УДК 65.011.56
DOI:

Информатика, вычислительная техника и управление

Ю.А. Кропотов, Н.Е. Холкина

ФУНКЦИИ РЕГРЕССИИ И МЕТОД НАИМЕНЬШИХ КВАДРАТОВ В ЗАДАЧАХ ПАРАМЕТРИЧЕСКОГО МОДЕЛИРОВАНИЯ В СИСТЕМАХ ОБМЕНА ИНФОРМАЦИЕЙ АКУСТИЧЕСКИМИ СИГНАЛАМИ

Исследуются решения задачи восстановления зависимостей методом построения функции регрессии, методом наименьших квадратов. Рассматривается оценивание параметров акустических сигналов методами параметрической оптимизации. Рассматриваются методы регуляризации при решении некорректных задач восстановления зависимостей линейной регрессии. Исследуются вопросы идентификации процессов и аппроксимации.

Ключевые слова: акустические сигналы, линейная регрессия, нелинейная регрессия, метод наименьших квадратов, параметрическое моделирование, квадратичная функция потерь.

Yu.A. Kropotov, N.E. Kholkina

REGRESSION FUNCTIONS AND LEAST-SQUARES METHOD IN PROBLEMS OF PARAMETRIC MODELING IN INFORMATION EXCHANGE SYSTEMS BY ACOUSTIC SIGNALS

This paper reports the solution investigations of dependencies recovery problems through the method of regression function formations with the least-squares method. The assessment of acoustic signal parameters by methods of parametric optimization is considered. Regulation methods at the solution of ill-conditioned problems of linear regression dependences recovery are under consideration. The identification problems of processes and approximation are under investigation. The investigation object is acoustic signals. The purpose of the work consists in the development of a signal parametric model and also in that of algorithms for the identification and assessment of processes against a background of interference.

Dependences recovery by methods of a linear regression and also belongs to the class of ill-conditioned problems. The incorrectness of a linear regression is explained by a possible growth of an error in continuous metrics, despite the fact that in a discrete metrics an error tends to zero.

A peculiarity of the generalization, based on the method of regression and least squares, consists in the realization of operators manifesting discrete data in the space of continuous or piecewise-continuous functions.

The obtained results on the development of methods and algorithms, directed on the recovery of smooth functions according to discrete data, open potentialities for the solution of problems of local processing and smoothing both stationary signals and non-stationary ones, the problems of the analysis of acoustic signals and speech dynamics, problems of the analysis of multi-extreme dependences.

Key words: acoustic signals, linear regression, non-linear regression, least-squares method, parametric modeling, loss quadratic function.

Введение

Оценка параметров акустических сигналов и помех основывается на эмпирических результатах измерений, полученных из эксперимента [1]. Известен ряд методов получения таких оценок. К ним относятся параметрические и непараметрические, прямые и косвенные методы.

Алгоритмы идентификации в исследованиях временных рядов и в оценивании процессов на фоне почти произвольных помех

Как известно, функция регрессии, которая зависит от некоторого числа параметров, может рассматриваться как отображение одного пространства переменных в другое [2 - 4]. При этом задача построения функции регрессии заключается в нахождении ее неизвестных параметров. Без ограничения общности эта функция может быть записана в виде

$$y(x) = \varphi(x, \theta),\ x \in R^n,\ \theta \in R^m,\ y \in R.$$

Обобщение задачи на случай, когда пространство зависимых переменных является многомерным, то есть $y \in R^q$, особых затруднений не вызывает.

На практике чаще всего используется линейная регрессия с функцией

$$y(x) = \langle \varphi(x), \theta \rangle = \sum_{k=1}^{m} \theta_k \varphi_k(x) = \phi^T(x)\theta.$$

Здесь $\varphi_k(x)$, $k = 1, \ldots, m$, образуют систему линейно независимых функций. Вектор $\phi(x) = (\varphi_1(x), \ldots, \varphi_m(x))^T$, $\theta = (\theta_1, \ldots, \theta_m)^T$. Известна также и квазилинейная регрессия [5], описываемая функцией

$$y(x) = g(\langle \varphi(x), \theta \rangle),$$

где функция $g(y)$ удовлетворяет необходимым условиям дифференцируемости.

Параметры $\theta_1, \ldots, \theta_m$ находятся в результате минимизации некоторой функции, заданной на дискретном множестве наблюдений. Эту задачу можно записать в виде

$$Q(y_k, \varphi(x_k, \theta), k = 1, \ldots, m) \to \min.$$

Обычно эта задача дополняется рядом ограничений, вытекающих из условий задачи или обеспечивающих ее корректное и робастное решение. При этом приходится иметь дело с задачей условной минимизации. (1)

В дальнейшем в качестве независимой переменной будет, как правило, выступать одномерная переменная – время. Соответственно и базисные функции $\varphi_k(x)$ будут зависеть от одной переменной.

В задачах линейной и квазилинейной регрессии часто используется квадратичная функция потерь

$$Q(\theta) = \sum_{j=1}^{N} \left(y_j - \langle \varphi(x_j), \theta \rangle\right)^2 \text{ или } Q(\theta) = \sum_{j=1}^{N} \left(y_j - g(\langle \varphi(x_j), \theta \rangle)\right)^2.$$

Минимизация подобных функций составляет содержание метода наименьших квадратов.

Метод наименьших квадратов и линейная регрессия применяются в адаптивных и обучающихся системах, алгоритмах идентификации, при анализе временных рядов и оценивании процессов на фоне почти произвольных помех [5; 6]. Во всех указанных областях наблюдаемый сигнал $y(t)$ рассматривается, как правило, в виде аддитивной смеси оцениваемого сигнала $x(t)$ и случайной помехи или шума $n(t)$:

$$y(t) = x(t) + n(t).$$

Оценка сигнала при этом представляется многочленом (1) по системе линейно независимых базисных функций $\{\varphi_k(t)\}_1^m$:

$$\hat{x}(t) = \sum_{k=1}^{m} \theta_k \varphi_k(t) = \boldsymbol{\theta}^T \varphi(t).$$

Если оценке подлежит векторный сигнал размерности p, то есть вектор $\mathbf{x}(t) = (x_1(t), \ldots, x_p(t))^T$, то его оценку можно записать в виде $\hat{\mathbf{x}}(t) = \boldsymbol{\Theta}^T \varphi(t)$.

Вектор наблюдаемых данных при этом можно аппроксимировать выражением

$$\hat{\mathbf{y}}(t) = \boldsymbol{\Theta}^T \varphi(t) + \mathbf{n}(t),$$

где $\mathbf{n}(t)$ – вектор помехи, а $\boldsymbol{\Theta}$ – матрица коэффициентов многочленов:

$$\boldsymbol{\Theta} = \begin{pmatrix} \theta_{11} & \cdots & \theta_{1p} \\ \theta_{21} & \cdots & \theta_{2p} \\ \cdots & \cdots & \cdots \\ \theta_{m1} & \cdots & \theta_{mp} \end{pmatrix}.$$

Вектор ошибки аппроксимации можно записать в виде

$$\mathbf{d}(t) = \hat{\mathbf{y}}(t) - \mathbf{y}(t) = \boldsymbol{\Theta}^T \varphi(t) - \mathbf{y}(t).$$

Его норма определяется выражением

$$\|\mathbf{d}(t)\| = \left\langle \boldsymbol{\Theta}^T \varphi(t) - \mathbf{y}(t), \boldsymbol{\Theta}^T \varphi(t) - \mathbf{y}(t) \right\rangle^{1/2}.$$

Отсюда следует, что задача аппроксимации заключается в нахождении матрицы $\hat{\mathbf{\Theta}}$ коэффициентов, обеспечивающей минимальное значение нормы ошибки на интервале T. Используя интеграл Стилтьеса, усредненное на указанном интервале значение ошибки можно записать в виде выражения, справедливого как для непрерывного, так и для дискретного сигнала:

$$g_T(\mathbf{\Theta}) = \int_T \left\langle \mathbf{\Theta}^T \varphi(t) - \mathbf{y}(t), \mathbf{\Theta}^T \varphi(t) - \mathbf{y}(t) \right\rangle \mu(dt).$$

Для дискретного сигнала это выражение принимает вид

$$g_T(\mathbf{\Theta}) = \sum_{t_k \in T} \left\langle \mathbf{\Theta}^T \varphi(t_k) - \mathbf{y}(t_k), \mathbf{\Theta}^T \varphi(t_k) - \mathbf{y}(t_k) \right\rangle. \qquad (2)$$

Тогда оценку матрицы параметров можно получить в результате минимизации функции (2):

$$\hat{\mathbf{\Theta}} = \arg\min_{\mathbf{\Theta}} g_T(\mathbf{\Theta}).$$

Следует отметить, что в принципе возможна и минимаксная оценка параметров:

$$\hat{\mathbf{\Theta}} = \min_{\mathbf{\Theta}} \min_{t \in T} \left\| \mathbf{d}(t) \right\|.$$

В целях обеспечения робастности задачи минимизации, то есть для снижения чувствительности решений к большим ошибкам измерения, рекомендуется применение квазилинейной регрессии

$$g_\psi(\hat{x}(t)) = \sum_{j=0}^{r} \sum_{k=0}^{p} \psi_j \left(\mathbf{\theta}^T \varphi^{(j)}(t_k) - y^{(j)}(t_k) \right),$$

где в качестве функций $\psi_j(x)$ могут выступать функции Хьюбера, Мешалкина и Демиденко [7], определяемые соответственно как

$$\psi_H(x) = \begin{cases} x^2, |x| \le c \\ 2c|x| - c^2, |x| > c \end{cases},$$

$$\psi_M(x) = \frac{1}{\lambda}\left(1 - e^{-\lambda x^2/2}\right), \lambda > 0,$$

$$\psi_D(x) = \frac{x^2}{x^2 + c}, c > 0.$$

Оценивание параметров функций регрессии осуществляется фактически методами параметрической оптимизации [7; 8]. Похожим способом в ряде случаев можно решить и задачу идентификации оператора $A(\mathbf{\theta})$ системы. Пусть, например, наблюдаемые данные $y(t)$ описываются выражением

$$y(t) = A(\mathbf{\theta}) x(t) + n(t).$$

Здесь, как и выше, шумы измерения обозначены как $n(t)$, а $\mathbf{\theta}$ – вектор оцениваемых параметров.

Критерий качества можно задать выражением

$$g_T(\mathbf{\theta}) = \int_T \left\langle A(\mathbf{\theta}) x(t) - y(t), A(\mathbf{\theta}) x(t) - y(t) \right\rangle \mu(dt).$$

При этом параметры находятся в результате его минимизации. В случае систем с переменными параметрами оператор $A(\mathbf{\theta})$ может зависеть от ряда функций $\mathbf{\theta}_k^T \varphi(t)$, то есть иметь вид $A(\mathbf{\theta}) = A\left(\mathbf{\theta}_k^T \varphi(t), k = 1, \ldots, q\right)$.

Методы регуляризации при решении некорректных задач восстановления зависимостей линейной регрессии

По своему определению функции регрессии – это функции непрерывного времени, независимо от характера наблюдаемых данных. Если последние имеют дискретный характер, то на интервалах между выборками поведение функции регрессии может оказаться неконтролируемым. Решение этой проблемы достигается применением соответствующего регуляризирующего функционала, определенного на всей временной оси. Регуляризирующими свойствами обладают, например, частотные и интегральные характеристики функций на соответствующих конечных интервалах. Так, регуляризацию можно обеспечить ограничениями на энергетический спектр восстановленного сигнала в области высоких частот и форму сигнала в целом.

Если преобразование Фурье взвешенной функции $x(t)$ с весом $w(t)=\begin{cases}\alpha_1, t\in[T_1,T_2]\\ \alpha_2, t\notin[T_1,T_2]\end{cases}$,
где $0\le\alpha_1\le\alpha_2\le 1$, обозначить как

$$\tilde{x}_w(\omega)=\int_{-\infty}^{\infty} w(t)x(t)e^{-j\omega t}dt,$$

то соответствующий функционал в частотной области можно записать в виде

$$G_W(\tilde{x}_w(\omega))=\int_{-\infty}^{\infty} W(\omega)\left|\tilde{x}_w(\omega)\right|^2 d\omega,$$

где весовая функция в частотной области удовлетворяет условию

$$G_W(\tilde{x}_w(\omega))=\int_{-\infty}^{\infty} W(\omega)\left|\tilde{x}_w(\omega)\right|^2 d\omega.$$

При аппроксимации с помощью всплесков в качестве ограничивающей функции можно воспользоваться высокочастотной частью разложения

$$\sum_{n=j}^{\infty}\sum_{k=-\infty}^{\infty}\langle x,\psi_{nk}\rangle\psi_{nk}(t),$$

записав эту функцию в виде

$$g_w=\int_{-\infty}^{\infty} w(t)\left|\sum_{n=j}^{\infty}\sum_{k=-\infty}^{\infty}\langle x,\psi_{nk}\rangle\psi_{nk}(t)\right|^2 dt.$$

Регуляризацию можно обеспечить и с помощью ограничений, налагаемых на интегро-дифференциальные характеристики функций регрессии в пределах соответствующих интервалов:

$$H(x(t))=\int_a^b\left(\left|d_0x(t)\right|^2+\sum_{j=1}^{r}\left|d_jx^{(j)}(t)\right|^2\right)dt,$$

где d_j – некоторые коэффициенты. Подобное выражение используется, например, при оценке качества систем автоматического регулирования.

Смысл этого ограничения можно рассмотреть на примере функционала

$$H(x(t))=\int_a^b\left(d_0^2x(t)^2+x(t)\right)dt.$$

Функция $x(t)$, обеспечивающая экстремальное значение функционала (3), является решением дифференциального уравнения $\ddot{x}(t)=d_0^2x(t)$, то есть решением уравнения Эйлера экстремальной задачи. Его общее решение имеет вид $x(t)=c_1e^{-d_0t}+c_2e^{d_0t}$. Если потребовать, чтобы оно в точках a и b удовлетворяло заданным граничным условиям, то искомое частное решение

$$\hat{x}(t)=\frac{1}{D}\left(x(a)e^{d_0b}-x(b)e^{d_0a}\right)e^{-d_0t}+\frac{1}{D}\left(-x(a)e^{-d_0b}+x(b)e^{-d_0a}\right)e^{d_0t},$$

где $D=e^{d_0(b-a)}-e^{-d_0(b-a)}$.

Принимая на этой функции минимальное значение, рассматриваемый функционал является характеристикой, отражающей близость функций к кривой экспоненциального типа, проходящей через заданные граничные точки. При стремлении коэффициента d_0 к нулю эта кривая превращается в прямую линию, проходящую через точки $x(a)$ и $x(b)$,

$$\hat{x}(t)=x(a)\frac{b-t}{b-a}+x(b)\frac{t-a}{b-a},$$

а функционал, определенный на произвольной функции $x(t)$, является мерой ее отклонения от этой линии. Аналогично функционалам общего вида соответствуют дифференциальные уравнения более высокого порядка, решения которых можно считать эталонами. Все это обусловливает возможность использования подобных функционалов при аппроксимации в целях придания результирующим функциям более регулярной формы.

Таким образом, существует множество различных способов регуляризации, приводящих к минимизации функционала вида

$$\hat{\boldsymbol{\theta}} = arg\,\underset{\boldsymbol{\theta}}{min}\left(Q(\boldsymbol{\theta}) + \varepsilon\Omega(x(t),\boldsymbol{\theta})\right),$$

что требует, как правило, применения численных методов оптимизации [9].

Техника регуляризации широко используется при решении и других некорректных задач, например задач восстановления сигналов [9; 10], связанных с решением интегральных уравнений

$$\int_{-\infty}^{\infty} h(t-\tau)x(\tau)d\tau + n(t) = y(t),$$

где $y(t)$ и $n(t)$ – наблюдаемый процесс и шум измерения соответственно; $x(t)$ – искомый процесс; $h(t)$ – импульсная функция (ядро уравнения). Задача в этом случае сводится к минимизации функционала

$$\left\|A\cdot x - y\right\|^2, \tag{4}$$

где $A\cdot x \equiv \int_{-\infty}^{\infty} h(t-\tau)x(\tau)d\tau$.

При дополнении функционала (4) регуляризирующим функционалом $\Omega(x)$ решение задачи сводится к минимизации функционала следующего вида:

$$\hat{x}(t) = arg\,\underset{x}{min}(\left\|A\cdot x - y\right\|^2 + \alpha\,\Omega(x)).$$

Восстановление зависимостей методами линейной регрессии также относится к классу некорректных задач [10]. Некорректность линейной регрессии объясняется возможным ростом ошибки в непрерывной метрике [3], при том что в дискретной метрике ошибка стремится к нулю.

Заключение

Особенность обобщения, основанного на методе регрессии и наименьших квадратов, заключается в реализации операторов, отображающих дискретные данные в пространстве непрерывных или кусочно-непрерывных функций.

Полученные результаты по разработке методов и алгоритмов, ориентированных на восстановление по дискретным данным в целом гладких функций, открывают возможности по решению задач локальной обработки и сглаживания как стационарных, так и нестационарных сигналов, задач анализа акустических сигналов и динамики речи, задач анализа многоэкстремальных зависимостей.

СПИСОК ЛИТЕРАТУРЫ

1. Вапник, В.Н. Восстановление зависимостей по эмпирическим данным / В.Н. Вапник. - М.: Наука, 1979. - 448 с.
2. Гроп, Д. Методы идентификации систем / Д. Гроп. - М.: Мир, 1979. - 304 с.
3. Демиденко, Е.З. Оптимизация и регрессия / Е.З. Демиденко. - М.: Наука, 1989. - 296 с.
4. Быков, А.А. Модель закона распределения вероятности амплитуд сигналов в базисе экспоненциальных функций системы / А.А. Быков, Ю.А. Кропотов // Проектирование и технология электронных средств. - 2007. - № 2. - С. 30-34.
5. Кропотов, Ю.А. Моделирование и методы исследований акустических сигналов, шумов и помех в системах телекоммуникаций: монография / Ю.А. Кропотов, В.А. Ермолаев. - М.-Берлин: Директ-Медиа, 2016. - 251 с.
6. Граничин, О.Н. Оценивание параметров линейной регрессии при произвольных помехах / О.Н. Граничин // Автоматика и телемеханика. - 2002. - № 1. - С. 30-41.
7. Кропотов, Ю.А. Алгоритм определения параметров экспоненциальной аппроксимации закона распределения вероятности амплитуд речевого сигнала / Ю.А. Кропотов // Радиотехника. - 2007. - № 6. - С. 44-47.
8. Кропотов, Ю.А. Методы оптимизации в задачах локальной аппроксимации сигналов / Ю.А. Кропотов, В.А. Ермолаев // В мире научных открытий. - 2010. - № 6.1. - С. 44-47.
9. Кропотов, Ю.А. Вопросы параметрического представления нестационарных сигналов / Ю.А. Кропотов, В.А. Ермолаев // Проектирование и технология электронных средств. - 2010. - № 1. - С. 31-35.
10. Морозов, В.А. Регулярные методы решения некорректно поставленных задач / В.А. Морозов. - М.: Наука, 1987. - 240 с.

1. Vapnik, V.N. Вапник, В.Н. *Dependences Recovery on Empirical Data* / V.N. Vapnik. – M.: Science, 1979. – pp. 448.
2. Grop, D. *Methods for Systems Identification* / D. Grop. – M.: World. 1979. – pp. 304.
3. Demidenko, E.Z. *Optimization and Regression* / E.Z. Demidenko. – M.: Science, 1989. – pp. 296.
4. Bykov, A.A. Model of signal amplitude probability distribution law in basis of system exponentials / A.A. Bykov, Yu.A. Kropotov // *Design and Techniques of Electronic Means*. – 2007. – No.2 – pp. 30-34.
5. Kropotov, Yu.A. *Modeling and Methods of Investigations of Acoustic Signals, Noises and Interference in TV Communication Systems:* monograph / Yu.A. Kropotov, V.A. Yerolayev. – M.: Berlin: Direct-Media, 2016. – pp. 251.
6. Granichin, O.N. Assessment of linear regression parameters at arbitrary interference / O.N. Granichin // *Automation and Telemechanics*. – 2002. – No.1. – pp. 30-41.
7. Kropotov, Yu.A. Algorithm for exponential approximation parameters definition of speech signal amplitude probability distribution law / Yu.A. Kropotov // *Radio Engineering*. – 2007. – No.6. – pp. 44-47.
8. Kropotov, Yu.A. Methods for optimization in problems of local signal approximation / Yu.A. Kropotov, V.A. Yerolayev // *In the World of Scientific Discoveries*. – 2010. – No.6.1. – pp. 44-47.
9. Kropotov, Yu.A. Problems in parametric presentation of non-stationary signal / Yu.A. Kropotov, V.A. Yermolayev // *Design and Techniques of Electronic Means*. – 2010. – No.1. – pp. 31-35.
10. Morozov, V.A. *Regular Methods for Solution of Ill-posed Problems* / V.A. Morozov. – M.: Science, 1987. – pp. 240.

Статья поступила в редколлегию 11.12.17.
Рецензент: д.т.н., профессор
Орлов А.А.

Сведения об авторах:

Кропотов Юрий Анатольевич, д.т.н., профессор, зав. кафедрой ЭиВТ Муромского института (филиала) Владимирского государственного университета им. А.Г. и Н.Г. Столетовых, тел. (49234) 77-2-72, e-mail: kaf-eivt@yandex.ru.

Холкина Наталья Евгеньевна, доцент кафедры ЭиВТ Муромского института (филиала) Владимирского государственного университета им. А.Г. и Н.Г. Столетовых, тел. (49234) 77-2-72 **,** e-mail: kaf-eivt@yandex.ru.

Kropotov Yury Anatolievich, D. Eng., Prof., Head of the Dep. of E&CT, Murom Institute (Branch) of Stoletovs State University of Vladimir, e-mail: kaf-eivt@yandex.ru.

Kholkina Natalia Yevgenievna, Assistant Prof. of the Dep. of E&CT, Murom Institute (Branch) of Stoletovs State University of Vladimir, e-mail: kaf-eivt@yandex.ru.