УДК 519.2

# ТЕКУЩЕЕ ОЦЕНИВАНИЕ ПРОИЗВОДНОЙ НЕСТАЦИОНАРНОГО ПРОЦЕССА НА ОСНОВЕ РЕКУРРЕНТНОГО СГЛАЖИВАЮЩЕГО СПЛАЙНА

**Е. А. Кочегурова, Е. С. Горохова**

*Томский политехнический университет,*
*634050, г. Томск, просп. Ленина, 30*
*E-mail: kocheg@mail.ru*

Представлен метод восстановления полезного сигнала и его низших производных в реальном масштабе времени на основе рекуррентного сглаживающего сплайна. Приведена расчётная схема сплайна, у которого число измерений каждого звена больше числа узлов, и с помощью вариационного подхода найдены его коэффициенты.

*Ключевые слова:* восстановление функции и её производных, рекуррентный алгоритм, вариационный сглаживающий сплайн.

DOI: 10.15372/AUT20160310

**Введение и постановка задачи.** Одной из форм представления исходных данных в системах мониторинга, управления и принятия решений являются временны́е ряды. При этом основная информация о наблюдаемом объекте часто формируется самим объектом и регистрируется измерительными приборами на некотором временно́м интервале. Наличие помех, искажающих параметры объекта, а также нестационарное поведение самого объекта приводят к тому, что наблюдаемый процесс становится нестационарным.

При обработке нестационарных случайных процессов возникает ряд задач, одна из которых — косвенная оценка производных на основе измерительной информации. Эта задача наиболее актуальна в таких прикладных областях, как управление объектами, мониторинг, принятие решений, прогноз показателей и поведения объектов и систем.

При этом особый интерес представляют методы восстановления полезного сигнала и его низших производных в реальном масштабе времени (РМВ).

В предлагаемой работе рассматривается класс алгоритмов, используемых в измерительных устройствах с дискретным входом и аналоговым выходом, где в моменты времени $t_i$, $i = 1, 2, \ldots$, регистрируется последовательность отсчётов $y(t_i)$ со случайной помехой $\xi(t_i)$:

$$\begin{gathered} y(t_i) = f(t_i) + \xi(t_i); \\ M\{\xi(t)\} = 0; \quad M\{\xi^2(t)\} = \sigma_\xi^2; \quad M\{f(t), \xi(t)\} = 0. \end{gathered} \tag{1}$$

Основу подобных устройств составляет преобразователь, аппроксимирующий входные отсчёты в аналоговый выходной сигнал.

**Подходы к решению задачи.** Сглаживание временны́х рядов — один из наиболее мощных инструментов их изучения. Сглаженный процесс можно рассматривать как идеальный вариант его поведения. Весьма перспективна при оценивании производных аппроксимация сглаженного ряда в виде аналитической функции. Полученная аналитическая функция, отражающая основные свойства наблюдаемого процесса, позволяет оценить производные заданного порядка. Наиболее известными для сглаживания являются методы регрессионного анализа, ортогональные полиномы, сплайны, вейвлет- и фурье-функции, а также другие методы частотно-временнóго анализа. Основные классы методов аппроксимации экспериментальных данных проанализированы в [1].

Нередко для восстановления зашумлённого сигнала применяются методы, базирующиеся на использовании сплайн-функций [2, 3], в сочетании со спектральными или генетическими методами.

Наиболее разработанный аппарат интерполяционных сплайнов связан с решением систем алгебраических уравнений, что приводит к значительным вычислительным затратам. Кроме того, интерполяционный сплайн не может быть использован для сглаживания при наличии значительных шумов в результатах измерений.

Известно, что решение задачи дифференцирования прямыми методами неустойчиво, так как скорость изменения функции существенно усиливает резкие флуктуации исходной функции, зарегистрированной с погрешностями. Для дискретных данных это проявляется в усилении погрешности входных данных. Одним из возможных подходов, приводящих к устойчивым результатам, является параметризация решений в классе сглаживающих сплайнов.

Применение сплайна, обеспечивающего минимальную норму производной, позволяет получить искомое гладкое решение [4–7]. Сглаживающий сплайн $S(t)$ основан на оптимизации специального вида функционала

$$J(S) = \lambda \int\limits_a^b [S''(t)]^2 dt + \sum_{i=0}^{n} [S(t_i) - y(t_i)]^2, \tag{2}$$

где первое слагаемое совместно со сглаживающим параметром $\lambda$ определяет штраф кривизны; второе слагаемое минимизирует остаточные суммы квадратов и соответствует классическому методу наименьших квадратов.

Получение производной по результатам наблюдений может быть интерпретацией прямых [8, 9] или косвенных [10] измерений. В последнем случае искомая производная является решением интегрального уравнения 1-го рода и связана с использованием метода регуляризации Тихонова.

Интерпретация прямых измерений основана на дифференцировании аппроксимирующей функции. Полиномиальный базис высокой степени обеспечивает вычисление $k$-й производной. Наличие у сплайна группы параметров: степень, дефект, узлы — позволяет варьировать точностные и временны́е характеристики сплайна. В то же время форма представления и способ получения сглаживающего сплайна во многом определяют возможные режимы оценивания производной. Для оценивания параметров наблюдаемого процесса в РМВ целесообразно формировать рекуррентные расчётные схемы. Метод последовательного сплайн-сглаживания [11, 12] не требует решения систем уравнений, однако обработка также ведётся апостериорно, а количество последовательных сглаживаний является параметром регуляризации.

В данной работе приведена расчётная схема рекуррентного сглаживающего сплайна (РСС), у которого число измерений каждого звена больше числа узлов. Рекуррентный сглаживающий сплайн $S(t)$ найден на основе оптимизации экстремального функционала, который для предлагаемого сплайна модифицирован и имеет следующий вид [13]:

$$J(S) = (1-\rho)(h\Delta t)^2 \int\limits_{t_0^i}^{t_h^i} [S''(t)]^2 dt + \rho \sum_{j=0}^{h} [S(t_j^i) - y(t_j^i)]^2, \tag{3}$$

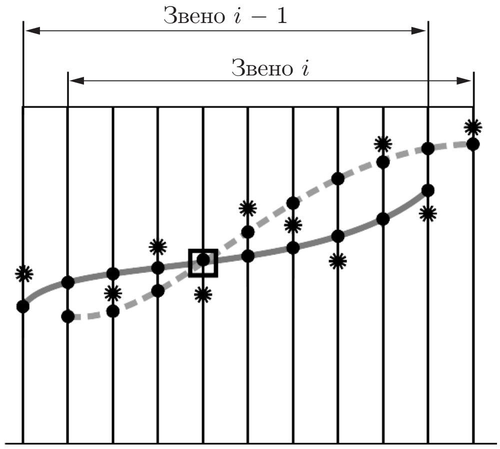


*Рис. 1.* Схема текущего режима функционирования РСС: $*$ — наблюдаемый процесс, $\bullet$ — вычисленные значения $\tau$, $\square$ — момент сопряжения звеньев РСС $q$

где $\rho$ — весовой коэффициент, устанавливающий баланс между сглаживающими и интерполяционными свойствами сплайна $S(t)$, $\rho \in [0,1]$; $\Delta t$ — интервал дискретизации наблюдаемого процесса; $t_0^i$, $t_h^i$ — номера начальной и конечной абсцисс $i$-го звена сплайна соответственно; $h = (t_h^i - t_0^i)/\Delta t$ — количество измерений внутри $i$-го звена сплайна, далее $h = \text{const}$ для всех звеньев сплайна на интервале наблюдения процесса; $y(t_i)$ — измерения, определённые формулой (1).

Для $i$-го звена анализируемый кубический сплайн $S_i(\tau) = a_0^i + a_1^i\tau + a_2^i\tau^2 + a_3^i\tau^3$ может быть вычислен в любой внутренней точке звена $\tau \in [1,h]$.

Условия сопряжения звеньев сплайна определяют формулы для его непрерывных коэффициентов. Разрывные коэффициенты сплайна находятся из условия минимума экстремального функционала (3) в зависимости от дефекта сплайна $d$. Для $d = 2$ разрывными являются коэффициенты $a_2^i$ и $a_3^i$, для $d = 1$ — коэффициент $a_3^i$.

В зависимости от соотношения моментов сопряжения $q \in [0, h-1]$ и вычисления $\tau \in [1,h]$ $i$-го сплайна выделяют несколько режимов функционирования РСС (рис. 1). Например, текущий режим оценивания, в котором звено сплайна вычисляется при поступлении очередного измерения с использованием предыдущих $h-1$ измерений. Сопряжение непрерывных производных в этом режиме осуществляется в момент времени $t_{q+1}^{i-1} = t_q^i$.

Формулы для оценки коэффициентов в таком режиме для $q = 0$ и $d = 2$ принимают наиболее простой вид:

$$\begin{aligned}
&a_0^i = a_0^{i-1} + a_1^{i-1} + a_2^{i-1} + a_3^{i-1}; && A = 6(1-\rho)h^4 + \rho H_5;\\
&a_1^i = a_1^{i-1} + 2a_2^{i-1} + 3a_3^{i-1}; && B = 4(1-\rho)h^3 + \rho H_4;\\
&a_2^i = \frac{\rho(F_1^i C - F_2^i A)}{BC - A^2}; && C = 12(1-\rho)h^5 + \rho H_6;\\
&a_3^i = \frac{\rho(F_2^i B - F_1^i A)}{BC - A^2}; && F_1^i = \sum_{k=0}^{h} y(t_k^i)k^2 - a_0^i H_2 - a_1^i H_3;\\
& && F_2^i = \sum_{k=0}^{h} y(t_k^i)k^3 - a_0^i H_3 - a_1^i H_4;\\
& && H_n = \sum_{k=0}^{h} k^n.
\end{aligned} \qquad (4)$$

Коэффициенты $a_0^i, a_1^i$ найдены из условия сопряжения звеньев $S^{(k)}(t_{q+1}^{i-1})_+ = S^{(k)}(t_q^i)_-$, $k = 0, 1$, а коэффициенты $a_2^i, a_3^i$ — из условий минимизации (3) $\frac{\partial J(S)}{\partial a_2^i} = 0$, $\frac{\partial J(S)}{\partial a_3^i} = 0$.

Достоинство данного подхода — наличие явных формул вычисления коэффициентов сплайна, что исключает использование численных алгоритмов и сокращает время расчётов при оценке производных в РМВ.

**Результаты вычислительного эксперимента.** Качество оценивания производных в большей степени определяют настроечные параметры РСС. К таким параметрам сплайна относятся: $\rho$ — балансовый множитель, $h$ — число измерений в звене сплайна. Качество восстановления функции и её производных при наличии помехи с заданной дисперсией $\sigma_\xi^2$ оценивалось по критерию MSE (Mean Squared Error) [14], который для наглядности выражен в процентах MSPE (Mean Squared Percentage Error):

$$\mathrm{MSE}(\widehat{x}_j(t)) = \sqrt{\frac{1}{K}\sum_{j=1}^{K}\frac{1}{n}\sum_{i=1}^{n}(x(t_i) - \widehat{x}_j(t_i))^2};$$
$$\mathrm{MSPE} = \frac{\mathrm{MSE}}{|x_{\max} - x_{\min}|}\cdot 100\ \%, \tag{5}$$

где $\widehat{x}_j(t_i)$ — оценка функции (производной) в момент времени $t_i$ в $j$-м эксперименте; $x(t_i)$ — истинное значение функции (производной) в момент $t_i$; $K$ — количество экспериментов; $n$ — число значений функции; $|x_{\max} - x_{\min}|$ — диапазон изменения тестовой функции (производной). При этом для сопоставления с результатами оценивания производной, приведёнными в [15], объёмы вычислительного эксперимента также были выбраны аналогичными значениям $K = 200$, $n = 100$.

Оптимальные значения настроечных параметров РСС находились на основе вычислительного эксперимента с использованием тестовых функций

$$f_1(t) = 10\sin\left(\frac{2\pi i}{100}\right), \qquad f_2(t) = \sin\left(\frac{\pi i}{20}\right)\mathrm{e}^{0{,}02t} + 3,$$

графики которых изображены на рис. 2, $a$ и $c$ соответственно.

Кроме полезного сигнала, обозначенного на рисунках точками, красными линиями показаны восстановленные функции (см. рис. 2, $a, c$) и восстановленные производные (рис. 2, $b, d$). Результаты приведены для частного случая $h = 10$ и $\sigma_\xi = 10$ % от максимального значения полезного сигнала. Из графиков видно, что восстановление производной визуально вполне удовлетворительно. Однако при близкой погрешности восстановления (менее 2 %) производная на рис. 2, $b$ более гладкая, чем на рис. 2, $d$. Гладкость восстановления производной регулируется параметром $\rho$. При этом её повышение приводит к возникновению систематического отклонения, которое также является нежелательным эффектом восстановления.

Проведённые исследования позволили оценить MSPE-погрешности восстановления функции $\sigma_0$ и её производной $\sigma_1$ в зависимости от величины шума. На рис. 3 представлены погрешности восстановления в диапазоне изменения шума $\sigma_\xi \in 0$–$20$ %. Погрешность оценки производных (синие линии) несколько выше, чем погрешность восстановления самой функции (красные линии), при этом значения погрешностей полностью зависят от вида и свойств исходной функции.

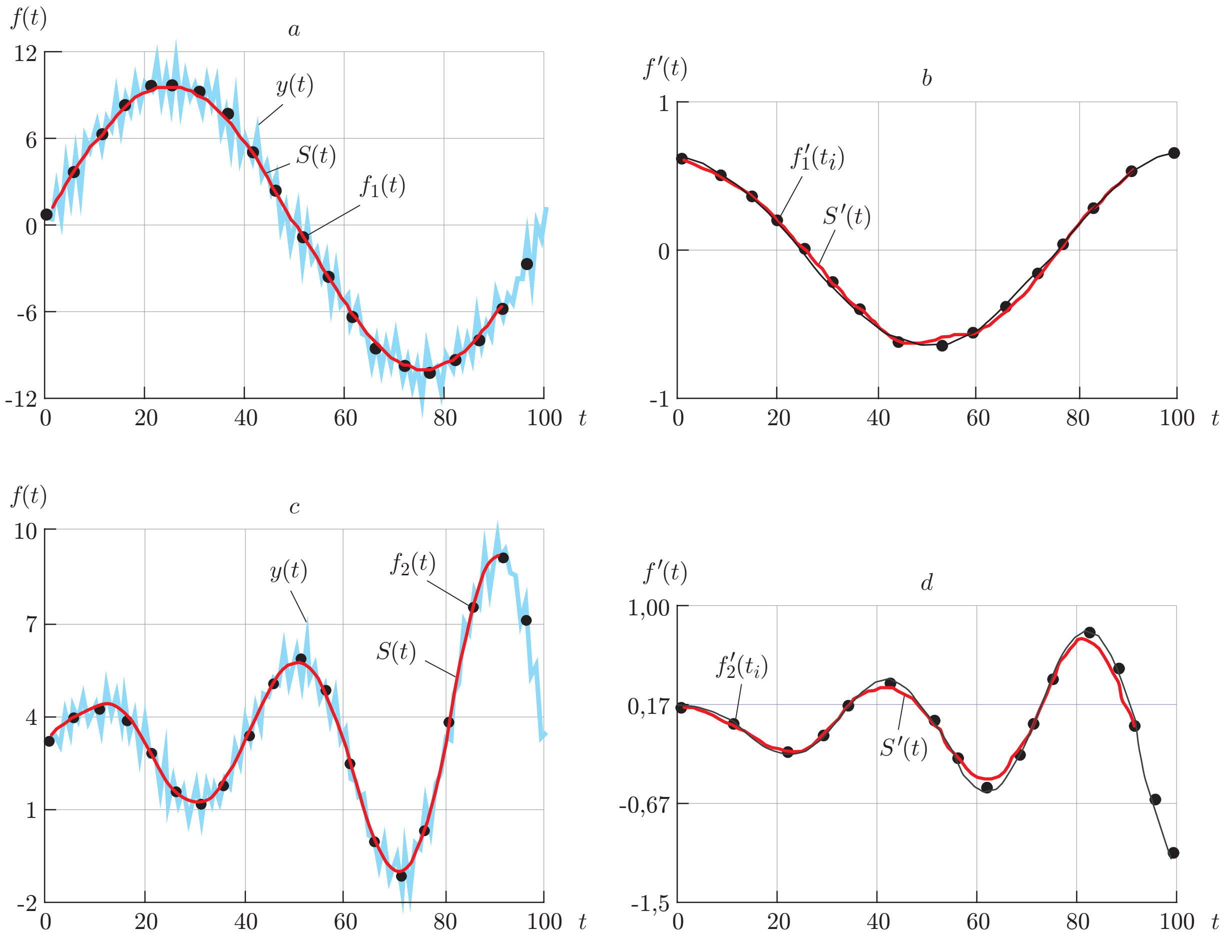


*Рис. 2.* Тестовые функции и восстановленные производные

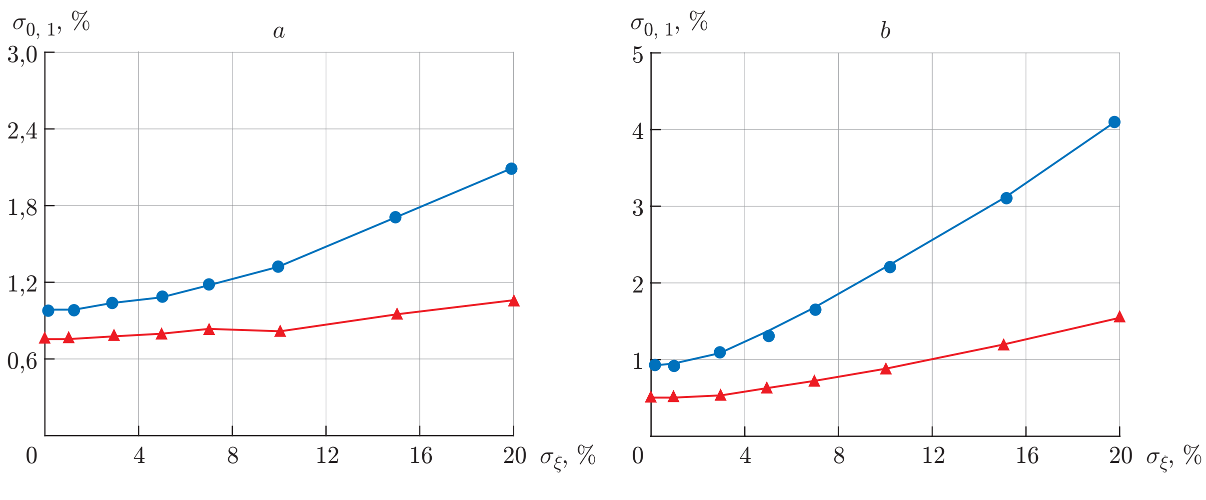


*Рис. 3.* Зависимости погрешностей восстановления функции $\sigma_0$ и её производной $\sigma_1$ от погрешности шума $\sigma_\xi$: $a$ — $f_1(t)$ и $b$ — $f_2(t)$

| Метод | $\sigma_\xi = 3$ % | $\sigma_\xi = 7$ % | $\sigma_\xi = 10$ % | Примечание |
|---|---|---|---|---|
| P-сплайн | 0,093<br>0,442 % | 0,204<br>0,972 % | 0,273<br>1,300 % | MSE-погрешность [15]<br>MSPE-погрешность |
| РСС | 1,09 % | 1,688 % | 2,218 % | $f_1(t)$, $h$, $\rho_{\text{opt}}$, $\tau = 1$ |
| РСС | 1,042 % | 1,176 % | 1,334 % | $f_2(t)$, $h$, $\rho_{\text{opt}}$, $\tau = 1$ |

Значения погрешностей приведены на рисунке для оптимальных значений настраиваемых параметров $h$ и $\rho$. Во всём оцениваемом интервале наблюдается заметное уменьшение погрешности исходных данных: MSPE-погрешность восстановления производной сокращается от 3 (при $\sigma_\xi < 10$ %) до 10 раз (при $\sigma_\xi > 15$ %).

С ростом погрешности шума $\sigma_\xi$ увеличивается и оптимальная длина сплайна $h$: от $h_{\text{opt}} = 8$–10 измерений при $\sigma_\xi = 1$ % до $h_{\text{opt}} = 14$–19 измерений при $\sigma_\xi = 20$ %, при этом $\rho_{\text{opt}} \in [0, 1]$.

В заключение сопоставим MSPE-погрешность восстановления производной предлагаемым методом (РСС) с результатами использования P-сплайна (Parametric Penalized Spline Smoothing), приведёнными в [15].

Из таблицы видно, что P-сплайн обеспечивает меньшую погрешность, но именно РСС осуществляет восстановление функции и её производной в реальном масштабе времени.

**Заключение.** В данной работе описана регрессионная модель восстановления функции и её производной, основанная на вариационном подходе. Главное требование, предъявляемое к сглаживающей функции, — возможность её использования в РМВ. Для этого была разработана рекуррентная схема расчёта, содержащая явные формулы. Сравнительный анализ предлагаемого РСС и P-сплайна показал, что погрешности восстановления РСС хотя и несколько выше (в среднем в 1,5 раза), но вполне применимы в инженерных приложениях. Следует отметить, что этот сплайн допускает реализацию в РМВ, а оценки таких процедур лишь асимптотически стремятся к оценкам апостериорных алгоритмов, каковым и является P-сплайн.

## СПИСОК ЛИТЕРАТУРЫ


1. **Голубинский А. Н.** Методы аппроксимации экспериментальных данных и построения моделей // Вестн. Воронежского института МВД России. 2007. № 2. С. 138–143.
2. **Завьялов Ю. С., Квасов Б. И., Мирошниченко В. Л.** Методы сплайн-функций. М.: Наука, 1980. 382 с.
3. **Вершинин В. В., Завьялов Ю. С., Павлов Н. Н.** Экстремальные свойства сплайнов и задача сглаживания. Новосибирск: Наука, 1988. 102 с.
4. **Роженко А. И.** Теория и алгоритмы вариационной сплайн-аппроксимации: Дис. … д-ра физ.-мат. наук /ИВМиМГ СО РАН. Новосибирск, 2003. 231 с.
5. **Воскобойников Ю. Е., Колкер А. Б.** Аппроксимация изолиний изображений сглаживающими сплайнами // Автометрия. 2003. **39**, № 4. С. 3–12.
6. **Денисов В. И., Фаддеенков А. В.** Сплайновая регрессия с переменными штрафными коэффициентами // Автометрия. 2015. **51**, № 3. С. 3–10.
7. **Хаймович И. Н., Клентак Л. С.** Усовершенствование методов сглаживания сложных поверхностей с использованием интерполяционных сплайнов // Фундаментальные исследования. 2013. № 10 (ч. 12). С. 2634–2638.
8. **De Brabanter K., de Brabanter J., de Moor B.** Derivative estimation with local polynomial fitting // Journ. Mach. Learning Res. 2013. **14**, N 1. P. 281–301.

9. **Жиляков Е. Г., Черноморец А. А., Паболкова Н. С.** Дифференцирование сигналов по дискретным отсчётам на основе частотных представлений // Научные ведомости Белгородского государственного университета. 2010. № 19. С. 88–92.
10. **Dmitriev V. I., Ingtem Zh. G.** Solving an integral equation of the first kind by spline approximation // Computat. Math. Modeling. 2004. **15**, N 2. P. 99–104.
11. **Гребенников А. И.** Метод последовательных сглаживаний и сплайн-алгоритмы // Численный анализ: методы, алгоритмы, приложения. М.: Изд-во МГУ, 1985. С. 82–95.
12. **Schmidt J. W., Scholz I.** A dual algorithm for convex-concave data smoothing by cubic $C^2$-splines // Numer. Math. 1990. **57**. P. 330–350.
13. **Кочегурова Е. А., Шебеко Е. В.** Использование вариационного сглаживающего сплайна в задаче краткосрочного прогнозирования // Изв. ТПУ. 2006. **309**, № 7. С. 36–39.
14. **Низамитдинов А. И.** Использование штрафного сплайна (P-spline) и кубического сглаживающего сплайна в прогнозировании временных рядов // Вестн. Таджикского государственного университета права, бизнеса и политики. Сер. Гуманитарные науки. 2012. № 4(52). С. 124–131.
15. **Cao J., Cai J., Wang L.** Estimating curves and derivatives with parametric penalized spline smoothing // Stat. Comput. 2012. **22**, N 5. P. 1059–1067.

*Поступила в редакцию 2 декабря 2015 г.*