Из этого выражения следует, что матрица $(Z^{\text{т}}Z)$ при $K < 2N + 1$ всегда вырождена. Поэтому мы должны потребовать, чтобы число измерений K было по крайней мере равно числу искомых параметров. Для того чтобы пользоваться рекуррентными формулами (13.81) – (13.83) с самого начала вычислений, когда $K < 2N + 1$, надо выбрать матрицу $Z(K_0)$, ранг которой равен числу искомых параметров. Тогда матрица $Z^{\text{т}}(K_0)Z(K_0)$ будет невырожденной, и мы можем определить первую оценку

$$P(K_0) = \alpha[Z^{\text{т}}(K_0)Z(K_0)]^{-1}, \tag{13.85}$$

$$\hat{\vartheta}(K_0) = [Z^{\text{т}}(K_0)Z(K_0)]^{-1}Z^{\text{т}}(K_0)y(K_0). \tag{13.86}$$

Можно также проводить вычисления по рекуррентным формулам (13.81) – (13.83) с первого шага $k = 1$, если выбрать некоторую невырожденную матрицу $P^{-1}(0)$ и вычислить $P(k)$ согласно выражению (13.78):

$$P^{-1}(k) = P^{-1}(0) + \frac{1}{\alpha} Z^{\text{т}}(k)Z(k) \tag{13.87}$$

при $k = 1, 2, \ldots$ Мы можем, конечно, положить

$$P(0) = \frac{1}{\epsilon} \text{E},$$

где $\epsilon > 0$ — достаточно малое число такое, что при $k > K$ в зависимости от выбора величины ϵ матрица $P^{-1}(k)$ близка к матрице $[Z^{\text{т}}(k)Z(k)]/\alpha$.

Различные модификации метода наименьших квадратов, приспособленные для оценивания параметров управляемого объекта, можно найти в обзорной статье В. Стрейца [77.28].

13.2.6. Свойства оценок метода наименьших квадратов. В п. 13.2.2 мы привели решение следующей задачи.

Пусть уравнение (13.57) описывает линейную модель объекта, который характеризуется матрицей параметров ϑ, и пусть $\{u(i - N),\ y(i - N),\ i = 1, 2, \ldots, K\}$ — измеренные значения входных и выходных переменных. Найти матрицу параметров ϑ такую, что функция потерь минимальна.

Мы приведем теперь некоторые важные свойства оценок метода наименьших квадратов. Для этой цели будем пользоваться математической моделью объекта в виде

$$y = Z\vartheta + e, \tag{13.88}$$

где ошибки $e(i - N)$, $i = 1, 2, \ldots, K$, — случайные переменные. Матрицу истинных параметров будем обозначать через ϑ, а матрицу оценок — через $\hat{\vartheta}$.

Мы теперь можем сформулировать следующую теорему.

Т е о р е м а 13.1. *Оценка матрицы параметров ϑ по методу наименьших квадратов несмещена, если среднее значение компонент вектора e равно нулю и если Z и e взаимно независимы.*

Доказательство. Оценка $\hat{\vartheta}$ согласно выражению (13.59) равна

$$\hat{\vartheta} = (Z^{\text{т}}Z)^{-1}Z^{\text{т}}y.$$

Подставляя y из уравнения (13.88), мы получим

$$\vartheta = (Z^{\text{т}}Z)^{-1}Z^{\text{т}}(Z\vartheta + e) = \vartheta + (Z^{\text{т}}Z)^{-1}Z^{\text{т}}e. \tag{13.89}$$

Если Z и e взаимно независимы, то имеет место

$$\mathcal{E}\hat{\vartheta} = \vartheta + \mathcal{E}[(Z^{\text{т}}Z)^{-1}Z^{\text{т}}e] = \vartheta + (Z^{\text{т}}Z)^{-1}Z^{\text{т}}\mathcal{E}e = \vartheta.$$

Т е о р е м а 13.2. *Если Z и e взаимно независимы,* $\mathcal{E}e(i) = 0$, $\mathcal{E}e(i)e(j) = 0$, $i \neq j$, $\mathcal{E}e^2(i) = \sigma^2$ *при* $i, j = 1, 2, \ldots, K$, *то*

$$\mathcal{E}[(\hat{\vartheta} - \vartheta)(\hat{\vartheta} - \vartheta)^{\text{т}}] = \sigma^2(Z^{\text{т}}Z)^{-1}. \tag{13.90}$$

Доказательство. Согласно выражению (13.59) имеем

$$\mathcal{E}[(\hat{\vartheta} - \vartheta)(\hat{\vartheta} - \vartheta)^{\text{т}}] = \mathcal{E}[\hat{\vartheta}\hat{\vartheta}^{\text{т}} - \vartheta\vartheta^{\text{т}}] =$$

$$= \mathcal{E}[(Z^{\text{т}}Z)^{-1}Z^{\text{т}}yy^{\text{т}}Z(Z^{\text{т}}Z)^{-1} - (Z^{\text{т}}Z)^{-1}(Z^{\text{т}}Z)\vartheta\vartheta^{\text{т}}(Z^{\text{т}}Z)(Z^{\text{т}}Z)^{-1}] =$$

$$= (Z^{\text{т}}Z)^{-1}Z^{\text{т}}\mathcal{E}[(y - Z\vartheta)(y - Z\vartheta)^{\text{т}}]Z(Z^{\text{т}}Z)^{-1} =$$

$$= (Z^{\text{т}}Z)^{-1}Z^{\text{т}}\mathcal{E}[ee^{\text{т}}]Z(Z^{\text{т}}Z)^{-1} = \sigma^2(Z^{\text{т}}Z)^{-1}.$$

В этом доказательстве мы воспользовались тем фактом, что компоненты случайного вектора e взаимно некоррелированы и имеют одинаковые дисперсии, так что

$$\mathcal{E}[ee^{\mathrm{T}}] = \sigma^2 \mathrm{E}. \tag{13.91}$$

Для многомерного случая можно также показать, что

$$\mathcal{E}[(\hat{\Theta} - \Theta)_j (\hat{\Theta} - \Theta)_s^{\mathrm{T}}] = \sigma^2 (Z^{\mathrm{T}} Z)^{-1}, \tag{13.92}$$

где $(\hat{\Theta} - \Theta)_j$ и $(\hat{\Theta} - \Theta)_s$ — j-й и s-й столбцы матрицы $(\hat{\Theta} - \Theta)$, $j, s = 1, 2, \ldots, p$. Более сложный вариант приводится в п. 13.3.2.

Следствие 13.1. *Матрица P, определяемая выражением* (13.78) *в случае, когда $\mathcal{E}[ee^{\mathrm{T}}] = \sigma^2 \mathrm{E}$, является ковариационной матрицей ошибок оценок параметров.*

Доказательство следует непосредственно из сравнения выражений (13.78) и (13.90).

Замечание 13.1. Условия, приведенные в теоремах 13.1 и 13.2, ограничивают справедливость этих теорем разомкнутыми системами, описываемыми простой регрессионной моделью, которая следует из обобщенной модели (13.37). В этом случае матрица Z содержит только входные переменные, которые не зависят от выходных переменных. Однако свойства оценок по методу наименьших квадратов для замкнутых систем, которые описываются обобщенной регрессионной моделью (13.37), в которой $y(k-i)$ и $u(k-i)$, $i = 0, 1, \ldots, N$, — наблюдаемые значения зависимых случайных величин, отличаются от свойств оценок параметров разомкнутых систем.

Дурбин [60.2] обобщил эти свойства следующим образом. Пусть $M = \mathcal{E}[(1/K) Z^{\mathrm{T}} Z]$ и предположим, что M^{-1} сходится к конечной положительно определенной матрице V при $K \to \infty$. Предположим также, что $M^{-1}[(1/K) Z^{\mathrm{T}} Z]$ сходится в стохастическом смысле к единичной матрице. Тогда асимптотически $K^{1/2}(\hat{\vartheta} - \vartheta)$ имеет нулевое среднее и ковариационную матрицу $\sigma_n^2 V$.

Если дисперсия компонент σ^2 случайного вектора не известна, то мы можем оценить ее по формуле

$$\hat{\sigma}^2 = \frac{1}{K-N} (y - Z\hat{\vartheta})^{\mathrm{T}} (y - Z\hat{\vartheta}). \tag{13.93}$$

Доказательство этого утверждения можно найти в литературе [68.2].

Из теоремы 13.2 следует, что оценка ковариационной матрицы ошибок оценок $\sigma^2 (Z^{\mathrm{T}} Z)^{-1}$, основанная на несмещенной оценке параметров, позволяет судить о точности оценок параметров, вычисленных по методу наименьших квадратов.

§ 13.3. Метод максимального правдоподобия

Мы займемся теперь более общей задачей идентификации замкнутых систем, чем задача идентификации разомкнутых систем с помощью метода наименьших квадратов. Для этой цели мы воспользуемся методом максимального правдоподобия, основанного на функции правдоподобия, которая определяется как совместная плотность распределения

$$L(\Theta, R;\ y_{(-N+1)}^{(K)}, u_{(-N+1)}^{(K)}) = f(y_{(-N+1)}^{(K)}, u_{(-N+1)}^{(K)}; \Theta, R) =$$

$$= \prod_{k=-N+1}^{K} f(y(k) \mid y_{(-N+1)}^{(k-1)}, u_{(-N+1)}^{(k)}; \Theta, R) \prod_{k=-N+1}^{K} f(u(k) \mid y_{(-N+1)}^{(k-1)}, u_{(-N+1)}^{(k-1)}), \tag{13.94}$$

где $[\Theta, R] = [P_N, Q_N, \ldots, P_1, Q_1, P_0, R] = \Theta^*$ — матрицы параметров регрессионной модели (13.39), а R — ковариационная матрица (13.38). С помощью метода максимального правдоподобия определяются оценки $\hat{\Theta}$ параметров Θ такие, при которых функция правдоподобия L достигает максимального значения. Поскольку $\ln L$ — возрастающая функция, которая достигает максимального значения при том же значении $\hat{\Theta}^*$, при котором достигает максимального значения функция L, то оценку $\hat{\Theta}^*$ часто находят, минимизируя $\ln L$. Это происходит потому, что для дифференцируемой

по Θ^* функции правдо[...]

$$\frac{\partial}{\partial \Theta^*} \ln L = 0,$$

чем из условия

$$\frac{\partial L}{\partial \Theta^*} = 0.$$

Любое решение $\hat{\Theta}^*$, к[...] *максимального правд*[...] *доподобия.*

Важность функции [...] К.Ф. Гауссу уже в т[...] нако сам метод мак[...] когда Р.А. Фишер пре[...]

В сумме логариф[...] имеет смысл рассмат[...] раметры объекта. Это [...]

$$\prod_{k=-N+1}^{K} f(y(k) \mid y_{(-}^{(k}\ldots$$

все сомножители ко[...] модели (13.37). Есл[...] пределение и не завис[...]

$$f[e(k) \mid y_{(-N+1)}^{(k-1)}, u\ldots$$

$$= (2\pi)^{-p/2} |R^{-1}|^{1/}\ldots$$

где p — число выход[...]

Выражая теперь [...] (13.39), найдем фун[...]

$$L(\Theta, R) = \prod_{k=1}^{K} f[y(\ldots$$

$$= (2\pi)^{-pK/2} |R^{-1}|\ldots$$

Следует отметить, [...] ности при $k < 1$. [...] как начальное услов[...]

Регрессионная м[...] рица $z(k)$ и, следо[...] $= 1, 2, \ldots, K$, содер[...] та и, следовательно [...] $= 1, 2, \ldots$ Кроме [...] зависят от предыду[...]

13.3.1. Вычислен[...] т.е. определение па[...] максимум этой фу[...] мер, в [58.1]. Здес[...] которое, как мы п[...] и к идентификации [...]

Натуральный лог[...]

$$\ln L(\Theta, R) = -\frac{pK}{2}\ldots$$

по Θ^* функции правдоподобия легче вычислить оценку $\hat{\Theta}^*$ из условия

$$\frac{\partial}{\partial \Theta^*} \ln L = 0, \tag{13.95}$$

чем из условия

$$\frac{\partial L}{\partial \Theta^*} = 0. \tag{13.96}$$

Любое решение $\hat{\Theta}^*$, которое удовлетворяет уравнению (13.95), называется *оценкой максимального правдоподобия* Θ^*, а уравнение (13.95) называется *уравнением правдоподобия.*

Важность функции правдоподобия и применение ее логарифма были известны К.Ф. Гауссу уже в то время, когда он предложил метод наименьших квадратов. Однако сам метод максимального правдоподобия оставался неизвестным до 1912 г., когда Р.А. Фишер предложил его обобщенную и точную формулировку.

В сумме логарифмов совместной плотности вероятностей в уравнении (13.95) имеет смысл рассматривать только те слагаемые, которые содержат неизвестные параметры объекта. Этому условию удовлетворяет произведение

$$\prod_{k=-N+1}^{K} f(y(k) \mid y_{(-N+1)}^{(k-1)}, u_{(-N+1)}^{(k)}), \tag{13.97}$$

все сомножители которого для $k \geqslant 1$ можно определить с помощью регрессионной модели (13.37). Если случайная переменная $e(k)$ этой модели имеет нормальное распределение и не зависит от переменных $y(k)$ и $u(k)$, то имеет место

$$f[e(k) \mid y_{(-N+1)}^{(k-1)}, u_{(-N+1)}^{(k)}] = f[e(k)] =$$

$$= (2\pi)^{-p/2} | R^{-1} |^{1/2} \exp\left\{-\frac{1}{2} e^{\mathrm{T}}(k) R^{-1} e(k)\right\}, \tag{13.98}$$

где p — число выходов объекта.

Выражая теперь случайную переменную $e(k)$ с помощью регрессионной модели (13.39), найдем функцию правдоподобия

$$L(\Theta, R) = \prod_{k=1}^{K} f[y(k) \mid y_{(-N+1)}^{(k-1)}, \quad u_{(-N+1)}^{(k)}; \Theta, R] =$$

$$= (2\pi)^{-pK/2} | R^{-1} |^{K/2} \exp\left\{-\frac{1}{2} \sum_{k=1}^{K} [y(k) - \Theta^{\mathrm{T}} z(k)]^{\mathrm{T}} R^{-1} [y(k) - \Theta^{\mathrm{T}} z(k)]\right\}. \tag{13.99}$$

Следует отметить, однако, что произведение (13.97) не определяет плотность вероятности при $k < 1$. Произведение (13.97) при этом условии должно рассматриваться как начальное условие.

Регрессионная модель (13.39) соответствует замкнутой системе, поскольку матрица $z(k)$ и, следовательно, матрица Z, строками которой являются $z^{\mathrm{T}}(k)$ при $k = 1, 2, \ldots, K$, содержат переменные, зависящие от предыдущих значений выхода объекта и, следовательно, от предыдущих значений случайной переменной $e(k-1)$, $i = 1, 2, \ldots$ Кроме того, входы u, которые также являются элементами $z(k)$, также зависят от предыдущих значений.

13.3.1. Вычисление оценок максимального правдоподобия. Общее решение задачи, т.е. определение параметров Θ функции правдоподобия (13.99), которые доставляют максимум этой функции, хорошо известно и было опубликовано в литературе, например, в [58.1]. Здесь мы остановимся в основном на численном решении этой задачи, которое, как мы покажем, может быть применено и к методу наименьших квадратов и к идентификации в реальном масштабе времени.

Натуральный логарифм функции правдоподобия (13.99) равен

$$\ln L(\Theta, R) = -\frac{pK}{2} \ln(2\pi) + \frac{K}{2} \ln | R^{-1} | - \frac{1}{2} \sum_{k=1}^{K} [y(k) - \Theta^{\mathrm{T}} z(k)]^{\mathrm{T}} R^{-1} [y(k) - \Theta^{\mathrm{T}} z(k)]. \tag{13.100}$$

Искомые параметры определяются последним слагаемым в правой части выражения (13.100).

Следует отметить, что ковариационная матрица R положительно определена, поскольку ни одна случайная компонента вектора $e(k)$ не имеет нулевую дисперсию. Факторизуем эту матрицу следующим образом:

$$R = (R^{1/2})^{\mathrm{T}} R^{1/2}, \tag{13.101}$$

где $R^{1/2}$ – верхняя треугольная матрица, которая является квадратным корнем Холецкого из матрицы R. Введем обозначение

$$e^*(k) = (R^{-1/2})^{\mathrm{T}}[y(k) - \Theta^{\mathrm{T}} z(k)] = (R^{-1/2})^{\mathrm{T}} e(k). \tag{13.102}$$

Последнее слагаемое в (13.100) теперь принимает вид

$$\frac{1}{2}\sum_{k=1}^{K} e^{*\mathrm{T}}(k) e^*(k) = \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{p} e_i^{*2}(k), \tag{13.103}$$

где $e_i^*(\cdot)$ – отдельные компоненты вектора-столбца $e^*(k)$.

Транспонируя теперь уравнение (13.102) и записывая его для всех значений $k = 1, 2, \ldots, K$, мы получим систему линейных алгебраических уравнений, которую можно записать в матричной форме

$$E_e^* = [Y - Z\Theta] R^{-1/2}. \tag{13.104}$$

Размерности отдельных матриц здесь совпадают с размерностями соответствующих матриц в уравнении (13.56). Теперь мы преобразуем уравнение (13.104) к виду

$$E_e^* = [Z, Y]\begin{bmatrix} -\Theta \\ \mathrm{E} \end{bmatrix} R^{-1/2} = D\tilde{\Theta} R^{-1/2}, \tag{13.105}$$

где размерности отдельных матриц равны $E_e^*(K;\ p)$, $D(K;\ \nu + p)$, $\tilde{\Theta}(\nu + p;\ p)$ и $R(p;\ p)$. Единичная матрица Е имеет ту же размерность, что и матрица R.

Отметим теперь, что последнее слагаемое в правой части выражения (13.100) можно представить как квадрат евклидовой нормы E_e^*, т.е.

$$\sum_{k=1}^{K}[y(k) - \Theta^{\mathrm{T}} z(k)]^{\mathrm{T}} R^{-1}[y(k) - \Theta^{\mathrm{T}} z(k)] = \|E_e^*\|^2. \tag{13.106}$$

Теперь мы воспользуемся тем фактом, что ортогональное преобразование не изменяет нормы матрицы, так что

$$\|E_e^*\| = \|TE_e^*\|, \tag{13.107}$$

если

$$T^{\mathrm{T}} T = \mathrm{E}.$$

Мы выберем матрицу преобразования таким образом, чтобы преобразовать матрицу D в верхнюю треугольную матрицу (см. приложение Г), т.е. такую, что

$$TD = \begin{bmatrix} D_\Delta \\ 0 \end{bmatrix}. \tag{13.108}$$

Таким образом, все элементы матрицы D оказываются собранными в верхней треугольной матрице D_∇ размерности $(\nu + p,\ \nu + p)$, которая, следовательно, не зависит от K и содержит всю информацию, необходимую для нахождения оценок $\hat{\Theta}$ и $\hat{R}$.

Поскольку с увеличением числа измерений, которые последовательно преобразуются в элементы матрицы D_∇, происходит рост элементов этой матрицы, то мы запоминаем в памяти ЦВМ так называемую *информационную матрицу* F, которая определяется соотношением

$$F = \frac{1}{\sqrt{K}} D_\nabla. \tag{13.109}$$

Разобьем эту матрицу

$$F = \begin{bmatrix} F_{\Theta\Theta} & F_{\Theta R} \\ 0 & F_{RR} \end{bmatrix},$$

где $F_{\Theta\Theta}$ и F_{RR} – снов
соответственно. С помо
(13.105) мы можем запи

$$\|E_e^*\|^2 = \left\| TD \begin{bmatrix} -\Theta \\ \mathrm{E} \end{bmatrix} \right.$$

$$= K\|[F_{\Theta R} - F_{\Theta\Theta}\Theta]$$

Для того чтобы функц
ны приравнять нулю
выражения, поскольку
отрицательный знак. От

$$F_{\Theta R} - F_{\Theta\Theta}\Theta = 0,$$

Величина логарифма ф

$$\ln L(\Theta, R) = -\frac{pK}{2}$$

Было показано (см., н
достигает своего макси

$$F_{RR} R^{-1/2} = \mathrm{E},$$

так что норма в выраж
ного корня Холецкого
бия принимает значени

$$L(\Theta, R) = \{\ (2\pi e)^{p/}$$

где $[\hat{R}^{1/2}]_{ii}$ – элемен

Все это позволяет
параметров регресси
ковариационной мат
метров в одном ре
положительно опред
правого квадратного
модели (13.104) опр
блоки информацион
(13.109), а сами бло

С помощью выраж

$$F^{\mathrm{T}} F = \frac{1}{K} D^{\mathrm{T}} D.$$

Применяя подстано

$$D = [Z, Y],$$

которая была вве
(13.110), мы полу
перь классическим

$$\hat{\Theta} = F_{\Theta\Theta}^{-1} F_{\Theta R} =$$

$$\hat{R} = F_{RR}^{\mathrm{T}} F_{RR} =$$

Разобьем эту матрицу на четыре блока

$$F = \begin{bmatrix} F_{\Theta\Theta} & F_{\Theta R} \\ 0 & F_{RR} \end{bmatrix}, \tag{13.110}$$

где $F_{\Theta\Theta}$ и F_{RR} — снова верхние треугольные матрицы размерности $(\nu;\ \nu)$ и $(p;\ p)$ соответственно. С помощью информационной матрицы вида (13.110) и соотношения (13.105) мы можем записать

$$\| E_e^* \|^2 = \left\| TD \begin{bmatrix} -\Theta \\ \mathrm{E} \end{bmatrix} R^{-1/2} \right\|^2 = K \left\| F \begin{bmatrix} -\Theta \\ \mathrm{E} \end{bmatrix} R^{-1/2} \right\|^2 =$$

$$= K \| [F_{\Theta R} - F_{\Theta\Theta}\Theta] R^{-1/2} \|^2 + K \| F_{RR} R^{-1/2} \|^2 . \tag{13.111}$$

Для того чтобы функция правдоподобия достигла максимального значения, мы должны приравнять нулю соответствующим выбором Θ первое слагаемое полученного выражения, поскольку в логарифме функции правдоподобия это слагаемое имеет отрицательный знак. Отсюда следует соотношение для матрицы параметров

$$F_{\Theta R} - F_{\Theta\Theta}\Theta = 0, \qquad \Theta = F_{\Theta\Theta}^{-1} F_{\Theta R}. \tag{13.112}$$

Величина логарифма функции правдоподобия при этих значениях параметров равна

$$\ln L(\Theta, R) = -\frac{pK}{2} \ln(2\pi) + \frac{K}{2} \ln |R^{-1}| - \frac{K}{2} \| F_{RR} R^{-1/2} \|^2 . \tag{13.113}$$

Было показано (см., например, [58.1, 73.18]), что логарифм функции правдоподобия достигает своего максимума при

$$F_{RR} R^{-1/2} = \mathrm{E}, \qquad F_{RR} = R^{1/2}, \tag{13.114}$$

так что норма в выражении (13.113) равна $\|\mathrm{E}\|^2 = p$, а F_{RR} — оценка правого квадратного корня Холецкого из ковариационной матрицы R. При этом функция правдоподобия принимает значение

$$L(\Theta, R) = \{ (2\pi e)^{p/2} \prod_{i=1}^{p} [\hat{R}^{1/2}]_{ii} \}^{-K}, \tag{13.115}$$

где $[\hat{R}^{1/2}]_{ii}$ — элементы главной диагонали матрицы $\hat{R}^{1/2}$.

Все это позволяет сделать следующий вывод: если $K \geqslant p\nu + (p + p^2)/2$, где $p\nu$ — число параметров регрессионной модели объекта, $(p + p^2)/2$ — число неизвестных элементов ковариационной матрицы R, p — число линейно независимых решений, а ν — число параметров в одном решении, и если матрица D полного ранга, а ковариантная матрица R положительно определена, тогда оценки максимального правдоподобия матрицы Θ и правого квадратного корня Холецкого из ковариационной матрицы R регрессионной модели (13.104) определяются выражениями (13.112) и (13.114), где $F_{\Theta\Theta}$, $F_{\Theta R}$ и F_{RR} — блоки информационной матрицы F, которая определяется выражениями (13.108) и (13.109), а сами блоки определяются выражением (13.110).

С помощью выражений (13.108) и (13.109) мы можем записать

$$F^{\mathrm{T}} F = \frac{1}{K} D^{\mathrm{T}} D. \tag{13.116}$$

Применяя подстановку

$$D = [Z, Y],$$

которая была введена в уравнении (13.105), и подставляя F согласно выражению (13.110), мы получим оценки $\hat{\Theta}$ и $\hat{R}$, которые неоднократно опубликованы и стали теперь классическими формулами:

$$\hat{\Theta} = F_{\Theta\Theta}^{-1} F_{\Theta R} = (Z^{\mathrm{T}} Z)^{-1} Z^{\mathrm{T}} Y, \tag{13.117}$$

$$\hat{R} = F_{RR}^{\mathrm{T}} F_{RR} = \frac{1}{K} [Y^{\mathrm{T}} Y - Y^{\mathrm{T}} Z (Z^{\mathrm{T}} Z)^{-1} Z^{\mathrm{T}} Y]. \tag{13.118}$$

Можно также показать, что полученные оценки максимального правдоподобия (13.112) и (13.114) или (13.117) и (13.118), вычисленные в предположении о нормальном распределении случайного вектора $e(k)$ в регрессионной модели (13.39), имеют естественный физический смысл даже тогда, когда случайный вектор не является нормальным и если матрица R в уравнении (13.111), например, произвольная весовая положительно полуопределенная матрица W.

Сравнивая полученное выражение (13.117) с формулой (13.59), можно заключить, что обе оценки совпадают. Это совпадение может произойти только в случае линейных систем. Мы можем, следовательно, заключить, что для линейных систем с гауссовым шумом оценки параметров регрессионной модели (13.39), получаемые с помощью методов наименьших квадратов и максимального правдоподобия, совпадают.

13.3.2. Свойства оценок максимального правдоподобия. При определении статистических свойств оценок максимального правдоподобия мы должны ограничиться, как и в случае метода наименьших квадратов, первым и вторым центральным моментами, т.е. средним значением оценки и ковариационной матрицей оценки. Мы приведем без доказательств, которые можно найти в [73.18], только наиболее важные результаты.

Вначале следует отметить, что полезно различить свойства оценок максимального правдоподобия для разомкнутых и замкнутых систем управления. Пусть соответствующая регрессионная модель имеет вид

$$Y = Z\Theta + E_e. \tag{13.119}$$

Подставляя Y (13.119) в выражение (13.117), мы получим

$$\Delta\hat\Theta = \hat\Theta - \Theta = (Z^T Z)^{-1} Z^T E_e = \mathcal{L}^T E_e. \tag{13.120}$$

В разомкнутых системах управления случайная переменная $e(k)$ и, следовательно, элементы матрицы E_e не зависят от элементов матрицы Z и поэтому

$$\mathcal{E}[E_e \mid Z] = \mathcal{E}[E_e] = 0, \tag{13.121}$$

и мы находим из выражения (13.120), что

$$\mathcal{E}\Delta\Theta = \mathcal{E}(\hat\Theta - \Theta) = 0, \tag{13.122}$$

т.е. среднее значение ошибок оценок параметров равно нулю.

Для ковариации оценки $\hat\Theta$ мы имеем соотношение

$$\mathcal{E}\Delta\hat\Theta_{ij}\Delta\hat\Theta_{rs} = \mathcal{E}\left\{\sum_{k=1}^{K} \mathcal{L}_{ki}\, e_j(k) \sum_{l=1}^{K} \mathcal{L}_{lr} e_s(l)\right\} =$$

$$= \mathcal{E}\left\{\sum_{k=1}^{K}\sum_{l=1}^{K} \mathcal{L}_{ki}\, \mathcal{L}_{lr}\, \mathcal{E}[e_j(k)\, e_s(l) \mid Z]\right\}, \quad i, r = 1, 2, \ldots, \nu, \quad j, s = 1, 2, \ldots, p, \tag{13.123}$$

где аргументы k и l соответствуют различным моментам времени, а индексы j и s соответствуют отдельным выходным переменным. Поскольку, как это уже отмечалось, случайные величины не зависят от значений элементов матрицы Z, то мы имеем

$$\mathcal{E}[e_j(k)\, e_s(l)] = \delta_{kl} R_{js}, \tag{13.124}$$

где δ_{kl} – символ Кронекера, а R_{js} – элементы ковариационной матрицы шума.

После ряда преобразований выражение (13.123) принимает вид

$$\mathcal{E}[\Delta\hat\Theta_j \Delta\hat\Theta_s^T] = R_{js}\mathcal{E}[(Z^T Z)^{-1}] = R_{js}\mathcal{E}\left[\frac{1}{K} F_{\Theta\Theta}^{-1} F_{\Theta\Theta}^{-T}\right]. \tag{13.125}$$

Из этого результата следует, что $\mathcal{E}[(Z^T Z)^{-1}]$ с точностью до коэффициента R_{js} является матрицей ковариации j-го и s-го столбца матрицы $\hat\Theta$.

При вычислении среднего значения оценки $\hat R$ мы снова воспользуемся регрессионной моделью (13.119) и выражением (13.118) для оценки R. Мы получим

$$\mathcal{E}\hat R = \frac{K-\nu}{K} R, \tag{13.126}$$

где ν – число параметров
$y(k)$. Из этого результата
чтобы она была несмещен
как это показало в следук

$$\hat R = \frac{K}{K-\nu} F_{RR}^T F_{RR}$$

или

$$\hat R^{1/2} = \left(\frac{K}{K-\nu}\right)^{1/2} F_{R}$$

Это означает, что при ч
ную матрицу (13.109) к

$$F = \frac{1}{(K-\nu)^{1/2}} D_\nabla$$

для того, чтобы для люб

Рассмотрим теперь з
Θ и R совпадают с оце
тистические свойства о

(а) матрица M^{-1}, г
положительно определ

(б) $M^{-1}[(1/K) Z^T Z$
це, то можно доказа
асимптотически.

13.3.3. Определение
задача идентификации
метров регрессионной
получения оценок пар
в виду, что у модели
сионной модели, и пр
ми и выходными пере
ров модели состояни
ными переменными
менных состояния м
дов решения был опи

Ниже мы приведе
мого объекта. В это
тояния, так и входн
модель (13.31) и
(13.31). Пусть слу

$\mathcal{E}[w(k)] = 0,$

$\mathcal{E}[w(k)\, w^T(k-\kappa$

$\mathcal{E}[w(k)\, u^T(k-\kappa$

$\mathcal{E}[w(k)\, y^T(k-\kappa$

Запишем уравнение

$x(k+1) = [A, B]$

где

$$z(k) = \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}$$

где ν — число параметров, которое соответствуют одной выходной компоненте вектора $y(k)$. Из этого результата следует, что оценка R не является несмещенной. Для того чтобы она была несмещенной, ее нужно умножить на соответствующий множитель так, как это показало в следующих соотношениях:

$$\hat{R} = \frac{K}{K-\nu} F^{\mathrm{T}}_{RR} F_{RR} \qquad (13.127)$$

или

$$\hat{R}^{1/2} = \left(\frac{K}{K-\nu}\right)^{1/2} F_{RR}.$$

Это означает, что при численных расчетах целесообразно преобразовать информационную матрицу (13.109) к виду

$$F = \frac{1}{(K-\nu)^{1/2}} D_{\nabla} \qquad (13.128)$$

для того, чтобы для любого момента времени иметь несмещенную оценку $\hat{R}^{1/2} = F_{RR}$.

Рассмотрим теперь замкнутые системы управления. Хотя оценки матриц параметров Θ и R совпадают с оценками этих матриц в разомкнутых системах управления, их статистические свойства оказываются различными. Если же, однако,

(а) матрица M^{-1}, где $M = \mathcal{E}\,[(1/K)\,Z^{\mathrm{T}} Z] = \mathcal{E}\,[F^{\mathrm{T}}_{\Theta\Theta} F_{\Theta\Theta}]$, сходится при $K \to \infty$ к положительно определенной матрице и

(б) $M^{-1}\,[(1/K)\,Z^{\mathrm{T}} Z] = M^{-1} F^{\mathrm{T}}_{\Theta\Theta} F_{\Theta\Theta}$ сходится по вероятности к единичной матрице, то можно доказать, что равенства (13.122), (13.125) и (13.126) выполняются асимптотически.

13.3.3. Определение параметров модели состояния. В предыдущих разделах § 13.3 задача идентификации управляемого объекта была сведена к задаче оценивания параметров регрессионной модели. Аналогичную процедуру можно также применить для получения оценок параметров модели состояния управляемого объекта. Следует иметь в виду, что у модели состояния в общем случае число параметров больше, чем у регрессионной модели, и при их определении едва ли возможно ограничиться только входными и выходными переменными объекта. Для того чтобы облегчить определение параметров модели состояния, кроме этих переменных нужно воспользоваться также измеренными переменными состояния модели. Одновременное оценивание параметров и переменных состояния модели объекта — очень трудная задача. Один из возможных методов решения был описан Л. Льюнгом [77.18].

Ниже мы приведем процедуру определения параметров модели состояния управляемого объекта. В этой процедуре предполагается, что измеряются как переменные состояния, так и входные и выходные переменные. Рассмотрим снова математическую модель (13.31) и (13.32). Сначала мы определим параметры уравнения динамики (13.31). Пусть случайный процесс $w(k)$ характеризуется следующими свойствами:

$$\begin{aligned}
&\mathcal{E}\,[w(k)] = 0,\\
&\mathcal{E}\,[w(k)\,w^{\mathrm{T}}(k-\kappa)] = 0; \quad \kappa = 1, 2, \ldots, k,\\
&\mathcal{E}\,[w(k)\,u^{\mathrm{T}}(k-\kappa)] = 0; \quad \kappa = 1, 2, \ldots, k,\\
&\mathcal{E}\,[w(k)\,y^{\mathrm{T}}(k-\kappa)] = 0; \quad \kappa = 1, 2, \ldots, k.
\end{aligned} \qquad (13.129)$$

Запишем уравнение (13.31) в виде

$$x(k+1) = [A, B]\; z(k) + w(k), \qquad (13.130)$$

где

$$z(k) = \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}.$$