

УДК 519.7

МЕТОД НАИМЕНЬШИХ МОДУЛЕЙ И МЕТОД НАИМЕНЬШИХ КВАДРАТОВ В ЗАДАЧАХ ОЦЕНИВАНИЯ ПРИ НАЛИЧИИ СКАЧКОВ ФАЗОВЫХ ТРАЕКТОРИЙ И СБОЕВ В ИЗМЕРЕНИЯХ

П.А. Акимов

Механико-математический факультет МГУ имени М.В. Ломоносова

РФ, 119991, Москва, Ленинские горы, 1

E-mail: akmpavel@rambler.ru

А.И. Матасов

Механико-математический факультет МГУ имени М.В. Ломоносова

РФ, 119991, Москва, Ленинские горы, 1

E-mail: alexander.matasov@gmail.com

Ключевые слова: метод наименьших модулей, метод наименьших квадратов, оптимизация, задачи сглаживания.

Аннотация: Рассматривается задача оценивания состояния линейной динамической системы по косвенным измерениям фазового вектора. Предполагается, что некоторые ошибки измерений могут быть аномально большими. Кроме того, возможное наличие редких импульсных возмущений в динамической системе приводит к скачкам в фазовых траекториях. В этом случае применение классических квадратичных методов приводит к «размытым» оценкам траекторий в окрестностях скачков. Предлагается при решении таких задач использовать метод наименьших модулей. Проводится сопоставление двух методов. Показано, что применение метода наименьших модулей позволяет получить «контрастные» оценки траекторий, отчетливо выделяющие моменты скачков. Более того, такие оценки мало чувствительны к аномально большим сбоям в измерениях.

1. Введение

Среди задач обработки измерений важное место занимают проблемы оценивания неизвестных параметров системы по имеющимся данным (измерениям, наблюдениям). Наиболее распространенные подходы к их решению основаны на квадратичных методах оптимизации – методе наименьших квадратов (МНК) или фильтре Калмана [1]. Однако в ряде ситуаций измерения могут содержать аномальные ошибки, а динамика рассматриваемой системы – допускать «скачки» фазовых переменных. В этом случае хорошую точность оценок доставляет метод

наименьших модулей (МНМ, l_1 -аппроксимация) [2–4], сравнению которого с МНК посвящена данная работа. Далее представлены вариационные проблемы, соответствующие МНМ и МНК, рассмотрены алгоритмы их решения, а также приведен численный пример, иллюстрирующий свойства получаемых оценок.

2. Постановка задачи

Применим МНК и МНМ для решения задачи сглаживания в случае дискретной динамической системы

$$(1) \quad x(k+1) = Fx(k) + Gq(k) + g(k), \quad k = 0, \dots, K-1,$$

где $x(k) \in \mathbf{R}^n$ – неизвестный вектор состояния системы в момент времени k , $q(k) \in \mathbf{R}^l$ – неизвестный вектор погрешностей динамической системы, $F \in \mathbf{R}^{n \times n}$, $G \in \mathbf{R}^{n \times l}$ – заданные матрицы (для простоты рассматривается стационарная система), $g(k) \in \mathbf{R}^n$ – известный вектор.

Измерения компонент фазового вектора $x(k)$ описываются линейными уравнениями: $z(k) = Hx(k) + r(k)$, $k = 0, \dots, K$. Здесь $z(k) \in \mathbf{R}^m$ – вектор измерений в момент времени k , $H \in \mathbf{R}^{m \times n}$ – заданная матрица, $r(k) \in \mathbf{R}^m$ – погрешность измерений. В начальный момент времени $x(0)$ известен с некоторой погрешностью $\bar{r}(0)$: $\bar{x}(0) = x(0) + \bar{r}(0)$, $\bar{x}(0), \bar{r}(0) \in \mathbf{R}^n$.

Необходимо оценить значения векторов состояния $x(k)$ на всем интервале времени $k = 0, \dots, K$, используя весь доступный массив измерений. Рассмотрим задачу в детерминированной постановке, в которой не вводятся гипотезы о вероятностных свойствах погрешностей, а заданы лишь их характерные величины R_i, Q_i, Π_i : $r_i(k) \sim R_i$, $q_i(k) \sim Q_i$, $\bar{r}_i(0) \sim \Pi_i$.

Введем краткие обозначения для неизвестных векторов в описанной системе $(x, q) = (x(0), \dots, x(K), q(0), \dots, q(K-1))$, и для весовых матриц $\Pi^{-1} = \text{diag}(\Pi_1^{-1}, \dots, \Pi_n^{-1})$, $Q^{-1} = \text{diag}(Q_1^{-1}, \dots, Q_l^{-1})$, $R^{-1} = \text{diag}(R_1^{-1}, \dots, R_m^{-1})$.

Вариационная проблема для метода наименьших квадратов (l_2 -сглаживание) для рассматриваемой системы формулируется следующим образом:

$$(2) \quad J(x, q) = \left(\left\| \Pi^{-1}(\bar{x}(0) - x(0)) \right\|_2^2 + \sum_{k=0}^{K-1} \left\| Q^{-1}q(k) \right\|_2^2 + \sum_{k=0}^K \left\| R^{-1}(z(k) - Hx(k)) \right\|_2^2 \right)^{\frac{1}{2}} \rightarrow \inf_{(x, q)}$$

при ограничениях (1).

Метод наименьших модулей (l_1 -сглаживание) определяется следующей проблемой негладкой оптимизации [5]:

$$(3) \quad I(x, q) = \left\| \Pi^{-1}(\bar{x}(0) - x(0)) \right\|_1 + \sum_{k=0}^{K-1} \left\| Q^{-1}q(k) \right\|_1 + \sum_{k=0}^K \left\| R^{-1}(z(k) - Hx(k)) \right\|_1 \rightarrow \inf_{(x, q)}$$

при ограничениях (1).

Здесь и далее используются обозначения для норм векторов: $\|y\|_1 = \sum_i |y_i|$, $\|x\|_2 = (\sum_i y_i^2)^{1/2}$. Иными словами, в детерминированной постановке необходимо минимизировать сумму l_1 - или l_2 -норм векторов невязок, соответствующих погрешностям начальной информации, динамической модели и измерений.

Наряду с вышеуказанными задачами, авторами была также исследована проблема со смешанными нормами, в которой минимизируются l_1 -нормы, соответствующие невязкам в уравнении динамической системы $\|Q^{-1}q(k)\|_1$, и l_2 -нормы, соответствующие невязкам в измерительной информации $\|R^{-1}(z(k) - Hx(k))\|_2^2$. В работе [6] показано, что данный подход обладает главным преимуществом МНМ – возможностью точнее отслеживать скачки траекторий фазового вектора.

3. Решение оптимизационных задач

Решение задачи l_2 -сглаживания может быть найдено при помощи нескольких алгоритмов, использующих идеи фильтра Калмана, например, согласно алгоритму Брайсона-Фрейзера или так называемым формулам Рауха-Тунга-Стрибела [1], которые позволяют в два прохода по массиву обрабатываемых измерений найти решение квадратичной задачи сглаживания.

В отличие от МНК, в МНМ минимизируется негладкая функция, и как правило, методы решения таких задач носят рекуррентный характер и направлены на поиск приближенного решения. Один из таких алгоритмов, идея которого предложена в [2] – метод весовой и временной рекурсий – рассмотрен авторами данной работы [5]. Его суть состоит в том, чтобы аппроксимировать модули всех невязок, например, $|Q_i^{-1}q_i(k)|$, $|R_j^{-1}(z_j(k) - H_jx(k))|$ квадратичными функциями вида $q_i(k) Q_{W_i}^{-2}(k, s) q_i(k)$, $(z_j(k) - H_jx(k)) R_{W_j}^{-2}(k, s) (z_j(k) - H_jx(k))$. Процедура решения состоит из двух вложенных циклов. Первый цикл представляет собой последовательность итераций, на каждой из которых ищется минимум квадратичной функции с весовыми коэффициентами, зависящими от решения, полученного на предыдущей итерации:

$$(4) \quad J_{(s+1)} = \inf_{(X,q)} \left((\bar{x}(0) - x(0))^T \Pi_W^{-2}(s) (\bar{x}(0) - x(0)) + \sum_{k=0}^{K-1} q^T(k) Q_W^{-2}(k, s) q(k) + \sum_{k=0}^K (z(k) - Hx(k))^T R_W^{-2}(k, s) (z(k) - Hx(k)) \right)^{\frac{1}{2}}$$

при ограничениях (1).

Здесь $\Pi_W^{-2}(s)$, $Q_W^{-2}(k, s)$, $R_W^{-2}(k, s)$ – диагональные матрицы весовых коэффициентов, элементы которых зависят как от момента дискретного времени k , так и от номера предыдущей итерации s . В простейшем случае диагональные элементы обратно пропорциональны модулям невязок, полученных на предыдущем шаге, например, $Q_{W_i}^{-2}(k, s) \sim Q_i^{-1} |q_i(k, s-1)|^{-1}$. В случае близких к нулю значений невязок применяется регуляризация значений весов.

Второй, внутренний цикл реализует формулы Брайсона-Фрейзера для поиска решения аппроксимирующей задачи сглаживания (4), (1). После нахождения приближенного решения $(x(s), q(s))$ веса пересчитываются и формируется новая проблема вида (4), (1). И так продолжается до выполнения выбранного критерия остановки. Для этой цели при помощи теории двойственности строится оценка уровня неоптимальности текущего приближенного решения [4–6]. Таким образом, вычислительная сложность решения задачи МНМ кратно выше, чем сложность решения задачи МНК, поскольку предполагает дополнительный цикл вычислений, количество итераций в котором, вообще говоря, трудно предсказать.

4. Сравнение методов

Рассмотрим пример применения вышеописанных методов для модельной динамической системы, возникающей при прогнозировании кредитных рисков. Задача состоит в оценке коэффициентов миграции (т.н. FlowRates), соответствующих средним по портфелю вероятностям перехода кредита за месяц из одного рейтинга или диапазона просрочки в следующий, более плохой (например, вероятность того, что просрочка увеличится с 60 до 90 дней). Необходимость в такой оценке зачастую возникает при прогнозировании кредитных потерь для продуктов банка [7].

Фазовый вектор $x(k) \in \mathbf{R}^{16}$ здесь состоит из 13 коэффициентов миграции, соответствующих тринадцати диапазонам просрочки (от 0 до 12 месяцев), а также еще 3 вспомогательных параметров, описывающих динамику кредитного портфеля (задолженности по кредитам) и «дрейфам» наиболее нестабильных коэффициентов. Матрицы динамической системы и измерений имеют вид: $G = \mathbf{I}_{16}$,

$$F = \mathbf{I}_{16} + \left(\begin{array}{ccc|ccc} \mathbf{0}_{1 \times 13} & 0 & 0 & 1 & & & \\ \mathbf{0}_{12 \times 13} & & \mathbf{0}_{12 \times 3} & & & & \\ \mathbf{0}_{1 \times 13} & 0 & 1 & 0 & & & \\ \mathbf{0}_{2 \times 13} & & \mathbf{0}_{2 \times 3} & & & & \end{array} \right) \in \mathbf{R}^{16 \times 16}, \quad H = \left(\begin{array}{cccc|ccc} -1 & 0 & \dots & 0 & -1 & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ & & \dots & & & & \\ 0 & 0 & \dots & 1 & 0 & 0 & 0 \end{array} \right) \in \mathbf{R}^{14 \times 16},$$

где \mathbf{I}_{16} , $\mathbf{0}_{12 \times 13}$ — единичная матрица и матрица из нулевых элементов соответственно.

Входными данными являются обезразмеренные объемы задолженности в каждом из 13 диапазонов, наблюдаемые в течение 42 месяцев подряд. Отметим, что решаемая здесь проблема является модельной в том смысле, что измерения для нее формировались искусственно. Типичные значения погрешностей r_i, q_i варьировались в пределах $[0,005; 0,1]$ в зависимости от коэффициента миграции, за исключением погрешностей в динамике вспомогательных переменных q_{14}, q_{15}, q_{16} , которые принимали значения порядка $5 \cdot 10^{-4}$ (все фазовые переменные безразмерные).

На практике в данных часто наблюдаются аномалии двух типов: отдельные всплески просрочки, объясняющиеся сезонностью и операционными сбоями, и ступенчатые изменения, вызванные, как правило, измерениями в бизнес-процессах. Поэтому проблема правильной идентификации скачков и, наоборот, игнорирования сбоев (их неучет в оценках), является актуальной. В рассматриваемом примере $n = 16, l = 16, m = 13, K = 42$, количество неизвестных параметров 1360, количество измерений 1290. В созданные временные ряды добавлены 2 сбоя (всплеск погрешности измерений в 3 – 5 раз) и одно ступенчатое изменение переменной (увеличение ее значения в 2 раза).

Проблема (3), (1) решалась методом весовой и временной рекурсий. Результаты оценки компонент, в которых были сбои и скачки, показаны на рис. 1. Здесь по оси абсцисс указаны дискретные моменты времени k , а по оси ординат – измерения и оценки двух компонент фазового вектора, в которые были внесены сбои. Как видим, в условиях наличия «аномальных» эффектов в измерениях и в динамике системы с одной стороны, и при небольшом по длительности отрезке времени, на котором заданы измерения, с другой, МНМ позволил избежать искажения оценок из-за сбоев и правильно выявить скачок в траектории. МНК в этой задаче дает возможность лишь оценить характерный масштаб переменной, в которой был скачок, и тенденцию ее изменения, а сбой приводит к существенному завышению оценки. При этом, как уже отмечалось, МНМ вычислительно сложнее: поиск решения

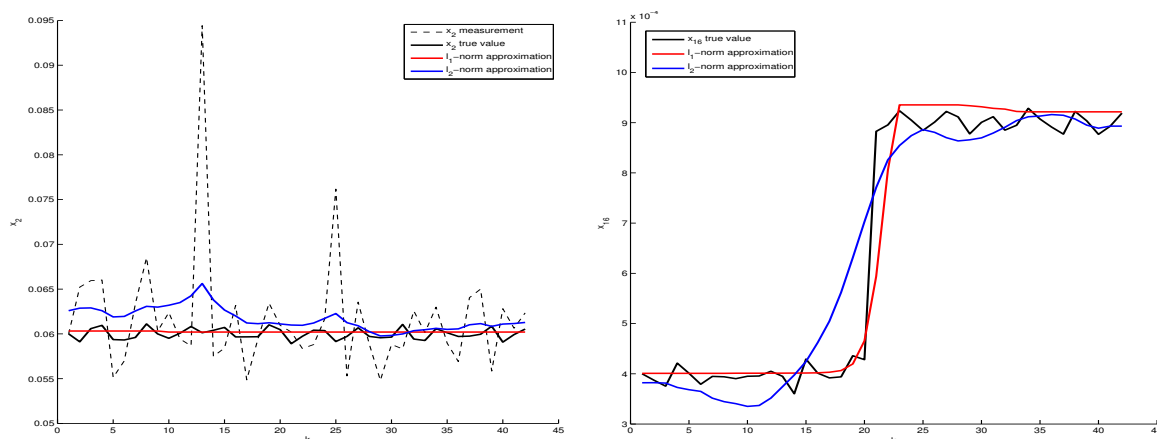


Рис. 1. Оценки компонент x_2, x_{16} фазового вектора

занял 14 секунд против 0,3 секунд для МНК. Преимущество МНМ по сравнению с МНК проявляется и во многих других примерах, в том числе, характеризующихся большими размерностями K, m, n, l [4–6].

5. Заключение

Сравнительный анализ метода наименьших модулей и метода наименьших квадратов показывает, что свойство МНМ формировать решения, малочувствительные к аномально большим ошибкам в данных, является его преимуществом в случаях, когда рассматриваемые системы допускают скачки и выбросы в траекториях, а МНК, особенно в условиях не слишком большого количества наблюдений, формирует излишне «осредненные» решения, не позволяющие идентифицировать реальный характер изменения параметров системы. Также стоит подчеркнуть, что разработанные численные алгоритмы позволяют решать задачи МНМ для разнообразных динамических систем и в случае измерительных интервалов большой длительности.

Список литературы

1. Kailath T., Sayed A.H., Hassibi B. Linear Estimation. Upper Saddle River, New Jersey: Prentice Hall, 2000.
2. Мудров В.И., Кушко В.Л. Метод наименьших модулей. М.: Знание, 1971.
3. Гильбо Е.П., Челпанов И.Б. Обработка сигналов на основе упорядоченного выбора (мажоритарное и близкие к нему преобразования). М.: Советское радио, 1976.
4. Акимов П.А., Деревянкин А.В., Матасов А.И. Гарантирующий подход и l_1 -аппроксимация в задачах оценивания параметров БИНС при стендовых испытаниях. М.: Изд-во МГУ, 2012.
5. Akimov P.A., Matasov A.I. An iterative algorithm for l_1 -norm approximation in dynamic estimation problems // Automation and Remote Control. 2015. Vol. 76, No. 5. P. 733-748.
6. Akimov P.A., Matasov A.I. Weight and time recursions in dynamic state estimation problem with mixed-norm cost function // IEEE Trans. Automatic Control. 2015. Vol. AC-60, No. 4. P. 1050-1063.
7. Стежкин А.А. Применение матриц миграции в задачах оценки и управления кредитным риском банков в рамках подхода на основе внутренних рейтингов заемщиков // Проблемы управления. 2016. №. 4. P. 37-44.