УДК 681.3

ОПТИМИЗАЦИЯ ПРОЦЕССА ИДЕНТИФИКАЦИИ В КРИТИЧЕСКИХ СИСТЕМАХ

к.т.н. В.А. Тимофеев
(представил д.т.н., проф. О.Г. Руденко)

Рассматривается задача оценивания параметров авторегрессионной модели для случая, когда отсутствует информация о статистических свойствах сигналов, а известен лишь уровень помехи. Выполнено сравнительное моделирование предлагаемого алгоритма и МНК.

Введение. Эффективность создаваемых систем управления реальными объектами в значительной степени зависит от качества используемых при этом математических моделей исследуемых объектов. Получаемые в результате решения задачи идентификации модели должны, с одной стороны, адекватно отражать свойства этих объектов, а с другой – быть достаточно простыми и удобными для синтеза алгоритма управления такой информации в виде известной плотности распределения помех либо информации о принадлежности неизвестной плотности какому-либо классу распределений. Однако зачастую сведения о статистических свойствах сигналов и помех отсутствуют, а исследователь обладает информацией лишь об их уровнях.

Целью настоящей работы является дальнейшее развитие подхода, обеспечивающего получение оценок параметров объекта при отсутствии априорной информации о статистических свойствах сигналов и помех. Рассматриваются пути оптимизации процесса идентификации.

Постановка задачи. Рассмотрим модель динамического объекта, описываемого уравнением псевдолинейной регрессии, имеющую вид

$$y(k) = \Theta^T \psi(k-1) + w(k), \ \left| w(k) \right| \le \delta(k), \qquad (1)$$

где $\psi(k-1)$ – вектор обобщенных входов; Θ – вектор неизвестных параметров; $y(k)$ – выходной сигнал; $w(k)$ – неопределенный, но ограниченный по амплитуде шум; $k = 1, 2, \dots, N$ – дискретное время.

Задача идентификации состоит в определении области D, внутри которой должны лежать параметры Θ исходной информацией, для построения которой имеется последовательность наблюдений $\{y(k)\}$ и векторов обобщенных входов $\{\psi(k)\}$.

© к.т.н. В.А. Тимофеев, 2004

Решение задачи. Уравнение (1) может быть переписано в виде пары неравенств

$$y(k)-\delta(k)\le\psi^{T}(k-1)\Theta\le y(k)+\delta(k)\,, \tag{2}$$

задающих границы области D, внутри которой должны лежать параметры θ. Можно заметить, что неравенства (2) задают пару гиперплоскостей в пространстве θ, между которыми находится искомый вектор параметров. Последовательность наблюдений $y(1), y(2), \dots, y(N)$ порождает N пар гиперплоскостей, «высекающих» в пространстве некоторую область D_N. Это и есть область оценок параметров θ, причем очевидно, что полученная область есть политоп [1, 2]. Каждое новое наблюдение изменяет эту область, относительно которой можно заметить, что все точки, принадлежащие этой области, равноправны в том смысле, что среди них нельзя выделить «наилучшую» оценку, хотя для удобства можно использовать некий «центр» области $D(N)$.

Пусть в качестве метрики в пространстве параметров Θ используется евклидово расстояние. Тогда расстояние от начала координат до гиперплоскости вида

$$\psi^{T}\theta=z \tag{3}$$

есть $z/\|\psi\|$, а расстояние между границами (3) есть $2\delta(k)/\|\psi(k)\|$.Очевидно, что чем меньше объем полученного политопа, тем меньше уровень неопределенности Θ. Таким образом видно, что вид области $D(N)$ зависит от выбора вектора обобщенных входов $\{\psi(k)\}$. Наилучший выбор $\psi(k)$ связан с максимизацией $\|\psi(k)\|$, что позволяет минимизировать расстояние и «стянуть» границы максимально близко друг к другу. При этом расстояние $2\delta(k)/\|\psi(k)\|$ можно интерпретировать как отношение сигнал/шум, которое может быть максимизировано соответствующим выбором $\psi(k)$. В случае, если в качестве расстояния используется неевклидова метрика, можно перейти от модели (1) к модели

$$y(k)=\Theta^{*T}\psi^{*T}(k-1)+w(k)\,,\ |w(k)|\le\delta(k)\,,\quad k=1,2,\dots,N, \tag{4}$$

где Θ^{*} и $\psi^{*}(k)$ связаны с Θ и $\psi(k)$ невырожденным линейным преобразованием G и G^{-1} соответственно и выбираемые так, что

$$\|\Theta^{*}\|=\left(\Theta^{T}G^{T}G\Theta\right)^{1/2}. \tag{5}$$

В общем случае проблема выбора последовательности векторов обобщенных входов $\{\psi(k)\}$ состоит в том, чтобы после N шагов уточнения сделать область $D(N)$ как можно более малой. Информативность каждого нового наблюдения $y(N)$ зависит от $D(N-1)$ и, следовательно, от всех уже реализованных $\psi(N-1)$ и $w(N)$. Первое и очевидное требование к векторам обобщенных входов состоит в том, что последовательность $\{\psi(N)\}$ должна образовывать базис пространства Θ. Если это не так, например, в силу существенной линейной зависимости входов, область $D(N)$ остается неограниченной, по крайней мере, в одном направлении. В том же случае, если набор $\psi(1), \psi(2), \ldots, \psi(N)$ – линейно независим, объем политопа $D(N)$ может быть минимизирован за счет взаимной ортогональности образующих его гиперплоскостей. Неортогональность входов увеличивает объем области $D(k)$ на величину $1 \Big/ \prod_{k=1}^{N} \cos\alpha(k)$, где последовательность $\alpha(1), \alpha(2), \ldots, \alpha(N)$ задает углы между $\psi(k)$. Отсюда видно, что оптимальный выбор векторов ψ также может быть связан с максимизацией произведения $\prod_{t=1}^{N} \|\psi(k)\| \cos\alpha(k)$.

До тех пор, пока на объект подаются линейно независимые векторы, форма и размер области настраиваемых параметров не зависят от длины амплитуд этих векторов. Здесь определяющим является только их взаимное расположение. Каждое последующее наблюдение, начиная с $y(n+1)$, где n – размерность пространства, уже существенным образом влияет как на объем, так и на форму области D.

Таким образом, необходимым является проведение планирования эксперимента в задачах множественной идентификации с использованием методов линейного программирования [3 – 5] для выбора векторов обобщенных входов.

Важной задачей в процессе идентификации является определение эмпирических характеристик действующих возмущений. Данная задача близка к проблеме статистического оценивания параметров шума на основе инновационной последовательности [6].

Рассмотрим ситуацию, когда область $D(k)$ достаточно мала и только одна гиперплоскость из новой пары $\psi^{T}(k)\Theta = y(k) \pm \delta(k)$ пересекает эту область. При этом либо вектор Θ, соответствующий данной реализации, принадлежит области $D(k)$ и можно воспользоваться для оценки шума выражением

$$y(k) - w(k) = \psi^T(k-1)\Theta \approx y(k) \pm \delta(k),$$

либо Θ не принадлежит области $D(k)$ и тогда фактический уровень шума выше, чем заданные границы. Так возникает задача оценки границ шумов. Ранее эта задача не была решена для динамических объектов. Для нахождения границ шумов рассмотрим ARMAX-модель

$$y(k) = \frac{B(q)}{1 + A(q)} u(k) + w(k) \tag{6}$$

с известными границами шума $\{w\}$. Поскольку выражение (6) можно переписать в виде

$$y(k) = -A(q)y(k) + B(q)u(k) + \big(1+A(q)\big)w(k) = -A(q)y(k) + B(q)u(k) + \varepsilon(k), \tag{7}$$

то связь с (1) выражается соотношениями

$$\psi(k) = \big(-y(k-1), \ldots, -y(k-m), u(k-1), \ldots, u(k-m)\big),$$

$$\Theta^T = \big(a_1, \ldots, a_i,\ b_1, \ldots, b_m\big), \quad \varepsilon(k) = w + a_1 w(k-1) + a_n w(k-n).$$

Видно, что неопределенность коэффициентов $\{a_i\}$ вызывает неопределенность в границах $\{\varepsilon\}$. Границы $\{\varepsilon\}$ могли бы быть получены рекуррентно путем использования значений $a_1, \ldots, a_n$ на предыдущей итерации, причем полученные границы не будут точными, поскольку зависят от $\{a_i\}$. Границы, полученные в k-й момент, могут быть записаны в виде

$$y(k) - \delta(k) + a_1 v(k-1) + \ldots + a_n v(k-n) \le b_1 u(k-1) + \ldots + b_m u(k-m) \le$$
$$\le y(k) + \delta(k) + a_1 v(k-1) + \ldots + a_n v(k-n),$$

где $\{v\}$ – “чистый” выход; $\{y - \varepsilon\}$ и $\pm\delta(k)$ – границы по $\delta(k)$.

Несложно видеть, что

$$y(k-i) - \delta(k-i) \le v(k-i) \le y(k-i) + \delta(k-i)\,, \quad i = 1, 2, \ldots, l. \tag{8}$$

Наименьшее из возможных значений $a_i(y(k-i) \pm \delta(k-i))$ $\forall\, a_i \in D(k)$ может использоваться в качестве оценки $a_i v(k-i)$ в левой части неравенства (8). При этом нужно помнить, что поскольку a_i может принимать множество значений, градиент каждой из ограничивающих гиперплоскостей (8) может в принципе изменяться скачком. В момент перехода a_i через ноль соответственно меняется знак в $a_i(y(k-i) \pm \delta(k-i))$. Однако изменения градиента могут привести к нарушению выпуклости области D, что делает неэффективной аппроксимацию области D некоторой фигурой. Данная проблема упрощается, если знаки a_i известны заранее.

Рассмотрим динамический объект, описываемый ARX-моделью

$$y(k) = -a_1 y(k-1) - \ldots - a_n y(k-n) + b_1 u(k-1) + \ldots + b_m u(k-m) + w(k), \quad (9)$$

и перепишем уравнение (9) в виде (4), где вектор истинных параметров записывается как

$$\Theta^* = \left(a_1, a_2, \ldots, a_n, b_1, \ldots, b_m\right)^T,$$

а вектор обобщенных входов имеет вид

$$\psi^*(k-1) = \left(-y(k-1), -y(k-2), \ldots, -y(k-N), u(k-1), \ldots, u(k-m)\right)^T.$$

Тогда область $D(k)$ является подмножеством евклидового пространства R^{n+m+1} и определяется как

$$D(k) = \left\{\Theta : \left(y(k) - \Theta^T \psi(k-1)\right)^2 \le \delta^2(k), \Theta \in R^{n+m}\right\}.$$

Алгоритм начинается с построения достаточно большого эллипсоида $E(0)$ в пространстве R^{n+m+1} и содержащего все возможные допустимые значения вектора параметров модели Θ^*. После получения первого наблюдения $y(1)$ может быть найден эллипсоид, построенный на пересечении $E(0)$ и выпуклого политопа $D(1)$. Для ускорения сходимости алгоритма, этот эллипсоид должен быть оптимизирован по какому-либо критерию, например: по критерию минимального объема или минимального следа. Обозначим оптимальный эллипсоид как $E(1)$. После получения второго наблюдения $y(2)$, аналогичным образом найдем эллипсоид, построенный на пересечении $E(1)$ и выпуклого политопа $D(2)$ и так далее для всех последующих наблюдений. Таким образом может быть получена последовательность оптимальных эллипсоидов $\{E(k)\}$, центры которых могут приниматься как оценка параметров в k-й момент времени и обозначаться как $\hat{\Theta}(k)$. Если в момент времени i результирующий оптимальный эллипсоид имеет наименьший размер по сравнению с предыдущими, то данные $y(i)$ указывают, что содержится некоторая новая "информация" относительно оценок параметров, что позволяет корректировать эти оценки. В противном случае $E(i)$ устанавливается равным $E(i-1)$, и оценки не уточняются. В сущности, рекуррентный алгоритм состоит из двух модулей: эволюционного модуля и модуля, где информация корректируется. В каждый момент времени полученные данные корректируются в случае, если они содержат некоторую новую информацию.

Эллипсоид в k-й момент времени может быть записан как

$$E(k) = \left\{ \Theta \in R^{n+m+1} : \left(\Theta - \hat{\Theta}^T(k) \right)^T P^{-1}(k) \left(\theta - \hat{\theta}(k) \right) \le \sigma^2(k) \right\},$$

где используются положительно-определенная матрица полуосей $P(k)$ и неотрицательная скалярная величина $\sigma^2(k)$, определяемая следующим образом:

$$\sigma^2(k) = (1 + \lambda(k))\sigma^2(k-1) + \lambda(k)\delta^2(k) - \frac{\lambda(k)(1 - \lambda(k))e^2(k)}{1 - \lambda(k) + \lambda(k)G(k)},$$

где ошибка идентификации $e(k)$ и коэффициент $G(k)$ имеют соответственно вид

$$e(k) = y(k) - \hat{\theta}^T \psi(k-1), \; G(k) = \psi^T(k-1)P(k-1)\psi(k-1).$$

Размер эллипсоида связан со скалярной величиной $\sigma^2(k)$ и матрицей $P(k)$, при этом алгоритм оценивания выглядит следующим образом:

$$\begin{cases} \hat{\Theta}(k) = \hat{\Theta}(k-1) + K(k)\sigma(k); \\ K(k) = \dfrac{\lambda(k)P(k-1)\psi(k-1)}{1 - \lambda(k) + \lambda(k)G(k)}; \\ P(k) = \dfrac{1}{1 - \lambda(k)} \left[I - K(k)\psi^T(k-1) \right] P(k-1). \end{cases} \tag{10}$$

Оптимальный эллипсоид $E(k)$ построен на пересечении $E(k)$ и $D(k)$ и определен в терминах коэффициента коррекции $\lambda(k)$, где $0 \le \lambda(k) \le \alpha < 1$, и где пользователь сам выбирает верхнюю границу этого коэффициента. Оптимальное значение $\lambda(k)$ может быть определено минимизацией $\varepsilon^2(k)$ в каждый момент времени. Процедура минимизации заканчивается последовательной процедурой корректировки. В частности, установим $\lambda(k)$ равным нулю (нет коррекции), если

$$\sigma^2(k-1) + e^2(k) \le \delta^2(k). \tag{11}$$

С другой стороны, если (11) не выполняется, то оптимальные значения $\lambda(k)$ вычисляются следующим образом:

$$\lambda(k) = \min(\alpha, \gamma), \tag{12}$$

где

$$\beta(k) = \frac{\delta^2(k) - \sigma^2(k-1)}{e^2(k)}; \tag{13}$$

$$\gamma = \begin{cases} \alpha, & \sigma^2(k) = 0; \\ \dfrac{1-\beta(k)}{2}, & G(k) = 1; \\ \dfrac{1}{1-G(k)}\left[1 - \left(\dfrac{G(k)}{1+\beta(k)\big(G(k)-1\big)}\right)^{1/2}\right], & 1+\beta(k)\big(G(k)-1\big) > 0; \\ \alpha, & 1+\beta(k)\big(G(k)-1\big) \le 0. \end{cases}$$

Соотношения (10) и стратегия коррекции наряду с начальными значениями:

$$\begin{aligned} &P^{-1}(0) = I; \\ &\hat{\Theta}(0) = 0; \\ &\sigma^2(0) = 1/\Delta; \\ &\Delta \ll 1 \end{aligned} \tag{14}$$

составляют алгоритм идентификации. Значение, выбранное для верхней границы $\delta(k)$, не должно быть тесно связанно с величиной шума, так как на оценки параметра не воздействуют оценки границ шума. Однако переоценивание границ шума может привести к тому, что «раздует» ограничивающий эллипсоид. Недооценивание границ шума может привести к тому, что $\sigma^2(k)$ станет отрицательным в некоторый момент времени, и тогда, ограничивающий эллипсоид может исчезнуть. В этом случае можно инициировать или увеличение размера эллипсоида $E(k-1)$, или увеличение ширины $D(k)$ увеличением $\delta(k)$.

Моделирование. В качестве примера рассмотрим следующую модель

$$y(k) = -0{,}4y(k-1) - 0{,}85y(k-2) - 0{,}2u(k-1) - 0{,}7u(k-2) + w(k),$$

где $u(k)$ – измеряемый вход, равномерно распределенный в интервале $[-1, 1]$; $w(k)$ – гармоническая составляющая в белом шуме.

Такая ситуация может возникнуть, когда на наблюдения воздействует, например, помеха от источника промышленной частоты питания или другое электромагнитное вмешательство. Примем следующую модель возмущения:

$$w(k) = (1-\beta)\,w(k-1) + \beta \sin\left(\frac{\pi k}{10}\right).$$

Значение β варьирует от 0 до 1 для каждого β; АМИ и РМНК-алгоритмы реализованы с числом данных 500 точек. Значение верхней границы для коэффициента усиления коррекции $\alpha = 0.5$ и верхняя граница для шума $\delta(k) = 1{,}0$. Конечная ошибка оценки параметра $\left(\Theta^* - \Theta(500)\right)^{\mathrm{T}}\left(\Theta^* - \Theta(500)\right)$ усреднена по десяти выборкам, и изображена на рис. 1 для β в пределах от 0 до 1.

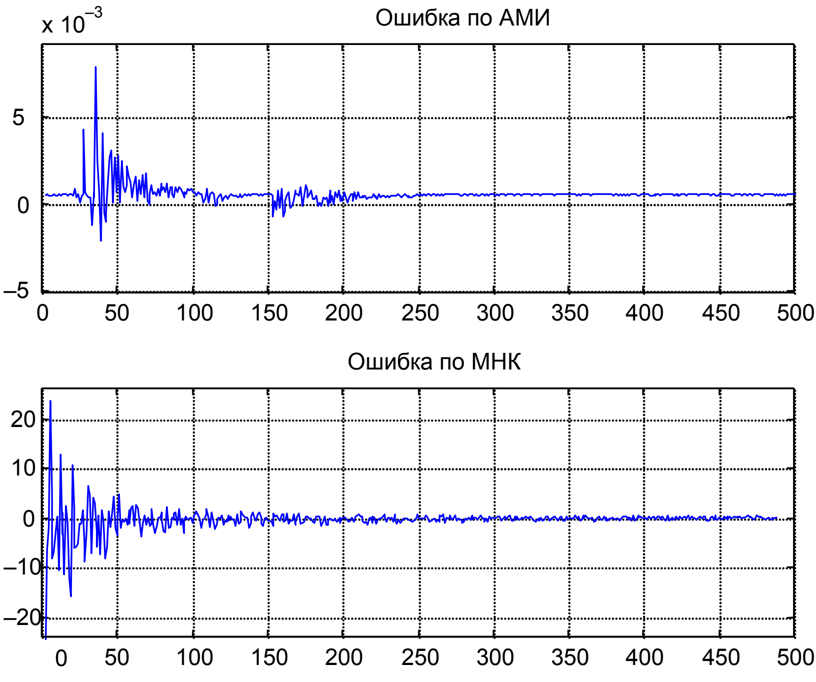

Рис. 1. Ошибка оценки настраиваемых параметров по АМИ и МНК

сли оценки параметров РМНК алгоритма недопустимы для больших начений β, то эффективность АМИ алгоритма – относительно постоянна в заданном диапазоне β.

Эффективность АМИ алгоритма также превосходит РМНК алгоритм и в других случаях, в которых шум является импульсным с резкими выбросами.

Заключение. Оценивание параметров объекта при отсутствии априорной информации о статистических свойствах сигналов и помех представляет собой сложную задачу, решение которой можно разбить на ряд этапов, связанных с проведением как активного, так и пассивного экспериментов. Если планирование эксперимента можно осуществить с привлечением стандартных методов линейного программирования, то для оценивания неизвестных параметров необходима разработка новых методов, которые, как показывают результаты исследований, могут базироваться на МНК. При этом, однако, точность оценивания зависит от параметров присутствующих в измерениях шумов, которые обычно и являются неизвестными. В связи с этим обработка данных в режиме пассивного эксперимента должна осуществляться с использованием двух процедур оценивания, одна из которых является необходимой для определения реальных границ шума, а другая, использующая результаты первой – для получения непосредственно искомых оценок. Несмотря на кажущуюся сложность процесса оценивания, удовлетворяющие исследователя результаты могут быть получены с меньшими вычислительными затратами по сравнению со статистическими методами.

ЛИТЕРАТУРА

1. *Mo S.H., Norton J.P. Fast and robust algorithm to complete exact polytope parameter bounds // Math. Comp. Simulation. – 1990. – 32. – P. 481 – 493.*
2. *Стрейц В. Метод пространства, состоящий в теории дискретных линейных систем. – М.: Наука, 1985. – 294 с.*
3. *Васильев Ф.Г. Численные методы решения экстремальных задач. – М.: Наука, 1985. – 520 с.*
4. *Ecker I.E. Geometric programming: method computations and applications // SIAM J. – 1980. – 22. – P. 338 – 362.*
5. *Arruda L.V.R., Favier G. A review and a comparison of robust estimation methods // Prepr. 9th IFAC / IFORS Symp. On Identification and System Parameter Estimation. – Budapest, Hungary. – 1991. – P. 1027 – 1032.*
6. *Черноусько Ф.Л. Оценивание фазового состояния динамических систем. Метод эллипсоидов. – М.: Наука, 1988. – 320 с.*

Поступила 20.04.2004

***ТИМОФЕЕВ Владимир Александрович**, канд. техн. наук, доцент, ведущий научный сотрудник Харьковского национального университета радиоэлектроники. В 1975 году окончил Харьковский институт радиоэлектроники. Область научных интересов – контроль и управление динамическими объектами.*