

УДК 681.5

Б.А.Скороход, профессор, д-р техн. наук*Севастопольский национальный технический университет**ул. Университетская 33, г. Севастополь, Украина, 99053**E-mail: b.skorohod@mail.ru***ИНИЦИАЛИЗАЦИЯ ЛИНЕЙНОЙ РЕКУРРЕНТНОЙ РЕГРЕССИИ**

Рассматривается поведение рекуррентного метода наименьших квадратов при отсутствии априорной информации относительно оцениваемых неизвестных параметров.

Введение. Задачи оценки параметров линейной регрессии в темпе реального времени возникают в самых различных областях, включающих, в частности, идентификацию объектов управления, адаптивную обработку сигналов, адаптивное управление [1-3].

Предполагается, что уравнение регрессии имеет вид

$$y_t = c_t x + \xi_t, \quad t = 1, 2, \dots, \quad (1)$$

где $y_t \in R^1$ – измерения, $x \in R^n$ – вектор неизвестных параметров, подлежащий оцениванию по наблюдениям, $c_t^T \in R^n$ – вектор регрессоров, ξ_t – последовательность некоррелированных случайных величин, $E(\xi_t) = 0$, $\text{var}(\xi_t) = \sigma_t^2$, $\sigma_t^2 > 0$. Здесь, R^1 и R^n – пространства действительных чисел и n -мерных векторов, соответственно, T – операция транспонирования, $E(\cdot)$, $\text{var}(\cdot)$ – математическое ожидание и дисперсия случайной величины, соответственно.

Для оценки неизвестного векторного параметра x используется стандартный рекуррентный алгоритм вида [2]

$$\hat{x}_t = \hat{x}_{t-1} + K_t (y_t - c_t \hat{x}_{t-1}), \quad (2)$$

где $\hat{x}_t$ – оценка x по измерениям доступным к моменту t , K_t – векторный коэффициент усиления. Для того чтобы воспользоваться этим алгоритмом мы должны выбрать K_t и задать начальное условие $\hat{x}_0 = x_0$. Один из общепринятых подходов к выбору K_t состоит в использовании соотношений [2]

$$K_t = P_{t-1} c_t^T (c_t P_{t-1} c_t^T + \sigma_t^2)^{-1}, \quad (3)$$

$$P_t = P_{t-1} - P_{t-1} c_t^T c_t P_{t-1} (c_t P_{t-1} c_t^T + \sigma_t^2)^{-1}. \quad (4)$$

Имеются, по крайней мере, два существенных преимущества алгоритма (2)–(4) по сравнению со стандартным нерекуррентным методом наименьших квадратов (МНК). Во-первых, не требуется обращение матрицы. Во-вторых, количество операций и объем памяти фиксирован в каждой итерации, а не растет по мере поступления новых наблюдений. Вместе с тем, для инициализации алгоритма, необходимо задать начальные условия x_0, P_0 для вектора $\hat{x}_t$ и матрицы P_t , причем, типичной является ситуация, когда априорная информация относительно x отсутствует. Стандартной является следующая, так называемая, диффузная инициализация алгоритма: $\hat{x}_0 = 0, P_0 = \lambda^2 I$, позволяющая учесть неопределенность в задании $\hat{x}_0$, где I – единичная матрица, $\lambda > 0$ – большой параметр. Однако открытым остается вопрос теоретического обоснования возможности его использования. В работе [4] получены асимптотические представления для K_t, P_t и показано на численном примере, что алгоритм сходится за конечное число шагов к оценке, даваемой нерекуррентным МНК. Эти представления справедливы в предположении, что $\sigma_t = \text{const}$ и $c_t P_{t-1} c_t^T \neq 0$, где P_{t-1}^0 – диффузная составляющая в разложении $P_t = P_t^* + \lambda^2 P_t^0$. Целью данной статьи является получение асимптотического разложения для K_t без использования каких-либо дополнительных предположений и доказательстве того, что при $\sigma_t = \text{const}$ и $\lambda \rightarrow \infty$ оценка $\hat{x}_t$, определяемая из предельного алгоритма, совпадает с оценкой нерекуррентного МНК.

2. Результаты. Покажем вначале, что

$$K_t = M_t^{-1} c_t^T / \sigma_t^2, \quad (5)$$

где

$$M_t = \begin{cases} P_0^{-1}, & t = 0 \\ P_0^{-1} + \sum_{s=0}^t c_s^T c_s / \sigma_s^2, & t = 1, 2, \dots \end{cases} \quad (6)$$

Из (6) следует, что $M_t > 0$ и

$$M_t = M_{t-1} + c_t^T c_t / \sigma_t^2, M_0 = P_0^{-1}, t = 1, 2, \dots \quad (7)$$

Найдем рекуррентное выражение для M_t^{-1} , используя матричное тождество

$$(S + H^T R^{-1} H)^{-1} = S - S H^T (H S H^T + R)^{-1} H S. \quad (8)$$

Полагая в (8) $S = M_{t-1}, H = c_t, R = \sigma_t^2$, находим

$$M_t^{-1} = (M_{t-1} + M_{t-1} + c_t^T c_t / \sigma_t^2)^{-1} = M_{t-1}^{-1} - M_{t-1}^{-1} c_t^T c_t M_{t-1}^{-1} (\sigma_t^2 + c_t^T M_{t-1}^{-1} c_t)^{-1}. \quad (9)$$

Сравнивая (4) и (9), устанавливаем, что $P_t = M_t^{-1}$. Умножая слева (3) на множитель $P_t P_t^{-1}$ и используя (9), находим

$$\begin{aligned} K_t &= P_t P_t^{-1} P_{t-1} c_t^T (c_t P_{t-1} c_t^T + \sigma_t^2)^{-1} = P_t (P_t^{-1} + c_t^T c_t / \sigma_t^2) P_{t-1} c_t^T (c_t P_{t-1} c_t^T + \sigma_t^2)^{-1} = \\ &= P_t (c_t^T + c_t^T c_t P_{t-1} c_t^T / \sigma_t^2) (c_t P_{t-1} c_t^T + \sigma_t^2)^{-1} = M_t^{-1} c_t^T / \sigma_t^2. \end{aligned}$$

Перейдем непосредственно к анализу асимптотического поведения K_t . Имеем

$$M_t = P_0^{-1} + \sum_{s=1}^t c_s^T c_s / \sigma_s^2 = (I + \lambda W_t) / \lambda, t = 1, 2, \dots \quad (11)$$

Откуда, с помощью формулы обращения возмущенных матриц [5]

$$(I + \lambda V)^{-1} = (I - V V^+) + V^+ (V^+ + \lambda I)^{-1}, \quad (12)$$

где V – произвольная симметрическая матрица, находим

$$M_t = (I + \lambda W_t)^{-1} \lambda = ((I - W_t W_t^+) + W_t^+ (W_t^+ + \lambda I)^{-1}) \lambda. \quad (13)$$

Представим матрицу W_t в виде $W_t = T_t^T D_t T_t$, где T_t – ортогональная, D_t – диагональная матрицы. Тогда

$$(W_t^+ + \lambda I)^{-1} \lambda = T_t (D_t^+ + \lambda I)^{-1} T_t^T \lambda = I + \sum_{i=1}^k (-1)^i (W_t^+)^i \lambda^{-i} + O(\lambda^{-k-1}), \lambda \rightarrow \infty \quad (14)$$

равномерно по $t \in T$. Подстановка (14) в (13), дает равномерное по $t \in T$ на любом ограниченном множестве $T = \{0, 1, \dots, N\}$ асимптотическое разложение

$$M_t = (I - W_t W_t^+) \lambda + W_t^+ + O(\lambda^{-1}), \quad \lambda \rightarrow \infty, \quad (15)$$

Из (3), (4) следует, что $K_t = 0$, если $c_t = 0$. Пусть $c_t \neq 0$ и $\tilde{C}_t = (\tilde{c}_1^T, \tilde{c}_2^T, \dots, \tilde{c}_t^T)^T$, где $\tilde{c}_i = c_i / \sigma_i$, $i = 1, 2, \dots, t$. Обозначим $l_{1t}, \dots, l_{k(t),t}$ любые линейно независимые столбцы матрицы $\tilde{C}_t^T$ такие, что $l_{k(t),t} = \tilde{c}_t^T$. Использование скелетного разложения для $\tilde{C}_t^T$ дает $\tilde{C}_t^T = L_t \Gamma_t$, где $L_t = (l_{1t}, \dots, l_{k(t),t}) \in R_{k(t)}^{n \times k(t)}$, $\Gamma_t \in R_{k(t)}^{k(t) \times t}$, где $R_r^{m \times n}$ – пространство $m \times n$ матриц ранга r .

Покажем, что

$$I - W_t W_t^+ = I - L_t (L_t^T L_t)^{-1} L_t^T.$$

Имеем

$$W_t = \tilde{C}_t^T \tilde{C}_t = L_t \tilde{\Gamma}_t \Gamma_t^T, \quad (16)$$

где $\tilde{\Gamma}_t = \Gamma_t \Gamma_t^T$. Так как $\tilde{\Gamma}_t > 0$ – как матрица Грамма, построенная по линейно независимым строкам матрицы $\tilde{\Gamma}_t$ и $\text{rank}(L_t) = \text{rank}(\tilde{\Gamma}_t \Gamma_t^T)$, то

$$W_t^+ = (L_t \tilde{\Gamma}_t \Gamma_t^T)^+ = (L_t^T)^+ \tilde{\Gamma}_t^{-1} L_t^+. \quad (17)$$

Далее, так как столбцы матрицы L_t линейно независимы, то [5]

$$L_t^+ = (L_t^T L_t)^{-1} L_t^T. \quad (18)$$

Используя (16)–(18), находим

$$\begin{aligned} I - W_t W_t^+ &= I - L_t \tilde{\Gamma}_t \Gamma_t^T (L_t \tilde{\Gamma}_t \Gamma_t^T)^+ = I - L_t \tilde{\Gamma}_t \Gamma_t^T (L_t^T)^+ \tilde{\Gamma}_t^{-1} L_t^+ \\ &= I - L_t \tilde{\Gamma}_t \Gamma_t^T L_t (L_t^T L_t)^{-1} \tilde{\Gamma}_t^{-1} (L_t^T L_t)^{-1} L_t^T = I - L_t (L_t^T L_t)^{-1} L_t^T. \end{aligned}$$

Откуда следует, что

$$(I - W_t W_t^+) c_t^T = (I - L_t (L_t^T L_t)^{-1} L_t^T) L_t e_n = 0,$$

где $e_k = (0, 0, \dots, 1)^T \in R^k$.

Таким образом, равномерно по $t \in T$

$$K_t = M_t^{-1} c_t^T / \sigma_t^2 = (I + \lambda W_t)^{-1} c_t^T / \sigma_t^2 \lambda = W_t^+ c_t^T / \sigma_t^2 + O(\lambda^{-1}) = K_t^\infty + O(\lambda^{-1}), \quad \lambda \rightarrow \infty. \quad (19)$$

Рекуррентный алгоритм (2) при $K_t = K_t^\infty$ будем называть предельным алгоритмом оценки параметров линейной регрессии, а K_t^∞ – предельным коэффициентом усиления.

Покажем, как может быть рекуррентно вычислен коэффициент усиления K_t^∞ . Воспользуемся формулой Гревилля [5]. Положим $Q_t = \tilde{C}_t^T$ и воспользуемся формулой Гревилля [5]

$$Q_t^+ = \begin{pmatrix} Q_{t-1}^+ (I - \tilde{C}_t^T k_t^T) \\ k_t^T \end{pmatrix}, \quad (20)$$

где

$$k_t = \begin{cases} \frac{(I - Q_{t-1}^+ Q_{t-1}^+) \tilde{C}_t^T}{\| (I - Q_{t-1}^+ Q_{t-1}^+) \tilde{C}_t^T \|^2}, & (I - Q_{t-1}^+ Q_{t-1}^+) \tilde{C}_t^T \neq 0 \\ \frac{(Q_{t-1}^+)^T Q_{t-1}^+ \tilde{C}_t^T}{1 + \| Q_{t-1}^+ \tilde{C}_t^T \|^2}, & (I - Q_{t-1}^+ Q_{t-1}^+) \tilde{C}_t^T = 0 \end{cases}. \quad (21)$$

Из (20) следует, что $k_t^T = e_n^T Q_t^+$. Но так как [5] $Q_t^+ = Q_t^T (Q_t Q_t^T)^+ W_t = Q_t Q_t^T$, то $k_t^T = e_n^T Q_t^T (Q_t Q_t^T)^+ = c_t W_t^+ / \sigma_t$. Таким образом.

$$K_t^\infty = k_t / \sigma_t. \quad (22)$$

Вводя обозначения $V_t = \sigma_t^{-1} S_t$, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_t)$, где $S_t = (c_1^T c_2^T \dots c_t^T)$, представим (21) в следующем эквивалентном виде

$$k_t = \begin{cases} \frac{(I - S_{t-1}^+ S_{t-1}^+) c_t^T \sigma_t}{\| (I - S_{t-1}^+ S_{t-1}^+) c_t^T \|^2}, & (I - S_{t-1}^+ S_{t-1}^+) c_t^T \neq 0 \\ \frac{(S_{t-1}^+)^T \Sigma_t^2 S_{t-1}^+ c_t^T / \sigma_t}{1 + \| S_{t-1}^+ \Sigma_t c_t^T / \sigma_t \|^2}, & (I - S_{t-1}^+ S_{t-1}^+) c_t^T = 0 \end{cases}. \quad (23)$$

Полагая в (22) $\sigma_t = \sigma$, устанавливаем, что предельный коэффициент усиления K_t^∞ совпадает с коэффициентом усиления рекуррентного алгоритма полученного в [5], оценки которого, в свою очередь, совпадают с оценками МНК.

Таким образом, доказано следующее утверждение.

Теорема. 1. Для векторного коэффициента усиления рекуррентной процедуры (2)-(4) оценки параметров линейной регрессии при $c_t \neq 0$ справедливо равномерное на любом ограниченном множестве асимптотическое представление (19).

2. Предельный коэффициент усиления может быть найден рекуррентно с помощью соотношений (21)–(23).

3. При постоянной интенсивности шума в измерениях оценка, полученная с помощью предельного алгоритма, совпадают с оценкой МНК.

Заключение. Полученные результаты могут рассматриваться, как теоретическое обоснование использования диффузных начальных условий в рекуррентной процедуре оценки параметров линейной регрессии. Дальнейшие исследования будут направлены на изучение свойств предельного алгоритма оценивания.

Библиографический список

1. Haykin S. Adaptive Filter Theory / S. Haykin. — Prentice-Hall, Upper Saddle River, NJ, 1996. — 978 p.
2. Ljung L. System Identification Theory for the User / L. Ljung. — Prentice Hall PTR, 1999. — 410 p.
3. Tao Gang Adaptive control and analysis / Gang Tao. — John Wiley & Sons, 2003. — 637 p.
4. Pollock D.S.G. Recursive estimation in econometrics / D.S.G. Pollock // Computational statistic. — 2003. — Vol. 44. — Issue 1–2. — P. 37–75.
5. Алберт А. Регрессия, псевдоинверсия и рекуррентное оценивание / А. Алберт. — М.: Наука, 1977. — 221 с.

Поступила в редакцию 4.07.2008 г.