МЕТОДЫ ОБНАРУЖЕНИЯ ЗАКОНОМЕРНОСТЕЙ С ПОМОЩЬЮ ЭВМ
(Вычислительные системы)
1981 год Выпуск 91

УДК 518:512.25:519.6

СМЕЩЕННОЕ ОЦЕНИВАНИЕ В ЛИНЕЙНЫХ МОДЕЛЯХ С ПОМОЩЬЮ ОБОБЩЕННЫХ ОБРАТНЫХ МАТРИЦ

Б.Н.Луценко

1. Введение. При построении автоматизированных систем обработки данных немаловажно обеспечение надежного и бесперебойного процесса вычислений. В ходе решения систем линейных уравнений возможно вырождение или плохая обусловленность матриц. Сама по себе задача линейного оценивания может возникнуть как в сугубо линейных, так и в нелинейных моделях, в которых прибегают к линеаризации на каждом шаге итерации. Обычно в случае плохой обусловленности матриц для получения устойчивого решения прибегают к методам регуляризации [1] или используют аналогичные методы из ридж регрессии [2,3]. При этом привлекаются операции обычного обращения матриц. Иной подход связан с переходом к операциям обобщенного обращения с регулируемым порогом точности при вычислении ранга матрицы [4]. Варьируя этим порогом и переходя в область смещенных оценок,можно не только обеспечить устойчивость решения, но и в ряде случаев повысить его точность. В некоторых ситуациях стабилизация решения за счет легко реализуемой редукции ранга матрицы предпочтительнее регуляризация, так как позволяет использовать достаточно грубую априорную информацию о векторе параметров.

2. Модель наблюдений. Будем полагать, что наблюдения осуществляются согласно схеме

$$y_i = \sum_{j=1}^{m} a_{ij} x_j + \xi_i ,$$

$$M(\xi_i) = 0, \quad M(\xi_i \xi_j) = \delta_{ij}\sigma^2; \quad i,j = \overline{1,n} , \tag{1}$$

где δ_{ij} - символ Кронекера.

Распределение ξ_i в остальном может быть произвольным. В матричной записи модель имеет вид $Y = AX + \Xi$, смысл символов которой очевиден из (I). Известные компоненты матрицы A могут представлять значения временных (или пространственных) базисных функций $a_{ij} = a_j(t_i)$, заданных на множестве n точек ($i = 1,n$). Базис может быть избыточен, линейно-зависим.

В природе и технике широко используется дублирование и резервирование элементов. Этим достигается высокая надежность функционирования систем и, кроме того, возможность получения значимых макроскопических эффектов за счет суммирования большого числа микроскопических эффектов. Линеаризованная модель такой избыточной системы может соответствовать схеме (I). Выбирая определенным образом значения x_j (усилия) при заданном базисе A, можно достичь $M(Y)$ (требуемого эффекта) при минимальной квадратичной норме вектора X. В энергетической трактовке это означает возможность достижения требуемого эффекта при минимальных энергетических затратах. Этот принцип весьма часто реализуется в природе и технике. Таким образом, неполнота ранга матрицы A не обязательно свидетельствует о неудачно выбранном базисе. Определенному классу задач вырожденность A присуща принципиально и может свидетельствовать как раз о достоинствах описываемой системы. Обобщенные обратные матрицы являются инструментом, рассчитанным на работу и с такими моделями. В частности, псевдообратная матрица Мура-Пенроуза позволяет получать решение с минимальной квадратичной нормой, удовлетворяющей нормальной системе уравнений. При этом исходная система уравнений может быть как переопределена, так и недоопределена.

3. Обобщенные обратные матрицы. Понятие обобщенной обратной матрицы было введено в 1935 г. Муром. Лишь в пятидесятые годы появились следующие публикации в этой области, и наибольшую известность среди них получила работа Пенроуза [5, с. 3-228; 7], которая и привлекла широкое внимание к этим вопросам. О динамике развития этой области некоторое представление дает нарастающий поток публикаций. Если в [7] библиография содержала 360 наименований, в [6, с. 135-174] - 420, то в тщательно структурированной по тематике библиографии, приведенной в [9], она насчитывала 1776.

Операция обобщенного обращения, в отличие от обычного, определена и для прямоугольных матриц. Мы ограничимся вещественными матрицами. Пусть A – исходная матрица порядка $n \times m$, а G – $m \times n$ такие, что удовлетворяются некоторые из приведенных ниже соотношений

$$A\,G\,A = A, \qquad (2)$$

$$G\,A\,G = G, \qquad (3)$$

$$(A\,G)^T = A\,G, \qquad (4)$$

$$(G\,A)^T = G\,A\,. \qquad (5)$$

Матрицу G, удовлетворяющую всем четырем условиям (2)-(5), называют псевдообратной матрицей Мура-Пенроуза. Она единственна. Будем в дальнейшем обозначать ее символом A^+. Можно потребовать выполнения некоторых условий, а именно: (2); (2) и (3); (2) и (4); (2) и (5); (2)-(4); (2), (3) и (5); (2), (4) и (5) [5, с. 26-30; 7]. Для решения ряда задач вполне достаточно таких обобщенных обратных или полуобратных матриц, задаваемых неоднозначно [7,8]. Эта неоднозначность обычно используется для упрощения алгоритмов их вычисления. Выделим полуобратную матрицу, удовлетворяющую (2)-(4), и обозначим ее A^-. Из перечисленных семи лишь она при $r\,k\,A = m$ совпадает с A^+. В последующем мы будем работать главным образом с A^+ и A^-, сопровождая индексами "+" и "–" получаемые с их помощью оценки. Приведем поэтому некоторые дополнительные сведения о них.

Из множества прямых методов вычисления псевдообратных матриц выделяются по своей устойчивости, точности и быстродействию методы, использующие гауссово исключение, преобразование отражения (Хаусхолдера) и двучленную ортогонализацию Грама-Шмидта [9, с. 245-302; 10, 11]. Мы будем ориентироваться главным образом на двучленную ортогонализацию Грама-Шмидта.

УТВЕРЖДЕНИЕ I. *Пусть A - матрица порядка $n \times m$ ранга r, факторизованная по схеме*

$$A = A_H \cdot C, \qquad (6)$$

где A_H - ортогональная матрица порядка $n \times r$, $A_H^T\,A_H = I_r$, а C - матрица порядка $r \times m$. Тогда

$$A^+ = C^+\,A_H^T\,. \qquad (7)$$

В частности, $A^+ = C^T(C\,C^T)^{-1}\,A_H^T$.

В последующем нам потребуется и другая форма представления A^+, основанная на сингулярном разложении A.

УТВЕРЖДЕНИЕ 2. Пусть A – матрица порядка $n \times m$ ранга r и

$$A = U \Lambda V^T, \tag{8}$$

где U и V – матрицы порядка $n \times r$ и $m \times r$ с ортогональными столбцами $U^T U = I_r$, $V^T V = I_r$, а Λ – диагональная матрица порядка $r \times r$ с положительными элементами. Тогда $A^+ = V \Lambda^{-1} U^T$.

Матрица A^+ обладает тем замечательным свойством, что при решении системы линейных уравнений приводит к вектору оценок с минимальной квадратичной нормой.

УТВЕРЖДЕНИЕ 3. Пусть A – матрица порядка $n \times m$ ранга r и A_H – ортогональная матрица $A_H^T A_H = I_r$, порождающая в R_n то же линейное многообразие (образ), что и A, т.е.

$$\mathcal{L}(A_H) = \mathcal{L}(A). \tag{9}$$

Тогда если θ – матрица порядка $r \times m$, такая что

$$A_H = A \cdot \theta, \tag{10}$$

то

$$A^- = \theta \cdot A_H^T, \tag{11}$$

а при $r = m$ $A^- = A^+$.

ДОКАЗАТЕЛЬСТВО. Существование θ вытекает из (9). Удостоверимся в выполнимости (2):

$$A A^- A = A \cdot \theta A_H^T A = A_H \cdot A_H^T A.$$

В силу того же предположения (9) существует некоторая матрица C – $r \times m$, удовлетворяющая (6). Следовательно, $A_H A_H^T A = A_H A_H^T A_H C = A_H C = A$.

Аналогично проверяем выполнимость (3) $A^- A A^- = \theta A_H^T (A \theta) A_H^T = \theta (A_H^T A_H) A_H^T = \theta A_H^T$ и (4) $A A^- = A \theta A_H^T = A_H A_H^T$. При $r = m$ θ квадратна и невырожденна и из (10) следует $A = A_H \theta^{-1}$, вследствие чего $A^- A = \theta A_H^T A_H \theta^{-1} = \dot{I}$, что и завершает доказательство утверждения 3.

Справедливо и обратное

УТВЕРЖДЕНИЕ 4. Л ю б а я A^- д о п у с к а е т п р е д с т а в л е н и е (II), в к о т о р о м A_H и θ с в я з а н ы с и с х о д н о й м а т р и ц е й A с о о т н о ш е н и е м (I0).

ДОКАЗАТЕЛЬСТВО. В этом легко убедиться, приняв в качестве исходной матрицу $B = (A^T)^-$ и использовав затем утверждение 3. В остальном выбор A_H, удовлетворяющей (9), произволен. При фиксированной A_H матрица θ определяется также неоднозначно, что можно использовать для упрощения алгоритма ее вычисления. В частности, θ можно вычислить опираясь на

УТВЕРЖДЕНИЕ 5. П у с т ь A_H в (6) и (I0) с о в п а д а ю т. Т о г д а $\theta = C^-$.

ДОКАЗАТЕЛЬСТВО. Из (6) и (I0) следует $A\,\theta = A_H\, C\, \theta = A_H$. Поскольку из строк A_H размерности $m \times r$ и ранга $r \le n$ всегда можно сформировать квадратную подматрицу ранга r (назовем ее A_H^*), и при этом $A_H^* C\, \theta = A_H^*$, то

$$C\,\theta = \dot{I}_r\,, \qquad (12)$$

откуда и следует утверждение 5.

4. О б у с л о в л е н н о с т ь. В настоящее время нет единого определения понятия обусловленности [5, с. I3-I6]. В общем число обусловленности должно характеризовать меру устойчивости решения по отношению к различным видам возмущений входных данных. Наиболее удобным для наших иллюстративных целей будет следующее определение числа обусловленности [5, с. I4; 9, с. 279; I0,II]: $\mathrm{cond}(A) = \|A\| \cdot \|A^+\| \ge 1$. Достоинством его является применимость к прямоугольным матрицам неполного ранга в случае различных норм, выбор которых, в свою очередь, может диктоваться методом вычисления псевдообратной матрицы. При квадратичной норме $\mathrm{cond}\,(A) = \|A\|_2 \cdot \|A^+\|_2 = \lambda_{max}/\lambda_{min}$, где λ_{max} и λ_{min} – максимальное и минимальное сингулярные значения матрицы A (см. (8)). При этом $\mathrm{cond}\,(A^T A) = (\mathrm{cond}\,(A))^2 \ge \mathrm{cond}(A)$. Тем самым при переходе к нормальной системе уравнений

$$A^T A\, X = A^T Y \qquad (13)$$

обусловленность ухудшается. Попытка же скомпенсировать это ухудшение повышением точности вычислений может привести к обратному результату, если априори не известен ранг матрицы A.

УТВЕРЖДЕНИЕ 6. Пусть действие возмущений, приведенных ко входу, таково, что вместо исходной матрицы A фактически имеется $B = A+\Delta A$, причем $\mathrm{rk}B > \mathrm{rk}A$. Тогда [9, с. 259] $\|B^+\|_2 \geq \dfrac{1}{\|\Delta A\|_2}$.

Таким образом, повышая точность вычислений и этим уменьшая $\|\Delta A\|_2$, можно еще более ухудшить точность оценки, так как число обусловленности выступает в качестве множителя в погрешностях, вызываемых возмущениями входных данных [10] и с ростом это числа растет случайная компонента погрешности оценки вектора параметров X.

Регуляризации обычно подвергается именно нормальная система уравнений. При этом как бы легализуется возможное расширение за счет возмущений размерности подпространства параметров $\mathcal{X}(A)$. Иной подход, о котором упоминалось во введении, связан с назначением некоторого порога точности при вычислении ранга A. В этом случае можно работать с исходной несовместной системой уравнений, не переходя к нормальной системе, т.е. искать оценки, скажем, вида $\hat{X}^+ = A^+Y$ вместо эквивалентных им теоретически (но не в вычислительном плане) $\hat{X}^+ = (A^TA)^+ A^TY$. Повышение точности вычислений здесь уже не вызовет обратного эффекта, а напротив, может облегчить определение истинного ранга.

5. Оценивание параметров. Будем в дальнейшем полагать систему уравнений переопределенной, т.е. $n \geq m$. Если A - полного ранга, то оценки

$$\hat{X}^+ = A^+ Y, \tag{14}$$

$$\hat{X}^- = A^- Y \tag{15}$$

теоретически тождественны $\hat{X}^+ = \hat{X}^- = \hat{X}_{МНК} = (A^T A)^{-1}A^TY$. Свойства оценок наименьших квадратов известны. В вычислительном плане более устойчивы оценки $\hat{X}^+$ и $\hat{X}^-$ (см. п.4). Возможна, однако, ситуация, например, в алгоритме рекуррентного уточнения оценок $\hat{X}$ по мере поступления новых данных, когда разумнее использовать оценку типа $\hat{X}_{МНК}$, осуществляя в случае необходимости вычисления с повышенной точностью.

УТВЕРЖДЕНИЕ 7. Пусть $\operatorname{rk} A = r < m$. Тогда оценки $\hat{X}^+$ и $\hat{X}^-$ в общем случае смещены

$$M(X - \hat{X}^+) = (\dot{I} - C^+ C)X, \qquad (16)$$

$$M(X - \hat{X}^-) = (\dot{I} - \theta C)X \qquad (17)$$

и характеризуются ковариационными матрицами

$$\operatorname{cov}(\hat{X}^+) = \sigma^2 C^+ (C^+)^T, \qquad (18)$$

$$\operatorname{cov}(\hat{X}^-) = \sigma^2 \theta\, \theta^T, \qquad (19)$$

причем

$$\operatorname{tr} \operatorname{cov}(\hat{X}^-) \geq \operatorname{tr} \operatorname{cov}(\hat{X}^+). \qquad (20)$$

ДОКАЗАТЕЛЬСТВО. Выражение для смещения (16) следует непосредственно из (1), (6) и (7), смещение представляет в R_m компоненту вектора X, принадлежащую $\mathcal{N}(C^T)$, т.е. ядру оператора C^T. Аналогично, (17) следует из (I), (6) и (II). Оператор θC идемпотентен, однако в общем случае несимметричен и, следовательно, не является оператором проектирования. Таким образом, смещения (16) и (17), вообще говоря, различны, и если нет никакой дополнительной информации, их нельзя сопоставить. Несмещенной каждая из оценок может оказаться при различных условиях. Так, при $X \in \mathcal{L}(C^T)$ смещение $M(X - \hat{X}^+) = 0$, в то время как $M(X - X^-) \neq 0$. Если же, с другой стороны, $X \in \mathcal{L}(\theta)$ то $M(X - \hat{X}^-) = 0$, в то время как $M(X - \hat{X}^+) \neq 0$.

Выражения для ковариационных матриц (18) и (19) вытекают непосредственно из (1), (7), (II), (14) и (15). Справедливость (20) следует из (12) и свойств следа произведения матриц. Поскольку $C\theta = I_r$, то $\operatorname{tr} \operatorname{cov}(\hat{X}^+) = \sigma^2 \operatorname{tr} C^+(C^+)^T = \sigma^2 \operatorname{tr} C^T(CC^T)^{-2}C = \sigma^2 \operatorname{tr}(CC^T)^{-1} = \sigma^2 \operatorname{tr} \theta^T C^T (C C^T)^{-1} C\, \theta$.

Оператор $P_{\mathcal{L}(C^T)} = C^T(CC^T)^{-1}C$ является оператором проектирования на $\mathcal{L}(C^T)$ в R_m, он симметричен и идемпотентен $P = P^T = P^T P$. Следовательно, $\operatorname{tr} \theta^T P \theta = \operatorname{tr} \theta^T P^T P \theta \leq \operatorname{tr} \theta^T \theta$, так как $\operatorname{tr} \theta^T \theta$ представляет сумму квадратов модулей векторов-столбцов матрицы θ, а $\operatorname{tr} \theta^T P \theta$ – аналогичную сумму квадратов проекций векторов-столбцов на $\mathcal{L}(C^T)$. Использование иных типов полуобратных матриц для получения оценок $\hat{X}$ еще более усугубит различие с оценками $\hat{X}^+$. Если же есть основание предполагать, что описываемая моделью (I) система построена исходя из принципа экономичности, то тем более предпочтение следует отдать оценке $\hat{X}^+$, которая будет в этом случае еще и несмещенной в силу упомянутого выше свойства матрицы Мура-Пенроуза.

Мы пока не касались вопроса определения ранга, а вместе с тем на практике достаточно сложно определить истинный ранг матрицы. Устойчивые процедуры построения псевдообратных матриц обычно используют некоторый порог точности вычислений [4,12]. Выбирая определенным образом этот порог, можно не только стабилизировать процедуру вычисления, но и, переходя в класс смещенных оценок, повысить их точность.

УТВЕРЖДЕНИЕ 6. Пусть $rkA = r < m$ и $k \le r$, а

$$\left.\begin{aligned} &A = A_k + A_l, \\ &rkA_k = k; \quad rkA_l = l = r - k, \end{aligned}\right\} \qquad (21)$$

причем компоненты разложения (21) порождают ортогональные подпространства

$$\mathcal{L}(A_k) \perp \mathcal{L}(A_l) \qquad (22)$$

и элементы матрицы Грама $\Gamma_l = A_l^T A_l$ не превышают по модулю ε_l^2

$$|\{\Gamma_l\}_{ij}| \le \varepsilon_l^2; \quad i,j = \overline{1,m}. \qquad (23)$$

Тогда оценка

$$\hat{X}_k^+ = A_k^+ Y \qquad (24)$$

является смещенной $M(X - \hat{X}_k^+) = (I - A_k^+ A_k)X$ с ковариационной матрицей $cov(\hat{X}_k^+) = \sigma^2 A_k^+ (A_k^+)^T$ и средним квадратом ошибки

$$Q_k^+ = X(I - A_k^+ A_k)X + \sigma^2 tr(A_k^+ (A_k^+)^T). \qquad (25)$$

Если $\|X\|_2 \le \sigma E$, где E - положительная константа, то смещенная оценка (24) предпочтительнее оценки (14) по критерию (25) по крайней мере при

$$\varepsilon_l^2 \le l^2/(q\,E^2), \qquad (26)$$

где $m > q > 1$.

ДОКАЗАТЕЛЬСТВО. Вывод выражений для смещения и ковариационных матриц очевиден и мы перейдем сразу к доказательству (26). Воспользуемся сингулярным разложением A (см. (8))

$$A = U\Lambda_r V^T = U\Lambda_k V^T + U\Lambda_l V^T = A_k + A_l. \qquad (27)$$

Здесь Λ_r, Λ_k, Λ_l – диагональные матрицы $r \times r$ соответственно с r, k и l отличными от нуля положительными элементами

$$\Lambda_r = \Lambda_k + \Lambda_l; \quad \Lambda_k \cdot \Lambda_l = 0. \qquad (28)$$

Обозначим Λ_k^- диагональную матрицу, полученную из Λ_k заменой ее ненулевых элементов их обратными величинами. При этом

$$A^+ = V\Lambda_r^{-1}U^T = V\Lambda_k^- U^T + V\Lambda_l^- U^T = A_k^+ + A_l^+ . \qquad (29)$$

Средний квадрат ошибки $\hat{X}^+$ имеет вид (см. (16),(18))

$$Q^+ = X\,(I - A^+A)X + \sigma^2 \mathrm{tr}\, A^+(A^+)^T = X^T P_{\mathcal{N}(A)} X + \sigma^2\, \mathrm{r}[(\Lambda_k^-)^2 + (\Lambda_l^-)^2].$$

Здесь $P_{\mathcal{N}(A)}$ – оператор проектирования на ядро A. Аналогично $\mathrm{tr}[A_k^+(A_k^+)^T] = \mathrm{tr}\, V(\Lambda_k^-)^2 V = \mathrm{tr}(\Lambda_k^-)^2$. Очевидно, $X^T P_{\mathcal{N}(A_k)} X \ge$ $\ge X\, P_{\mathcal{N}(A)}\, X$, поскольку $\mathcal{N}(A_k) \supset \mathcal{N}(A)$.

Нас будет интересовать ситуация, при которой $Q^+ - Q_k^+ \ge 0$, т.е. когда

$$\mathrm{tr}(\Lambda_l^-)^2 \ge \frac{X^T A_l^T (P_{\mathcal{N}(A_k)} - P_{\mathcal{N}(A)})X}{\sigma^2} = \frac{X\, P_{\mathcal{L}(A_l^T)}\, X}{\sigma^2}. \qquad (30)$$

Поскольку $\mathrm{tr}(\Lambda_l)^2 = \mathrm{tr}(A_l^T A_l) = \mathrm{tr}\,\Gamma_l = q\,\varepsilon_l^2$, где $q < m$, а минимальное значение $\mathrm{tr}(\Lambda_l^-)^2$ достигается в случае равенства всех l отличных от нуля диагональных элементов Λ_l, т.е. когда каждый из них будет равен $1/(q\,\varepsilon_k^2)$, получаем $\mathrm{tr}(\Lambda_l^-)^2 \ge l^2/(q\,\varepsilon_l^2)$. Очевидно также, что $X\,P_{\mathcal{L}(A_l^T)}X \le X^T X \le \sigma^2 E^2$. Заменяя теперь в (30) левую часть неравенства его нижней оценкой, а правую часть неравенства его верхней оценкой, огрубляем неравенство и приходим к (26). Разумеется, реальная область, в которой смещенная оценка $\hat{X}_k^+$ предпочтительнее оценки $\hat{X}^+$ (возможно, также смещенной) существенно шире, нежели это следует из условия (26), поскольку, как явствует из доказательства, это неравенство достаточно грубое. Тем не менее оно может использоваться для ориентировочного предварительного выбора ϵ_k. Разбиение (21) достигается в процессе двучленной ортогонализации Грама-Шмидта [5, с.36; , с.274-276], при этом $q \le m - k$.

6. Сглаживание и прогнозирование. Под сглаживанием будем понимать оценивание $M(Y) = AX$, а под прогнозированием - линейной формы $a^T(\tau)\,X$, допускающей оценивание [8]. Результат будет зависеть от того, строится ли оценка на базе исходной системы (1) или нормального уравнения (13).

УТВЕРЖДЕНИЕ 9. Пусть $rkA=r$; $r \ge k \ge 1$ и $A = A_k + A_1$, где A_k и A_1 удовлетворяют (22), (23), тогда оценки

$$\hat{Y}_k^+ = A \cdot A_k^+ Y = A_k A_k^+ Y, \tag{31}$$

$$\hat{Y}_k^- = A\, A_k^- Y = A_k A_k^- Y \tag{32}$$

тождественны:

$$\hat{Y}_k^+ = \hat{Y}_k^- = \hat{Y}_k = A_{H_k}\, A_{H_k}^T\, Y, \tag{33}$$

где $A_{H_k} A_{H_k} = \dot{I}_k$ и $\mathcal{L}(A_k) = \mathcal{L}(A_{H_k})$.

При $k=r$ в классе линейных несмещенных оценок они обладают минимальными дисперсиями в совокупности. При $k<r$ оценки смещены

$$M(AX - \hat{Y}_k) = A_1 X, \tag{34}$$

$$\operatorname{cov}(\hat{Y}_k) = \sigma^2 A_{H_k} \cdot A_{H_k}^T. \tag{35}$$

Если $\|X\|_2 \le \sigma E$ и $Q_k = \frac{1}{\sigma_2} M[(\hat{Y}_k - AX)^T(\hat{Y}_k - AX)]$, то

$$\min_{r \ge k \ge 1}\ \max_{\|X\| \le \sigma E} Q_k = \min_{r \ge k \ge 1} (k + \lambda_1 E^2),$$

где λ_1 - максимальное собственное значение матрицы Грама $\Gamma_1 = A_1^T A_1$. Смещенная оценка предпочтительнее несмещенной по критерию (34) по крайней мере при

$$\varepsilon_1^2 < \left(\frac{r - k}{m - k}\right)\frac{1}{E^2}. \qquad (36)$$

ДОКАЗАТЕЛЬСТВО. Чтобы убедиться в тождественности (3I) и (32), обратимся вновь к сингулярному разложению (см. (27),(28), (29)). Очевидно, $A_1 \cdot \Lambda_k^- = 0$ и, следовательно, $A A_k^+ = (U\Lambda_k V^T + U\Lambda_1 V^T) V\Lambda_k^- U^T = A_k A_k^+$. Воспользуемся теперь иным представлением A (см. (6))

$$A = A_k + A_1 = (A_{H_k}, A_{H_1})\begin{pmatrix} C_k \\ C_1 \end{pmatrix},$$

$$A_k = A_{H_k} C_k; \quad A_k^+ = C_k^+ A_{H_k}^T.$$

Следовательно, $A_k A_k^+ = A_{H_k} A_{H_k}^T$, поскольку C_k – матрица полного ранга и $C_k C_k^+ = I_k$. Далее, согласно (10),

$$A_H = (A_{H_k}, A_{H_1}) = (A_k, A_1)\begin{pmatrix} \theta_k, & 0 \\ 0, & \theta_1 \end{pmatrix},$$

$$A_{H_k} = A_k \cdot \theta_k$$

и в соответствии с (II) $A_k^- = \theta_k A_{H_k}^T$. С учетом (12) получаем, наконец,

$$A A_k^- = (A_{H_k} \; A_{H_1})\begin{pmatrix} C_k \\ C_1 \end{pmatrix}\cdot\theta_k A_{H_k}^T = A_{H_k} A_{H_k}^T,$$

т.е. доказательство тождественности оценок $\hat{Y}_k^+$ и $\hat{Y}_k^-$. Выражения для смещения (34) и ковариационной матрицы (35) следуют из (I), (2I) и (33). Фактически нет необходимости искать A_k^+ или A_k^-, а достаточно факторизовать A согласно (6) и найти A_{H_k}. Интересно, что при $k = r$ оценки $\hat{Y}_r$ получаются несмещенными несмотря на то, что оценки вектора параметров $\hat{X}^+$ могут быть смещены.

Пусть, далее, некоторый оператор B обеспечивает также получение линейных несмещенных оценок $\hat{Y}_B$, т.е. $M(\hat{Y}_B) = M(ABY) = ABAX = AX$. Поскольку это должно выполняться $\forall X$, отсюда следует

$$A B A = A, \tag{37}$$

т.е. B должна быть одной из полуобратных матриц, перечисленных в п.3. Покажем далее, что $\operatorname{cov}(\hat{Y}_B) - \operatorname{cov}(\hat{Y})$ представляет неотрицательно определенную матрицу, т.е. $\forall Z \neq 0$ имеет место

$$Z^T A B B^T A^T Z \geq Z^T A_H A_H^T Z . \tag{38}$$

Используем для A в (37) факторизацию (6) $A_H C B A_H C = A_H C$. Выделим далее из строк A_H квадратную невырожденную подматрицу A_H^*. При этом $A_H^* C B A_H C = A_H^* C$, откуда следует $C B A_H C = C$. Аналогичную операцию проделаем со столбцами C, сформировав из них квадратную невырожденную матрицу C^*, $C B A_H C^* = C^*$, откуда $C B A_H = I_r$, что позволяет представить I_r в виде $\dot{I}_r = C B A_H A_H^T B^T C^T$. Возвращаясь к (38) и учитывая свойства оператора проектирования $P_{\mathcal{N}(A^T)} = I_n - A_H A_H^T = P_{\mathcal{N}(A^T)} P_{\mathcal{N}(A^T)}$, получаем

$$Z^T A_H C B B^T C^T A_H^T Z - Z^T A_H C B A_H A_H^T B^T C^T A_H^T Z = W^T W \geq 0,$$

где $W = P_{\mathcal{N}(A^T)} B^T C^T A_H^T Z$, что и завершает доказательство (37). В частности, $\operatorname{tr} \operatorname{cov} (\hat{Y}) \leq \operatorname{tr} \operatorname{cov} (\hat{Y}_B)$. Отметим, что использование любой из полуобратных матриц, приведенных в п.3, кроме A^+ и A^-, приводит к несмещенным оценкам $\hat{Y}_G$, точность которых не превосходит точности $\hat{Y}^+$ или $\hat{Y}^-$. Удостоверимся в этом относительно G, удовлетворяющей (2), (3) и (5). Примем для простоты, что $A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$, где A_1 - матрица полного ранга, $rk\,A_1 = rk\,A$. Легко убедиться, что G можно представить в виде $G = (A_1^+ \ \ 0_{m,n-r})$. Ковариационная матрица оценки $\hat{Y}_G = AGY$ имеет вид

$$\operatorname{cov} (\hat{Y}_G) = \sigma^2 \begin{pmatrix} \dot{I}_r & D^T \\ D & D D^T \end{pmatrix},$$

где $D = A_2 A_1^+$. И, очевидно $\operatorname{tr} \operatorname{cov} (\hat{Y}_G) = \sigma^2(r + \operatorname{tr} D D^T) \geq \operatorname{tr} (\hat{Y}^+) = \sigma^2 r$. Точность оценок $\hat{Y}_G$, построенных с помощью полуобратных матриц, удовлетворяющих еще меньшему числу ограничений, будет тем более не выше. Однако если строить оценки сглаживания, опираясь на нормальную систему уравнений, т.е. в виде

$$\hat{Y}_G = A\, G\, A^T\, Y\,, \tag{39}$$

где G - полуобратная матрица к $A^T A$, то все типы полуобратных матриц приведут к тождественным результатам [8]. Следовательно,если обусловленность матрицы невелика, может оказаться целесообразнее использовать оценки вида (39), прибегая в случае необходимости к вычислениям с повышенной точностью.

В заключение рассмотрим средний нормированный квадрат отклонений оценок сглаживания от истинных значений

$$Q_k = \frac{1}{\sigma^2} M\,[(\hat{Y}_k - M(Y))^T (\hat{Y}_k - M(Y)] = \frac{1}{\sigma^2} X^T A_1 A_1^T X + k.$$

Известно (см. например, [8]), что

$$\max_{\|X\|_2 \le \sigma E} X^T \Gamma_1 X = \sigma^2 E^2 \lambda_1\,,$$

где λ_1 - максимальное собственное значение матрицы Грама Γ_1. Выбор $\min\limits_{r \le k \le 1} \max\limits_{\|X\|_2 \le \sigma E} Q_k$ гарантирует, что точность смещенной оценки в наихудшем случае будет не ниже точности $\hat{Y}_k$. Минимум может достигаться на границе при $k = 1$ или $k = r$. Можно получить более грубую оценку, потребовав, чтобы уменьшение случайной компоненты Q_k, равное $r - k$, превышало систематическую компоненту Q_k, т.е. $\frac{1}{\sigma^2} X^T \Gamma_1 X \le E^2 \lambda_1 \le E^2 \mathrm{tr}\, \Gamma_1 \le E^2 (m-k) \epsilon_1^2$. Смещенная оценка будет точнее при $r-k > E^2\, \epsilon_1^2 (m - k)$, откуда и следует (36). Разумеется, реально смещенная оценка будет точнее несмещенной и при большем значении ϵ_1. Это завершает доказательство утверждения 9.

В случае прогнозирования мы ограничимся лишь формулировкой утверждений, доказательство которых может быть проведено аналогично предыдущему. Напомним, что под прогнозированием здесь понимается оценивание линейной формы $a^T(\tau)X$ в том случае, когда для нее существует несмещенная линейная оценка [8]. Это будет иметь место при $a(\tau) \in \mathcal{X}(A^T) = \mathcal{X}(A^T A) = \mathcal{X}(C^T)$. (Здесь C удовлетворяет (6).)

УТВЕРЖДЕНИЕ 10. Пусть $a(\tau) = A^T A \cdot d(\tau)$, тогда оценка $\hat{y}(\tau) = d^T(\tau) A^T Y$ является несмещенной оценкой с минимальной диспер -

сией в классе всех линейных несмещенных оценок для $a^T(\tau) X$, причем [8]

$$D(\hat{y}(\tau)) = \sigma^2 a^T(\tau) \cdot d(\tau) .$$

Если использование $A^T A$ нежелательно в силу возможной потери точности, можно опереться на представление $a(\tau) = C^T b(\tau)$; $b(\tau) \in R_r$.

УТВЕРЖДЕНИЕ II. Пусть

$$a(\tau) = C^T b(\tau) = (C_k, C_1)\begin{pmatrix} b_k(\tau) \\ b_1(\tau) \end{pmatrix}.$$

Тогда оценки $\hat{y}^+_k(\tau) = a^T(\tau) A^+_k Y$ и $\hat{y}^-_k(\tau) = a^T(\tau) A^-_k Y$ тождественны:

$$\hat{y}^+_k(\tau) = \hat{y}^-_k(\tau) = b^T_k(\tau) A^T_{H_k} Y , \qquad (40)$$

они смещены $M(a^T(\tau)X - \hat{y}_k(\tau)) = b^T_1(\tau)C_1 X$ и характеризуются дисперсией $D(\hat{y}_k(\tau)) = \sigma^2 b^T_k(\tau) b_k(\tau)$ и средним нормированным квадратом отклонения

$$Q_k = b^T_k(\tau)\, b_k(\tau) + \frac{1}{\sigma^2}(b^T_1(\tau)\, C_1 X)^2 .$$

При $\|X\|_2 \le \sigma E$

$$\min_{r \ge k \ge 1} \max_{\|X\| \le \sigma E} Q_k = \min_{r \ge k \ge 1} [b^T_k(\tau)\, b_k(\tau) + E^2 \nu_1^2] ,$$

где $\nu_1^2 = b^T_1(\tau)\, C_1\, C^T_1 b_1(\tau)$.

Смещенная оценка (40) будет предпочтительнее несмещенной (при $k = r$) по крайней мере при

$$\varepsilon_1^2 \le \frac{1}{(m-k)^2 E^2} . \qquad (41)$$

ДОКАЗАТЕЛЬСТВО. Покажем справедливость (41). Для несмещенной оценки при $k=r$ $Q_r = b^T_r(\tau) \cdot b_r(\tau)$. Потребуем, чтобы

$$Q_r - Q_k = b^T_1(\tau) \cdot b_1(\tau) - \frac{1}{\sigma_2}(b^T_1(\tau)\, C_1\, X)^2 \ge 0.$$

Принимая во внимание (23) и полагая, что C_1 получены в процессе двучленной ортогонализации Грама-Шмидта для A, имеем

$$\frac{1}{\sigma^2}(b_1^T(\tau)\ C_1X)^2 \le E^2 b_1^T(\tau)\ C_1\ C_1^T\ b_1(\tau) \le$$

$$\le E^2 b_1^T(\tau)\ b_1(\tau)\cdot \mathrm{tr}\ C_1\ C_1^T \le E^2 b_1(\tau)\ b_1(\tau)\cdot(m-k)\varepsilon_1^2\ ,$$

откуда и следует (41). Как видно из вывода (41), реально смещенные оценки будут точнее при значительно большем значении ε_1.

7. Резюмируем основные выводы.

Существует класс задач, которому принципиально присуща вырожденность. Поскольку реально приходится иметь дело с возмущенной системой, ранг ее чаще всего может оказаться завышенным. Если он заранее не известен, то повышение точности вычислений может привести к обратному эффекту - понижению результирующей точности.

Чтобы не ухудшать обусловленность, предпочтительно не переходить к нормальной системе уравнений. При этом, однако, в задачах сглаживания и прогнозирования уже не все полуобратные матрицы приводят к эквивалентным оценкам. Лучшую точность обычно обеспечивает использование A^+ и A^-, вычислять которые при этом также нет надобности. Достаточно осуществить факторизацию $A = A_H \cdot C$ и использовать затем ортонормальную A_H и C. При оценивании параметров преимущество уже на стороне только A^+.

Если ранг матрицы не известен заранее, его определение может доставить некоторые трудности. Способ его вычисления будет зависеть от используемого метода полуобращения. Алгоритм, основанный на сингулярном разложении, был рассмотрен в [4]. В данной работе предполагается использование при полуобращении двучленной ортогонализации Грама-Шмидта. При вычислении ранга приходится вводить некоторый порог точности. В этом методе он сопоставляется с квадратами норм последовательно ортогонализуемых столбцов модифицируемой матрицы A. Спектр этих норм, подобно сингулярному спектру матрицы A, содержит отличное от нуля число компонент, равное рангу. Под действием ошибок возмущения этот спектр может расширяться. Если компоненты ранжированного по убыванию спектра плавно приближаются к нулю, величина ранга теряет свою определенность. Ранг становится относительным, зависящим от выбранного порога. Повышение этого порога улучшает обусловленность отвечающей ему матрицы и этим стабилизирует решение. Это сопровождается уменьшением случайной компоненты погрешности и увеличением смещения (если относительный ранг меньше истинного).

Итак, в принципе возможны три ситуации:

1. Исходная матрица полного ранга и хорошо обусловлена. Возмущение не меняет ранга и выбор порога в некотором достаточно широком диапазоне приводит к тому же рангу. Получаются обычные оценки метода наименьших квадратов.

2. Исходная матрица хорошо обусловлена, но имеет неполный ранг, возмущенная матрица имеет больший ранг. Привлекая достаточно грубую априорную информацию о модуле решения и в случае необходимости повышая точность вычислений, можно получить относительный ранг, совпадающий с истинным.

3. Исходная матрица плохо обусловлена. Она может быть как полного, так и неполного ранга. Случай уже описан выше. Варьирование порогом точности приводит к перераспределению между величиной смещения и случайной компонентой погрешности. Опираясь на дополнительную априорную информацию о норме решения, можно достичь определенного оптимума. Разумеется, и в первых двух случаях, завышая порог, можно перейти к смещенным оценкам, однако это может оказаться нецелесообразным.

В отличие от методов, использующих регуляризацию Тихонова или ридж регрессию, в которых привлечение априорной информации ограничивает область поиска решения (обычно сферой или эллипсоидом) не меняя размерности пространства параметров, при данном подходе редуцируется размерность пространства параметров, область же поиска внутри него не ограничивается.

По-видимому, данный подход может оказаться предпочтительнее при наличии достаточно грубой априорной информации в первом и особенно во втором случаях, а также иногда в третьем, когда исходная матрица будет неполного ранга. Если в методах регуляризации Тихонова любое привлечение априорной информации уже обеспечивает повышение точности смещенных оценок, здесь этого может и не произойти, если не осуществится редукция исходного ранга, как в случаях 1 и 2. Количественное сопоставление точности различных методов выходит за рамки данной работы.

Привлекательность рассматриваемого подхода в простоте его реализации и в обеспечении устойчивости решения при использовании достаточно грубой априорной информации о модуле решения.

Литература

1. ТИХОНОВ А.Н., АРСЕНИН В.Я. Методы решения некорректных задач. - М.: Наука, 1979. - 285 с.

2. DRAPER N.R., Van NORSTRAND R.C. Ridge regression and James-Stein estimation: review and comments.- Technometrics,1979, v.21, N 4,p.451-466.

3. SMITH G., CAMPBELL F. and other. A critique of some ridge regression methods.(Discus.)-JASA,1980,v.75,N 369,p.74-103.

4. MARQUARDT D.W. Generalized inverses,ridge regression,biased linear estimation and nonlinear estimation.- Technometrics, 1970,v.12, N 3,p.591-612.

5. ФАДДЕЕВ Д.К., ФАДДЕЕВА В.Н. Вычислительные методы линейной алгебры.-Л., 1975. - 228 с. (Зап. науч. семин. ЛОМИ АН СССР, т. 54).

6. Вычислительные методы линейной алгебры. Библ. указатель 1828-1974. Под ред. В.В.Воеводина. - Новосибирск, 1976. - 418 с.

7. RAO C.R., MITRA S.K. Generalized inverse of matrices.New York: Wiley and Sons,1971.-240 p.

8. РАО С.Р. Линейные статистические методы и их применения. -М.: Наука, 1968. - 548 с.

9. Generalized inverses and applications.- In Proc.of an advanced seminar sponsored by the mathematics research center the university of Wisconsin-Madison, October 8-10, 1973. Ed.by M.Zuhair Nashed.- New York, Academic Press,1976.

10. ВОЕВОДИН В.В. Вычислительные основы линейной алгебры. -М.: Наука, 1977. - 303 с.

11. ФОРСАЙТ Дж., МАЛЬКОЛЬМ М., МОУЛЕР К. Машинные методы математических вычислений. - М.: Мир, 1980. - 279 с.

12. МЕЛЕШКО В.И. Устойчивое к возмущениям псевдообращение замкнутых операторов. - Журн. вычислит. математики и мат. физики, 1977, т. 17, № 5, с. 1132-1143.

Поступила в ред.-изд.отд.
29 сентября 1981 года