<Оптика атмосферы и океана>, **7**, №9 (1994)

ОБРАТНЫЕ ЗАДАЧИ ОПТИКИ АТМОСФЕРЫ И ОКЕАНА

УДК 519.24,551

Ю.Ф. Демьянов

УСТОЙЧИВЫЙ АЛГОРИТМ ВОССТАНОВЛЕНИЯ ФОРМЫ СИГНАЛА ИМПУЛЬСНОГО ЛИДАРА

Представлен метод решения задачи восстановления формы сигнала импульсного лидара в условиях, когда закон распределения ошибок измерения отличен от нормального. Предложены устойчивый <выделяющий> алгоритм и результаты его исследований методом Монте-Карло. Приведена теорема, позволяющая для частного случая оценки параметра сдвига оценить оптимальные свойства алгоритма.

При дистанционном определении различных параметров атмосферы с помощью импульсного лидара важной является информация о форме отраженного лазерного импульса. В частности, информация об искажении регулярной составляющей зондирующего импульса позволяет определить концентрацию и спектр частиц атмосферного аэрозоля, объемные коэффициенты рассеяния облаков и некоторые другие параметры [1]. Однако стандартные методы обработки отраженных сигналов, базирующиеся на идеях линейной фильтрации, эффективны лишь при соблюдении явных или неявных предположений о форме (в частности, о нормальном виде) законов распределения ошибок измерений. На практике, в силу влияния помех различного происхождеия, например, шумов видеодатчика, ошибок каналов цифрового преобразования и передачи информации, и других причин эти законы отличаются от нормального.

Малый объем информации, а также сам характер статистических критериев проверки на адекватность законов распределения и ненулевая вероятность присутствия грубых ошибок (<выбросов>) не позволяют достаточно надежно установить вид закона ее распределения. Вследствие этого применение стандартных методов может привести к существенному искажению результатов интерпретации измерений формы сигнала [2–4], и возникает проблема разработки таких методов обработки измерений, которые были бы малочувствительны (устойчивы) к виду закона распределения и, в частности, не так сильно реагировали бы на <выбросы>.

Будем полагать, что участки наблюдения ограничены по длительности, а регулярная составляющая сигнала ограничена по первой и второй производным. В таком случае, поскольку любая непрерывная на замкнутом отрезке функция сколь угодно точно равномерно приближается полиномом (теорема Вейерштрасса), регулярную составляющую сигнала можно считать полиномиальной, а в качестве стохастической математической модели решения задачи выбрать линейную регрессию на полином.

Пусть описывающий регулярную составляющую сигнала полином

$$F(t) = C_0 + C_1 t + \ldots + C_{p-1} t^{p-1},$$

где $p - 1$ – степень полинома (считается известной); $C_{p-1} \neq 0$, $0 \leq t \leq T$ (T – длительность сигнала). Необходимо определить неизвестные коэффициенты C_j, $j = 1,2,\ldots, p - 1$ по результатам b_i измерений в моменты времени t_i, $i = 1,2, \ldots , n$. Видно [5], что поставленная задача сводится к трем математическим моделям в виде систем линейных алгебраических уравнений, через решение которых определяются искомые коэффициенты.

1. Стохастическая модель. Для этого случая коэффициенты регрессии определяются как результат решения стохастической системы линейных алгебраических уравнений (СЛАУ) вида

$$A\, x = b, \qquad (1)$$

где $A = (a_{ij})$ – $n{\times}p$ – матрица; $x = (x_j)$ – $1{\times}p$ – столбец; $b = (b_i)$ – $1{\times}n$ – столбец. Элементы a_{ij} и b_i системы – случайные переменные с неизвестным законом распределения. Далее всюду полагаем, что $n \geq p$ и

$$\operatorname{rang} A = p \tag{2}$$

почти наверное.

2. Детермированная модель. Это детермированная СЛАУ вида

$$\bar{A}\,\bar{x} = \bar{b}, \tag{3}$$

элементы $(\bar{a}_{ij})$ и $(\bar{b}_i)$ известны точно. Так как измерений без ошибок нет, эта наиболее разработанная теоретически модель не рассматривается.

3. Промежуточная модель. Это стохастическая СЛАУ вида

$$\bar{A}\,x = b, \tag{4}$$

где $\bar{A} = (\bar{a}_{ij})$, $\bar{a}_{ij} = \bar{t}_i^{\,j-1}$, $\bar{t}_i$ – моменты измерений, известны без ошибок, $b_i = \bar{b}_i + \Delta b_i$, причем закон распределения ошибок измерения сигнала Δb_i неизвестен.

Пусть $\bar{b}_{i+1} - \bar{b}_i = \overline{\mathrm{Q}}$ ($i = 1,\dots, n-1$). Если число уравнений велико, то последовательным вычитанием их друг из друга через одно можно получить систему

$$\bar{D}\,y = e, \tag{5}$$

где

$$\bar{D} = (d_{2l-1,j}) = (\bar{a}_{2l,j} - \bar{a}_{2l-1,j});$$

$$l = 1,\dots, [n/2];\quad j = 1,\dots, p-1;\quad y_1 = x_2,\dots, y_{p-1} = x_p;$$

$$e = (e_{2l-1}) = (\bar{\theta} + \Delta e_{2l-1});\quad \Delta e_{2l-1} = \Delta b_{2l} - \Delta b_{2l-1}.$$

Неизвестное x_1 можно найти после неизвестных $x_2,\dots, x_p$ из дополнительного уравнения.

Оценка точности решения СЛАУ предлагаемым ниже алгоритмом проводилась с помощью метода Монте-Карло, а реализованное в ходе модельной оценки выражение для ошибок задания матрицы (правой части) систем (1) и (или) (4) имеет вид

$$\Delta t(\Delta b) = \gamma_{a(b)} + \delta_{a(b)}\,\xi_{a(b)}, \tag{6}$$

где $\gamma_i \in \Phi(0,\sigma_{\mathrm{U}})$, $\delta_i \in B(\varepsilon)$, $\xi_i \in H(0,\sigma_{\mathrm{H}})$ – нормальное, Бернулли и симметричное унимодальное распределения соответственно;

$$E(\Delta t_i) = E(\Delta b_i) = 0;\; E(\Delta t_i \Delta b_j) = E(\Delta t_i)\,E(\Delta b_j);$$

$$E(\Delta t_i \Delta t_j) = E(\Delta t_i)\,E(\Delta t_j);\; E(\Delta b_i \Delta b_j) = E(\Delta b_i)\,E(\Delta b_j),$$

$$i \neq j,\; \varepsilon \in [0;0{,}5],$$

где E – математическое ожидание (среднее значение).

Рассмотрим следующий (названный выделяющим) алгоритм решения задачи (4).

А1. Сформируем правую часть так называемой <базовой> системы – вектор b_{M}. Это вектор, компоненты которого – порядковые статистики $b(r)$, $b(r+1),\dots, b(s)$ вектора b. Величины r, s определяются из уравнения [6]:

$$\pi\left(r, n-r+1, n, \frac{1}{2}\right) = 2\, I_{1/2}(r, n-r+1) - 1 = 2^{-n} \sum_{i=r}^{n-r} C_n^i,$$

где $\pi\left(r, s, n, \frac{1}{2}\right)$ – доверительная вероятность; $s' = n - r + 1$; n – объем выборки b; $I_{1/2}$ – неполная β-функция.

Величины r, s представляют собой фактически непараметрические доверительные границы с уровнем значимости $\alpha_0 = 1 - \pi(\cdot)$ для медианы независимых одинаково распределенных случайных величин, соответствующих выборочным значениям b_i, $i = 1,..., n$.

Производим далее формирование матрицы <базовой> системы

$$\overline{A}_{\text{М}} = \{\overline{a}_{ij}\},$$

где i – индекс строки исходной матрицы $\overline{A}$, соответствующий k-му элементу вектора $b_{\text{М}}$, $j = 1,..., p$; $k = r, r + 1,..., s$.

Предположим, что нам точно известны степень аппроксимирующего полинома p и количество уравнений n.

А2. Производим оценку МНК компонент вектора решения $\hat{x}_{\text{М}}$.

А3. Рассчитываем оценку величины СКО закона распределения модели σ_{Φ}:

$$\hat{S}_{\Phi} = \frac{1}{s - r - p}\,(b_{\text{М}} - \overline{A}_{\text{М}}\,\hat{x}_{\text{М}})^{\text{T}}\,(b_{\text{М}} - \overline{A}_{\text{М}}\,\hat{x}_{\text{М}}),$$

а также $100(1 - \alpha)$-процентного доверительного интервала для b_* – предсказанного значения отклика согласно выражению [7] $\hat{\Delta} b_{*2} = t_{s-r-p}\hat{S}_{\Phi}(\upsilon_* + 1)^{1/2}$, где t_{s-r-p} – верхняя $100(\alpha/2)$-процентная точка распределения t_{s-r-p}, т.е.

$$P\left\{\frac{\overline{b}_* - b_*}{\hat{S}_{\text{F}}(\upsilon_* + 1)^{1/2}} > t^{\text{a}/2}_{s-r-p}\right\} = \alpha/2;$$

$$\upsilon_* = \overline{a}^{\text{T}}_*(\overline{A}_{\text{М}}{}^{\text{T}}\,\overline{A}_{\text{М}})^{-1}\overline{a}_*;\quad \overline{a}_* = (1, \overline{t}_*, \overline{t}^{\,2}_*, \dots, \overline{t}^{\,p}_*),$$

$*$ – значение индекса строки исходной матрицы $\overline{A}$, соответствующего порядковой статистике $b_{(r-i)}$, либо $b_{(s+i)}$ (левый и правый концы интервала берутся поочередно; $i = 1,2,..., r - 1$ для левого конца; $i = 1,2,..., n - s$ – для правого). Величина i изменяется на единицу на шаге 4 алгоритма. Если величина $\sigma_{\Phi,}{}^2$ известна априорно, то рассчитываем

$$\hat{\Delta} b_{*1} = \psi(1 - \alpha/2)\;\sigma_{\Phi}\,(\upsilon_* + 1)^{1/2},$$

где $\psi(1 - \alpha/2) = N^{-1}(1 - \alpha/2;0,1)$.

А4. Производим проверку на принадлежность найденному на шаге 3 доверительному интервалу элемента $b_{(r-i)}$ либо $b_{(s+i)}$. Если ответ положителен, то элемент $b_{(r-i)}b_{(s+i)}$ включается в состав вектора $b_{\text{М}}$, матрица $\overline{A}_{\text{М}}$ пополняется строкой $\overline{a}^{\text{T}}_*$, величины r, s заменяются на $r - i$ и $s + i$ соответственно; осуществляется переход на шаг 2. В противном случае элемент $b_{(r-i)}\,b_{(s+i)}$ отбрасывается, величина i увеличивается на единицу и производится возвращение на начало шага 3.

При обеих альтернативах производим проверку: $i = r - 1$ и $i = n - s$. Если эти условия выполнены, то конец работы алгоритма. В принципиальном плане алгоритм рассматривается только для модели (4), поскольку случай системы (1) сводится к поэтапному применению одного и того же алгоритма вначале к матрице, а затем к правой части, с последующим анализом МНК-решения системы, составленной из множества пересечений.

В [5] подробно исследован случай задачи оценки параметра сдвига (модель (5)), и доказана теорема.

<u>Теорема.</u> Пусть выполнено соотношение (5), а условия, накладываемые на ошибки задания правых частей Δe_i системы (5) такие же, как условия (6) для ошибок задания правых час-

тей Δb_i системы (4). Тогда решение системы (5) выделяющим алгоритмом при $\varepsilon \neq 0$, $0 < \varepsilon \leq 1/2$ является наиболее *B*-робастной и наиболее *V*-робастной оценкой, причем имеют место следующие соотношения: $\gamma^* = \sqrt{\pi/2}$, $\kappa^* = 2$, $\varepsilon = 1/2$. Более того, предложенный алгоритм определяет оптимальную *B*- и *V*-робастную оценку решений.

Если же $\varepsilon = 0$, то выделяющий алгоритм определяет единственную несмещенную оценку с минимальной дисперсией среди всех несмещенных оценок. Здесь γ^* – чувствительность к большой ошибке; κ^* – чувствительность к изменению дисперсии, а ε – пороговая точка [8].

Неизвестный вид закона распределения ошибок (6) в условиях конечного размера выборки разнораспределенных случайных величин делает невозможным теоретическое изучение результатов оценивания решения. Поэтому разработан конкретный алгоритм метода Монте-Карло, реализующий численный подход к анализу результатов исследований.

Методика оценки точности восстановления сигнала и конкретные параметры алгоритма моделирования приведены в [5]. Отметим, что огибающая сигнала была аппроксимирована полиномом второй степени (объем выборки – 100 наблюдений), параметр масштаба аномальностей на два порядка превышал СКО основного шума, а количество независимых реализаций, имитирующих сигнал, было выбрано равным 100. Результаты оценки точности алгоритма представлены в таблице, в которой для сравнения приведены полученные на этой же модели точностные характеристики линейного алгоритма метода наименьших квадратов (РМНК), а также известного робастного алгоритма П. Хьюбера (РХ). В числителе приведенных данных – оценка среднего значения отклонения оценки формы сигнала от истинной, в знаменателе – оценка СКО.

Результаты оценки точности исследуемых методов

Метод	Модель											
	Стохастическая						Промежуточная					
	0	0,1	0,2	0,3	0,4	0,5	0	0,1	0,2	0,3	0,4	0,5
РМНК	$\frac{1{,}0}{9{,}5}$	$\frac{0{,}02}{4{,}9}$	$\frac{-0{,}05}{4{,}1}$	$\frac{0{,}3}{4{,}4}$	$\frac{-0{,}06}{4{,}3}$	$\frac{0{,}2}{4{,}4}$	$\frac{0{,}1}{1{,}0}$	$\frac{2}{50}$	$\frac{8}{65}$	$\frac{17}{87}$	$\frac{13}{107}$	$\frac{-3}{114}$
РХ	–	–	–	–	–	–	$\frac{0{,}2}{1{,}03}$	$\frac{-0{,}03}{1{,}1}$	$\frac{-0{,}3}{1{,}8}$	$\frac{0{,}3}{1{,}8}$	РНП*	РНП
РА**	$\frac{0{,}7}{9{,}9}$	$\frac{2{,}0}{11{,}2}$	$\frac{1{,}3}{11{,}7}$	$\frac{-1{,}8}{13{,}3}$	$\frac{0{,}8}{12{,}3}$	$\frac{0{,}3}{13{,}8}$	$\frac{0{,}1}{1{,}0}$	$\frac{-0{,}05}{1{,}1}$	$\frac{0{,}08}{1{,}3}$	$\frac{0{,}1}{1{,}3}$	$\frac{0{,}04}{1{,}3}$	$\frac{0{,}08}{1{,}5}$

* Результат не получен.

** РА – робастный алгоритм, предложенный авторами.

Анализ результатов моделирования показывает, что для обеих моделей (1) и (4) и интенсивности <засорения>, изменяющейся от 0 до 0,5, лучшие характеристики имеет построенный устойчивый алгоритм. Некоторое увеличение (по сравнению с оптимальной) дисперсии оценки предложенным алгоритмом при $\varepsilon = 0{,}5$ может быть объяснено конечностью обрабатываемой выборки.

В настоящее время производится модификация алгоритма с привлечением в качестве базисных функций сглаживающих сплайнов.

1. Зуев В.Е. Лазер покоряет небо. Новосибирск: Западно-Сибирское книжное изд-во, 1972. 192 с.
2. Стогов Г.В., Макшанов А.В., Мусаев А.А. // Зарубежная радиоэлектроника. 1982. N 9. С. 3–46.
3. Хьюбер П. Робастность в статистике. М.: Мир, 1984. 304 с.
4. Бронников А.В., Воскобойников Ю.Е. // Автометрия. 1990. N 1. С. 21–26.
5. Демьянов Ю.Ф. Исследование и разработка устойчивых методов полиномиальной регрессии: Дис. на соиск.уч.степ. канд. физ.-мат. наук (05.13.16) // Научн.рук. Л.Я. Савельев. Приозерск, 1992. 88 с.
6. Дейвид Г. // Порядковые статистики. М.:Наука, 1979. 335 с.
7. Себер Дж. // Линейный регрессивный анализ. М.:Мир, 1980. 456 с.
8. Хампель Ф. и др. Робастность в статистике. (Подход на основе функций влияния) // Ф. Хампель, Э. Рончетти, П. Рауссеу, В. Штаэль. М.: Мир, 1989. 512 с.

Поступила в редакцию
29 июля 1993 г.

Yu. F. Dem'yanov. **Stable Algorithm of Reconstructing the Shape of a Pulsed Lidar Return Signal.**

In this paper a method is proposed for reconstructing the shape of a pulsed lidar return signal under conditions when the law of error distribution differs from the normal distribution law. A stable algorithm of the reconstruction is proposed and its investigation with Monte-Carlo method is discussed. A theorem that allows one, for a paticular case of estimating the shift parameter, to assess optimal properties of the algorithm is presented.