можно, но квадраты в данном случае занимают выделенное положение.

Дело в том, что выбор метода оценивания определяется видом функции распределения случайных ошибок. Если случайные ошибки распределены по нормальному закону, то условная плотность вероятности (4.1) представит в виде

$$p(h|x) = \frac{1}{(2\pi)^{N/2} \sigma_0} \exp \left[-\frac{1}{2} \sum_{i=1}^N \frac{(h - Ax)_i^2}{\sigma_i^2} \right]. \quad (6.5)$$

В формуле (6.5) суммируются квадраты ошибок $w = h - Ax$ в соответствии с выражением для нормального закона распределения случайных величин. Применим метод максимального правдоподобия, который дает оценки (4.5), совпадающие при принятом условии с оценкой МНК (3.3). Очевидно, что после дифференцирования квадраты в (6.5) из-под экспоненты "спустятся" вниз, и условие для вычисления оценки ММП (4.4) будет совпадать с (6.4). Таким образом, классический метод наименьших квадратов "вытекает" из нормального закона распределения случайных величин. Если

бы по каким-то причинам случайные ошибки имели другой вид распределения, отличающийся от нормального, то условная плотность вероятности (6.5) представлялась бы другой формулой, и оценка ММП отличалась бы от оценки МНК. В этом случае какой-то другой метод оценивания может оказаться более предпочтительным. Например, при большом количестве грубых ошибок функция распределения будет напоминать гауссиану с высоко поднятыми крыльями. В этом случае при использовании МНК обнаруживается много "выбросов", которые могут привести к смещению оценок, т.е. к неверному результату. Чтобы избежать этого применяют метод наименьших модулей, который менее чувствителен к наличию "выбросов". Тем не менее, как уже было сказано, допустимо применять МНК даже в тех случаях, когда форма распределения случайных ошибок незначительно отличается от гауссианы. По-видимому, успех МНК в начале 19 века и его широкое распространение в дальнейшем связано с тем, что реальные наблюдения достаточно редко содержат ошибки, которые придают распределению аномальный характер.

С другой стороны, ничто формально не мешает использовать какой-то иной метод (суммируя модули или четвертые степени случайных ошибок вместо суммы квадратов в числителе (6.4)), даже если случайные ошибки удовлетворяют нормальному закону распределения. Следует только помнить о том, что полученные экзотические оценки уже нельзя будет интерпретировать в смысле общепринятых статистических терминов (математическое ожидание, дисперсия и т.д.).

ЛИТЕРАТУРА

1. Браммер К., Зиффлинг Г., Фильтр Калмана-Бьюси, М.; Наука, 1982.
2. Валеев С.Г., Регрессионное моделирование при обработке наблюдений. М.; Наука, 1991.
3. Гаусс К.Ф., Избранные сочинения. Т.1; Способ наименьших квадратов. М.; Геодезиздат, 1957.
4. Гнеденко Б.В., Курс теории вероятностей. М.; Физматгиз, 1961.

5. Губанов В.С., Титов О.А., Оценивание стохастических параметров обобщенным методом наименьших квадратов. Сообщения ИПА РАН, 60, 1994.
6. Губанов В.С., Финкельштейн А.М., Фридман П.А., Введение в радиоастрометрию. М.; Наука, 1983.
7. Идельсон Н.И., Способ наименьших квадратов, Л.; Гостехиздат, 1932.
8. Мориц Г., Современная физическая геодезия. М.; Недра, 1980.
9. Мудров В.И., Кушко В.Л., Методы обработки измерений: квазиподобные оценки. М.; Наука, 1983.
10. Рао С.Р., Линейные статистические методы и их применения. М.; Наука, 1968.
11. Себер Д., Линейный регрессионный анализ. М.; Мир, 1980.
12. Сейдж Э., Мелса Дж., Идентификация систем управления. М.; Наука, 1974.
13. Сейдж Э., Мелса Дж., Теория оценивания и ее применение в связи и управлении. М.; Связь, 1976.
14. Титов О.А., Применение среднеквадратической коллокации для обработки РСДБ наблюдений. Сообщения ИПА РАН, 96, 1996.
15. Brosche, P., Oceanic tides and the rotation of the Earth, In W. Fricke, G. Teleki and D. Reidel (eds.), Sun and Planetary System, pp. 179-184, 1982.
16. Brosche, P., U. Seiler, J. Sündermann, J. Wünsch, Periodic changes in Earth's rotation due to oceanic tides, Astron. Astrophys., 220, pp. 318-320, 1989.
17. Charlot P., IERS Tech. Note 14, Int. Earth Rotation Service., Paris Observatory., Paris, 1993.
18. Gipson, J., 1996, Very long baseline interferometry determination of neglected tidal terms in high-frequency Earth orientation variation, Journal of Geoph. Res., 101, pp. 28051-28064.
19. Gontier A.-M., Orientation de la Terre par mesure VLBI Contribution a la chaine d'analyse de VLBI GLORIA. Ph.D. Thesis, Paris Obs., Paris, 1992.
20. Herring T.A., Dong D., Measurement of diurnal and semidiurnal rotational variations and tidal parameters of Earth. Journal of Geoph. Res., 99, pp. 18051-18071, 1994.
21. Herring T.A., Davis J.L., Shapiro I.I., Geodesy by Radio Interferometry: The application of Kalman Filtering to the analysis of Very Long Baseline Interferometry data. Journal of Geoph. Res., 95, pp. 12561-12581, 1990.

22. Himwich W.E., MacMillan D.S., Herring T.A., Ma C., Ryan J.W., Consistency of Geodetic Information From Kalman Filtering and Batch Least-Squares, Proc. of AGU Chapman Conference on Geodetic VLBI: Monitoring Global Change. pp. 106-110, 1991.
23. Jazwinski A.H., Stochastic processes and filtering theory. Academic Press Inc., New York and London, 1970.
24. Ma C., Sauber J., Bell L., Clark T., Gordon D., Himwich W., Ryan J., Measurement of Horizontal Motions in Alaska Using Very Long Baseline Interferometry. Journal of Geoph. Res., 95, pp. 21991-22011, 1990.
25. Ma C., Ryan J., NASA Space Geodesy Program GSFC, Data Analysis – 1995, VLBI Geodetic Results 1979-1995.5. GSFC, Greenbelt, Maryland, 1995.
26. MacMillan D.S., Gipson, J.M., Atmospheric pressure loading parameters from very long baseline interferometry observations. Journal of Geoph. Res., 99, pp. 18081-18087, 1994.
27. McCarthy D.D., (ed.), IERS Conventions 1996, Paris Observatory, 1996.
28. Moritz H., Least-Squares Collocation, Deutsche Geodätische Kommission, Reihe A, 75, München, 1973.
29. Niell A.E., Global mapping functions for the atmosphere delays at radio wavelengths. Journal of Geoph. Res., 101, pp. 3227-3246, 1996.
30. Schuh H., Tesmer, V., VLBI data analysis with a Full Variance-Covariance Matrix, XXII IUGG General Assembly, Birmingham, 1999.
31. van Dam T.M., Herring T. A., Detection of atmospheric pressure loading using very long baseline interferometry measurements. Journal of Geoph. Res., 99, pp. 4505- 4517, 1994.