# АЛГОРИТМ АВТОМАТИЧЕСКОГО ПОИСКА СИГНАЛЬНЫХ ИНТЕРВАЛОВ СКАНА В ПРИСУТСТВИИ ШУМА НЕИЗВЕСТНОЙ ИНТЕНСИВНОСТИ

*В. Л. Горохов, В. Н. Прокофьев*

На основе принципов подобия и несмещенности синтезирован устойчивый алгоритм автоматической регистрации зон в исходных наблюдениях, содержащих слабые сигналы в шумах. Алгоритм предназначен для программ автоматической обработки радиоастрономических наблюдений на РАТАН-600. Получены и экспериментальные и теоретические рабочие характеристики алгоритма.

The steady algorithm of automatic interval registration in arecorded observations containing weak signals in noises, is sinthasized on the principles of unbiasment and similarity. This algorithm is intended for automatic data-acquisition programms of radio-observations on RATAN-600. Theoretical and experimental work characteristics of the algorithm are obtained.

В настоящее время существует тенденция к обработке астрономических наблюдений на ЭВМ. Часто такая обработка связана с использованием статистических методов. В некоторых задачах радиоастрономии решающие правила приема отыскиваются среди хорошо известных байесовских правил [1]. Однако получаемые при этом алгоритмы требуют полных априорных сведений о параметрах наблюдаемых случайных процессов. Для успешной автоматизации наблюдений необходимы алгоритмы обработки, малочувствительные (или даже нечувствительные) к тем или иным изменениям условий наблюдения (т. е. средних характеристик процессов).

Такие устойчивые (по отношению к условиям наблюдения) алгоритмы имеют устойчивые показатели качества и в реальных условиях работы оказываются более эффективными, чем алгоритмы, не обладающие подобной устойчивостью (последние могут оказаться даже просто неработоспособными в тех или иных ситуациях). Одним из путей преодоления априорной неопределенности в радиоастрономических задачах является использование стохастической аппроксимации [2]. Получаемые таким образом алгоритмы носят асимптотический характер: их оптимальные свойства имеют место при неограниченном возрастании времени наблюдения (так же как и свойства устойчивости); их оправданное применение требует глубокой стационарности; наконец, при конечном времени наблюдения их качественные характеристики неясны. Можно отметить также, что эти алгоритмы часто бывают весьма сложными и трудными для реализации.

Возможен другой путь преодоления «априорной трудности» за счет синтеза алгоритмов на базе статистических принципов подобия, несмещенности и инвариантности [3, 4, 5, 6, 7]. Развитые здесь современные регулярные методы статистического синтеза конструктивны и во многих практических задачах позволяют получить при любом конечном времени анализа оптимальные решающие правила с высокими свойствами устойчивости их качественных показателей по отношению к условиям наблюдения, пригодные для реализации в виде подпрограмм систем автоматизации обработки наблюдений на ЭВМ. Как уже отмечалось, реальная эффективность таких правил выше, чем эффективность неустойчивых правил (например, байесовских), которые, будучи реализованными в реальных условиях, могут давать результаты, весьма далекие от потенциально достижимых, вследствие несовпадения фактических параметров наблюдаемых случайных процессов с расчетными значениями (записанными в алгоритмы при синтезе). В данной работе рассматривается радиоастрономическая задача, характеризующая недостаток априорных сведений. Для ее решения используются отмеченные принципы подобия и несмещенности. Полученное правило обладает, например, такими полезными свойствами устойчивости, как независимость вероятности ложной тревоги от уровня мешающих шумов (без каких-либо подстроек порога), независимость вероятности правильного обнаружения от масштабных изменений наблюдений. Перейдем к формулировке задачи и исходных предпосылок.

Пусть осуществляется сканирование участка неба диаграммой направленности радиотелескопа с целью обнаружения и выделения источника радиоизлучения. Часто этот источник слаб и необходимо накопление — многократное суммирование многих сканов одного и того же участка неба [8]. При этом вследствие различного уровня фона от скана к скану последние суммируются после их центрирования относительно своих средних значений («проведение нуля»). Причем эту процедуру следует производить только по той части скана, которая обязана шуму, т. е. исключив «сигнальные области» в скане. В противном случае вносятся искажения, и выделение сигнала затрудняется. Возникает задача обнаружения «сигнальной области» в сканах. В дальнейшем считается, что длительность такой сигнальной области (интервал разрешения сигнала) задана и определяется для точечных источников параметрами диаграммы направленности. Пусть реализация скана имеет $n$ интервалов разрешения. Принимается, что местоположение сигнального интервала одинаково на всех сканах и задача решается для двух смежных сканов. При этом предполагается, что уровни шума в сканах различны и неизвестны. В каждом скане уровень шума не является постоянным, но может меняться вдоль скана достаточно плавно и одинаков для каждой пары смежных интервалов разрешения. Осуществляется анализ последовательных пар таких интервалов для двух сканов, начиная с первой пары, по методу «контраста» [9]. Сигнал может появиться во втором интервале анализируемой пары интервалов. Наличие сигнала во втором интервале пары приводит к возрастанию мощности наблюдаемого процесса здесь по отношению к первому интервалу (как отмечалось, в отсутствие сигнала шум имеет одинаковую мощность в обоих интервалах анализируемой пары, воможно различную от пары к паре).

Пусть в качестве наблюдаемых данных выступают отсчеты двух реализаций (два обрабатываемых скана) на выходе радиометра, взятые в $n$ выделенных интервалах разрешения. В ряде случаев можно полагать эти отсчеты независимыми случайными величинами с показательной плотностью распределения. На основе принятых предпосылок сформулируем задачу более точно.

Пусть $x$, $y$ и $z$, $t$ — отсчеты в выделенной паре интервалов для первого и второго наблюдаемых сканов соответственно. Их совместное распределение, согласно предпосылкам:

$$P(x,\ y,\ z,\ t) = (ab\sigma^4\tau^4)^{-r} \exp(-x/\sigma^2 - y/a\sigma^2 - z/\tau^2 - t/b\tau^2); \qquad (1)$$

$$x = \sum_{j=1}^{r} x_j;\quad y = \sum_{j=1}^{r} y_j;\quad z = \sum_{j=1}^{r} z_j;\quad t = \sum_{j=1}^{r} t_j;\quad r = 1;\quad (x,\ y,\ z,\ t) > 0,$$

где $\sigma^2$, $\tau^2$ — неизвестные мощности шума для первого и второго сканов (в анализируемой паре); $a = 1 + q_1$; $b = 1 + q_2$; $q_1 = \nu_1^2/\sigma^2$; $q_2 = \nu_2^2/\tau^2$; $\nu_1^2$, $\nu_2^2$ —

мощность сигналов для первого и второго сканов; $q_1$, $q_2$ — соответствующие с/ш.

В отсутствие сигналов $a = b = 1$. Введем «сигнальный» (полезный, проверяемый) параметр $\theta$:

$$\theta = (1/\sigma^2 - 1/a\sigma^2) + (1/\tau^2 - 1/b\tau^2),$$

тогда распределение (1) можно записать в виде

$$P(x,\ y,\ z,\ t) = A \exp[-\theta x + \gamma_1(x+y) + \gamma_2(x+t) + \gamma_3(z-x)];\ (x,\ y,\ z,\ t) > 0, \quad (2)$$

где

$$A = -\gamma_1\gamma_2\gamma_3(\theta + \gamma_3 - \gamma_1 - \gamma_2);\ \gamma_1 = -1/a\sigma^2;\ \gamma_2 = -1/b\tau^2.\ \gamma_3 = -1/\tau^2. \quad (3)$$

Задача обнаружения сигнала заключается в проверке гипотез на базе экспонентного семейства распределений (2) $H_0 : \theta = 0$, $H_1 : \theta > 0$; $(\gamma_1, \gamma_2, \gamma_3)$ — неизвестные (мешающие) (3).

Наличие мешающих параметров в формулировке задачи (3) соответствует незнанию уровней шума (а также их различной величине) в сканах, произвольности масштаба данных, т. е. отражает фактическую ситуацию при наблюдении. Ясно, что для задачи (3) нельзя воспользоваться байесовым решением, или оно будет включать в себя неизвестные параметры $\gamma_1$, $\gamma_2$, $\gamma_3$ и окажется поэтому нереализуемым. Для получения практического решения воспользуемся принципом подобия [3, 4]. Решающие правила со свойствами подобия (подобные при $H_0$ в данном случае) имеют неизменную вероятность ложной тревоги при любых значениях мешающих параметров $\gamma_1$, $\gamma_2$, $\gamma_3$, что является, конечно, очень ценным для практики их свойством. Выбор наилучшего (оптимального в смысле вероятности правильного обнаружения) правила среди подобных осуществляется переходом к неймановским структурам [4, 6] при выполнении определенных условий для рассматриваемого семейства распределений.

Семейство (2), согласно теореме факторизации [3, 4], допускает достаточные статистики: $U = x$, $T_1 = x + y$, $T_2 = x + t$, $T_3 = z - x$ для параметров $\theta$, $\gamma_1$, $\gamma_2$, $\gamma_3$, и можно ограничиться далее только ими. На базе (2) легко найти совместное распределение достаточных статистик $(U,\ T)$, $T = (T_1,\ T_2,\ T_3)$ в виде:

$$P(U,\ T) = A \exp\left(-\theta U + \sum_{i=1}^{3} \gamma_i T_i\right);$$
$$\max_{-\infty < T_3 < +\infty}(0,\ -T_3) \leqslant U \leqslant \min_{T_1,\ T_2 > 0}(T_1,\ T_2). \quad (4)$$

Соотношения между величинами $U_1$, $T_1$, $T_2$, $T_3$ непосредственно следуют из их определения и условия $x$, $y$, $z$, $t > 0$. Переход к правилам неймановской структуры требует вычисления условного семейства распределений $P(U/T)$ при фиксированной статистике $T$, достаточной для мешающего параметра $\gamma = (\gamma_1,\ \gamma_2,\ \gamma_3)$. По свойствам достаточности тогда семейство $P(U/T)$ является однопараметрическим, зависящим лишь от полезного параметра $\theta$ (и не зависящим от мешающего параметра $\gamma$). В этом заключается важный смысл перехода к неймановским правилам, поскольку условное (при заданной $T$) решение задачи (3) на базе однопараметрического семейства $P(U/T)$ получить бывает легко. Усредняем (4) по $U$, получаем $P(T)$. Затем получаем распределение $P(U/T) = P(U,T)/P(T)$:

$$P(U/T) = \frac{\theta \exp(-\theta U)}{\exp[-\theta \max(0,\ -T_3)] - \exp[-\theta \min(T_1,\ T_2)]};$$
$$\max(0,\ -T_3) \leqslant U \leqslant \min(T_1,\ T_2). \quad (5)$$

Семейство (5) действительно зависит только от одного проверяемого параметра $\theta$ ($T$ фиксирована) и является, как видно, экспонентным. Поэтому для проверки гипотез (3) существует равномерно наиболее мощное (РНМ) [4] относительно $\theta$ решающее правило с критической областью вида

$$U < C(T). \tag{6}$$

Порог $c(T)$, зависящий от фиксированной статистики $T$ (но не от мешающих параметров $\gamma_1$, $\gamma_2$, $\gamma_3$!), определяется заданной вероятностью ложной тревоги $\alpha$ из условия

$$\alpha = \int\limits_{\max(0,\,-T_3)}^{C(T)} P_0(U/T)\, dU, \tag{7}$$

где $P_0(U/T)$ — распределение (5) при чистом шуме, т. е. при $\theta = 0$. Из (5) находим

$$P_0(U/T) = \lim P(U/T) = 1/\min(T_1,\ T_2) - \max(0,\ -T_3);\ \theta \to 0;$$
$$\max(0,\ -T_3) \leqslant U \leqslant \min(T_1,\ T_2). \tag{8}$$

Подставив (8) в (7), получаем порог $C(T)$ правила (6) в виде

$$C(T) = \alpha \min(T_1,\ T_2) + (1-\alpha)\max(0,\ -T_3). \tag{9}$$

Правило (6), (9) можно считать искомым решением задачи (3). Оно не зависит от неизвестных параметров $\gamma_1, \gamma_2, \gamma_3$ и при любых их значениях обеспечивает заданную вероятность ложной тревоги. Удобно выразить критическую область (6) через исходные данные $x$, $y$, $z$, $t$, заменяя в (6) и (9) статистики $U$, $T$ через $x, y, z,$ , нетрудно получить эту область в виде

$$\omega = \min(y,\ t)/\min(x,\ z) > C;$$
$$C = (1-\alpha)/\alpha. \tag{10}$$

Правило (10) позволяет дать ясную физическую интерпретацию его свойств устойчивости. Так, например, в отсутствие сигнала числитель и знаменатель тестовой статистики $\omega$ в (10) имеют одинаковые распределения [ибо пары $(x,\ z)$ и $(y,\ t)$ имеют при шуме одинаковые плотности]; поэтому плотность $\omega$ при шуме вообще не зависит от параметров $\sigma^2$, $\tau^2$ и, следовательно, вероятность $\alpha$ ложной тревоги также не зависит от них (а пороговая константа $C$ определяется только вероятностью $\alpha$). Вероятность правильного обнаружения (мощность) правила (10) легко найти, вычислив распределение $P_0(\omega)$ статистики при $H_0$ (при чистом шуме). Пусть $u = \min(y, t)$, $v = \min(x, z)$, $P_0(v) = 1/\mu_0 \exp(-v/\mu_0)$, $P(u) = 1/\mu \exp \times \times (u/\mu)$. Здесь $(1/\mu_0) = (1/\sigma^2 + 1/\tau^2)$, $(1/\mu) = (1/a\sigma^2 + 1/b\tau^2)$. При чистом шуме $\mu = \mu_0$, и стандартным путем замены переменных получаем $P_0(\omega) = 1/(1 + + \omega)^2$. При сигнале точно такое же распределение имеет величина $(\mu_0\omega/\mu)$, поэтому мощность правила (10)

$$\beta = P\{\omega > C/H_1\} = P\{(\mu_0/\mu)\,\omega > (\mu_0/\mu)\,C/H_1\} = \int\limits_{C_1}^{\infty} P_0(\omega)\, d\omega;\quad C_1 = (\mu_0/\mu)\,C.$$

Подставляя сюда плотность $P_0(\omega)$ и $C = (1-\alpha)/\alpha$, находим мощность в виде

$$\beta = [\alpha ab\,(1+\Delta)]/[\alpha ab\,(1+\Delta) + (1-\alpha)(\alpha + \Delta b)];$$
$$\Delta = \tau^2/\sigma^2;\quad a = 1 + q_1;\quad b = 1 + q_2. \tag{11}$$

Итак, в рассматриваемой задаче обнаружения получено устойчивое к условиям наблюдения решающее правило (6), (9), (10), которое не зависит от мешающих параметров наблюдения, выдерживает неизменной вероятность ложной тревоги при любых уровнях шума $\sigma^2$, $\tau^2$ и имеет мощность (11), не зависящую от общих масштабных изменений наблюдений. Правило является простым и пригодно для реализации в виде подпрограммы системы автоматизации обработки результатов наблюдений.

В заключение отметим, что устойчивый к условиям наблюдения алгоритм обнаружения может быть получен в более общем случае обработки $m > 2$ сканов.

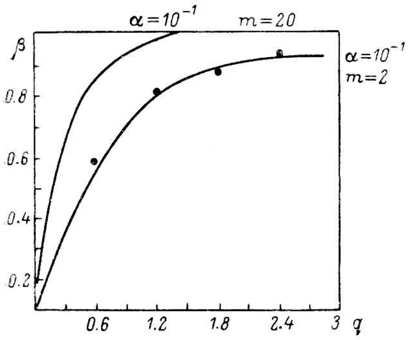


Рис. 1.

На рис. 1 показаны зависимости вероятности правильного обнаружения $\beta$ от отношения (с/ш) $q$ для $m=20$ и 2, $r=5$ при заданном уровне ложной тревоги $\alpha$. Точками отмечены результаты экспериментальной проверки теста на реальных наблюдениях. Вид наблюдательного материала и результаты работы теста показаны на рис. 2.


Рис. 2.

**Список литературы**


1. Измерение радиотепловых и плазменных излучений. Под. ред. А. Е. Башаринова. М., «Сов. радио», 1969.
2. Фридман П. А. Рекурентные алгоритмы восстановления радиоастрономических изображений. — Сообщ. САО, 1972, **6**.
3. Линник Ю. В. Статистические задачи с мешающими параметрами. «Наука», 1966.
4. Леман Э. Проверка статистических гипотез. Л., «Наука», 1964.
5. Закс Ш. Теория статистических выводов. М., «Мир», 1975.
6. Богданович В. А. Применение принципа несмещенности в задачах обнаружения с априорной неопределенностью. — Радиоэлектроника, 1972, **15**, № 4.
7. Богданович В. А. Применение принципа инвариантности в задачах обнаружения с априорной неопределенностью. — Радиоэлектроника, 1973, **16**, № 1.
8. Краус Дж.-Д. Радиоастрономия. М., «Сов. радио», 1973.
9. Прокофьев В. Н. Некогерентный обнаружитель флуктуирующих сигналов в шумах неизвестной интенсивности. — Радиоэлектроника, 1970, **13**, № 2.