УДК 519.95

# МЕТОД МАКСИМАЛЬНОЙ НАДЕЖНОСТИ В ЗАДАЧЕ ВЫБОРА МОДЕЛИ

**Ю. П. Пытьев, М. Л. Сердобольская**
*(кафедра физики атмосферы и математического моделирования)*

**Предложен новый метод выбора модели, рассмотрен способ наилучшего тестирования модели, проведен численный анализ метода на примере задачи определения неизвестного параметра модели.**

Как известно, для решения задачи интерпретации косвенных измерений необходима математическая модель эксперимента. При этом особо остро встает вопрос соответствия модельных представлений и экспериментальных данных, поскольку интерпретация, основанная на модели, не отвечающей реальной измерительной процедуре, может повлечь за собой значительные ошибки. В настоящей работе предложен новый метод выбора модели измерения, а также рассмотрен способ наилучшего тестирования при дискриминации моделей. Возможности метода проиллюстрированы в вычислительном эксперименте.

1. **Задача редукции экспериментальных данных.** Рассмотрим линейную схему измерения:

$$\xi = Af + \nu \in \mathcal{R}_n, \tag{1}$$

где $\mathcal{R}_n$ — $n$-мерное евклидово пространство; $\xi$ есть результат измерения сигнала $f \in \mathcal{R}_m$ от исследуемого объекта, $\nu$ — случайный сигнал (шум), линейный оператор $A \in \mathcal{R}_m \to \mathcal{R}_n$ моделирует измерительный прибор. Будем считать, что задана модель $[A, \Sigma]$ схемы измерения (1), т. е. известны оператор $A$ и невырожденный корреляционный оператор шума $\Sigma$, причем $\nu$ имеет нормальное распределение $\mathcal{N}(0, \Sigma)$. Линейное преобразование измерения (1) $R\xi = RAf + R\nu$ будем интерпретировать как искаженный шумом $R\nu$ результат измерения $f$ на приборе $U$, выходной сигнал которого равен параметрам исследуемого объекта. Величина $\mathbf{E}\|R\xi - Uf\|^2$ определяет среднеквадратичную погрешность такой интерпретации, и если $RA \neq U$, то в силу произвольности $f$ значение погрешности оценить невозможно. В самом деле, $\mathbf{E}\|R\xi - Uf\|^2 = \mathbf{E}\|(RA - U)f + R\nu\|^2$, и $\sup\{\mathbf{E}\|R\xi - Uf\|^2 \mid f \in \mathcal{R}_m\} = \infty$, если $RA \neq U$.

Определим оператор $R$ из условия

$$h = \mathbf{E}\|R\xi - Uf\|^2 = \inf\{\mathbf{E}\|R'\nu\|^2 \mid R',\ R'A = U\}, \tag{2}$$

тогда $R\xi$ — вектор параметров объекта $Uf$, искаженный шумом $R\nu$ минимальной интенсивности, $R\xi$ — наиболее точная «несмещенная версия» $Uf$ в классе всех оценок, в том числе и нелинейных, поскольку речь идет о нормальном распределении. Задача (2) называется задачей редукции, а $R\xi$ — редукцией измерения $\xi$ к прибору $U$. Если невырожден оператор $A^*\Sigma^{-1}A$, то решение (2) существует для любого $U$ и имеет вид $R\xi = U(A^*\Sigma^{-1}A)^{-1}A^*\Sigma^{-1}\xi$ [1]. Погрешность редукции

$$h = \mathrm{Tr}\, U(A^*\Sigma^{-1}A)^{-1}U^*. \tag{3}$$

Если модель неверна, причем ошибочно задан оператор $A$, так что на самом деле $\xi = (A+B)f + \nu$, то

$$\mathbf{E}\|R\xi - Uf\|^2 = \|U(A^*\Sigma^{-1}A)^{-1}A^*\Sigma^{-1}Bf\|^2 + \mathbf{E}\|R\nu\|^2, \tag{4}$$

и формула (3) не характеризует реальную погрешность.

При анализе редукции важна проблема не только соответствия модели экспериментальным данным, но и возможности использования ее для редукции измерения. Этот вопрос также частично рассмотрен в работе.

2. **Понятие надежности модели.** Если и в самом деле имеет место модель $[A, \Sigma]$, то $\xi \sim \mathcal{N}(a, \Sigma)$, где $a \in \mathcal{R}(A)$. В качестве альтернативы примем $\xi \sim \mathcal{N}(a, \Sigma)$, $a \overline{\in} \mathcal{R}(A)$, так как в этом случае измерение не согласуется с моделью. Эта задача подробно рассмотрена в [2], где показано, что семейство распределений $\{\mathcal{N}(a, \Sigma), a \overline{\in} \mathcal{R}(A)\}$, а также задача проверки гипотезы $a \in \mathcal{R}(A)$ при альтернативе $a \overline{\in} \mathcal{R}(A)$ инвариантны относительно группы движений $\mathcal{H} = \{(H, h)\}$ пространства $\mathcal{R}_n$ таких, что $(H, h)x = Hx + h$, $x \in \mathcal{R}_n$, где $H = \Sigma^{1/2} G \Sigma^{-1/2}$, $G \in \mathcal{G}$, $\mathcal{G}$ — группа ортогональных преобразований $\mathcal{R}_n$, оставляющих инвариантным $\mathcal{R}(\Sigma^{-1/2}A)$, $h \in \mathcal{R}(A)$. При этом в классе $\mathcal{H}$-инвариантных критериев существует равномерно наиболее мощный, критическое множество которого имеет вид $Z_\alpha = \{\xi : t(\xi) = \|(I - Q_A)\Sigma^{-1/2}\xi\|^2 \geqslant t_\alpha\}$, где $Q_A$ — ортогональный проектор на $\mathcal{R}(\Sigma^{-1/2}A)$. Величина $t_\alpha$ связана с вероятностью $\alpha$ ошибочно отвергнуть гипотезу равенством $\alpha = \int\limits_{Z_\alpha} p_{\chi^2}(\mu)\, d\mu$.

Здесь $p_{\chi^2}(\mu)$ — плотность $\chi^2$-распределения, согласно которому распределена статистика $t(\xi)$, если верна гипотеза.

Надежностью модели $[A, \Sigma]$ называется случайная величина $\alpha_A(\xi) = \inf\{\alpha \mid \xi \in Z_\alpha\}$, которую с известными оговорками можно интерпретировать как вероятность ошибочно отвергнуть модель $[A, \Sigma]$ на основании $\xi$. Нетрудно видеть, что

$$\alpha(\xi) = \int\limits_{t(\xi)}^{\infty} p_{\chi^2}(\mu)\, d\mu. \tag{5}$$

Согласно [2], если модель верна, то надежность распределена равномерно на $[0, 1]$, в противном случае $\alpha_A(\xi)$ принимает преимущественно малые значения. Чем меньше $\alpha_A(\xi)$, тем сильнее наблюдение $\xi$ свидетельствует против модели, следовательно, если надежность мала, то модель должна быть признана не отвечающей реальному измерению. Однако, возвращаясь к формуле (4) для погрешности, нетрудно заметить, что ответ на вопрос о применимости модели для редукции не получен. Будем говорить, что модель $[A, \Sigma]$ можно использовать для редукции измерения $\xi = (A + B)f + \nu$, если вклад ошибки модели в полную погрешность редукции не превосходит погрешности, вызванной шумом, т. е.

$$\|U(A^*\Sigma^{-1}A)^{-1}A^*\Sigma^{-1}Bf\|^2 \leqslant \operatorname{Tr} U(A^*\Sigma^{-1}A)^{-1}U^*. \tag{6}$$

К сожалению, оказывается, что надежность и, более того, само измерение $\xi$ не позволяют судить о том, пригодна ли модель для редукции. Имеет место следующий результат.

Теорема 1. Пусть согласно (6)

$$\mathcal{D}_A = \{b \in \mathcal{R}_n : d(b) = \|U(A^*\Sigma^{-1}A)^{-1}A^*\Sigma^{-1}b\|^2 \leqslant \operatorname{Tr} U(A^*\Sigma^{-1}A)^{-1}U^*\}$$

— множество тех $b \in \mathcal{R}_n$, для которых модель $[A, \Sigma]$ пригодна для редукции $\xi = a + b + \nu$, $b = Bf$, $a = Af \in \mathcal{R}(A)$. Тогда

1) Для любого $b \in \mathcal{D}_A$ найдется $b' \in \mathcal{R}_n \setminus \mathcal{D}_A$ такое, что статистика $\alpha(\xi) = \alpha(a + b + \nu)$ имеет то же распределение, что и статистика $\alpha(\xi') = \alpha(a + b' + \nu)$.

2) Для любой статистики $\beta(\xi)$ справедливо утверждение, анало-

гичное утверждению 1, т. е. $\forall b \in \mathscr{D}_A \ \exists b' \in \mathscr{R}_n \setminus \mathscr{D}_A$: $\beta(\xi)$ распределена так же, как и $\beta(\xi')$.

*Доказательство.*

1) Поскольку $d(b) = \|U(A^*\Sigma^{-1}A)^{-1}A^*\Sigma^{-1/2}Q_A\Sigma^{-1/2}b\|^2$, то вклад ошибки модели в погрешность редукции зависит от проекции вектора $\Sigma^{-1/2}b$ на $\mathscr{R}(\Sigma^{-1/2}A)$, в то время как распределение статистики $t(\xi)$ в (5) и, следовательно, распределение $\alpha_A(\xi)$ определяются $(I-Q_A)\times\Sigma^{-1/2}b$ — проекцией $\Sigma^{-1/2}b$ на $\mathscr{R}^{\perp}(\Sigma^{-1/2}A)$. Поэтому, если $Q_Ab'$ порождает такую ошибку модели, что $d(b') > \mathrm{Tr}\, U(A^*\Sigma^{-1}A)^{-1}U^*$ и при этом $(I-Q_A)\Sigma^{-1/2}b' = (I-Q_A)\Sigma^{-1/2}b$, $b$ — произвольный вектор из $\mathscr{D}_A$, то $b' \in \mathscr{R}_n \setminus \mathscr{D}_A$ и распределения $\alpha(\xi)$ и $\alpha(\xi')$ совпадают.

2) В силу произвольности $f$ по измерению $\xi$ нельзя различить операторы $A$ и $A+B$, если $\mathscr{R}(B) \subset \mathscr{R}(A)$, при этом ошибка модели может давать неопределенно большой вклад в погрешность редукции. Следовательно, распределение любой статистики $\beta(\xi)$, основанной на измерении $\xi$, в общем случае не несет информации о погрешности редукции.

Обратим внимание на то, что в теореме, по существу, идет речь только о таких ошибках модели, при которых $Q_A\Sigma^{-1/2}b = 0$ или $(I-Q_A)\Sigma^{-1/2}b = 0$. Подобные ошибки носят весьма специфический характер и вряд ли могут встретиться на практике. В реальной ситуации следует ожидать, что низкая надежность скорее всего свидетельствует о неприменимости модели для редукции*. Для иллюстрации этих выводов рассмотрим вычислительный эксперимент.

3. **Метод максимальной надежности.** Численный эксперимент. Пусть априори известно, что истинная модель принадлежит классу моделей $\mathscr{M}$. Методом максимальной надежности назовем метод выбора модели $[A, \Sigma]$, основанный на измерении $\xi$, при котором $\alpha_A(\xi) = \max\{\alpha_{\tilde{A}}(\xi) \mid [\tilde{A}, \Sigma] \in \mathscr{M}\}$.

Пусть в (1) $A = A(\Delta)$ моделирует простейшую оптическую систему, $\Delta$ — параметр расфокусировки, который равен числу ненулевых диагоналей матрицы $A(\Delta)$. В численном эксперименте неизвестная априори величина расфокусировки определялась по методу максимальной надежности из пяти возможных значений $\Delta_i = 2i+1$, $i = 1, \ldots, 5$ на основании значений статистики $t_{ij} = \|(I-Q_i)\Sigma^{-1/2}\xi_j\|^2$, $Q_i$ — ортогональный проектор на $\mathscr{R}(\Sigma^{-1/2}A(\Delta_i))$, $\xi_j = A(\Delta_j)f + \nu_j$, $i, j = 1, \ldots, 5$. Одновременно проводился расчет редукции $R_i\xi_j$ к $U = I$ в рамках модели $[A(\Delta_i), \Sigma]$. Расчет показал, что при малом шуме метод максимальной надежности определял верную модель (минимум $t_{ij}$ достигался при $i = j$). Если шум увеличивался, то в $t_{ij} = \|(I-Q_i)\Sigma^{-1/2}A(\Delta_j)f + (I-Q_i)\Sigma^{-1/2}\nu_j\|^2$ вклад первого слагаемого уменьшался, что в ряде случаев приводило к выбору неверной модели. Результаты расчета приведены в таблице.

Обратим внимание на случай $i = 4$, $j = 1$, когда минимум достигается на неверной модели даже при низком шуме. В согласии с теоремой 1 это объясняется тем, что при данном $f$ вектор $\Sigma^{-1/2}b = \Sigma^{-1/2}(A(\Delta_1) - A(\Delta_4))f \in \mathscr{R}(\Sigma^{-1/2}A(\Delta_4))$, и статистика $t_{41}$ распределена так же, как и $t_{44}$. Этот случай проиллюстрирован на рис. 1, *а*. На рис. 1, *б* и 2 приведены результаты редукции при низком $(\Sigma = 10^{-4}\cdot I)$ и высоком $(\Sigma = I)$ шумах.

4. **Проблема тестирования.** Согласно теореме 1, если модель $[A, \Sigma]$ используется для редукции $\xi = (A+B)f + \nu$, то при $\mathscr{R}(\Sigma^{-1/2}B) \subset$

---

* Высокая надежность также не означает, что модель верна и тем более что ее можно использовать для редукции.

$\subset\mathcal{R}^{\perp}(\Sigma^{-1/2}A)$ ошибка модели не дает вклада в погрешность редукции. С другой стороны, если $\mathcal{R}(B)\subset\mathcal{R}(A)$, то при неизвестном $f$ по измерению $\xi$ невозможно различить модели $[A, \Sigma]$ и $[A+B, \Sigma]$. Для того чтобы различить подобные модели, необходимо использование тестовых сигналов, причем последние следует выбирать так, чтобы дискри-

**Значения статистики $t_{ij}$**

| $i$ | $j=1$ | $j=2$ | $j=3$ | $j=4$ | $j=5$ |
|---|---|---|---|---|---|
| | | | $\Sigma=10^{-4}\cdot I$ | | |
| 1 | $4{,}2\cdot10^{1}$ | $1{,}4\cdot10^{2}$ | $3{,}0\cdot10^{3}$ | $4{,}4\cdot10^{3}$ | $5{,}3\cdot10^{3}$ |
| 2 | $1{,}2\cdot10^{3}$ | $4{,}7\cdot10^{1}$ | $9{,}1\cdot10^{2}$ | $2{,}0\cdot10^{3}$ | $2{,}5\cdot10^{3}$ |
| 3 | $6{,}4\cdot10^{1}$ | $2{,}6\cdot10^{2}$ | $3{,}5\cdot10^{1}$ | $6{,}1\cdot10^{2}$ | $1{,}3\cdot10^{3}$ |
| 4 | $3{,}9\cdot10^{1}$ | $7{,}2\cdot10^{1}$ | $3{,}3\cdot10^{2}$ | $6{,}8\cdot10^{1}$ | $5{,}4\cdot10^{2}$ |
| 5 | $4{,}8\cdot10^{1}$ | $7{,}6\cdot10^{1}$ | $2{,}9\cdot10^{2}$ | $5{,}4\cdot10^{2}$ | $4{,}3\cdot10^{1}$ |
| | | | $\Sigma=I$ | | |
| 1 | 42,3 | 26,2 | 52,1 | 42,8 | 32,2 |
| 2 | 42,5 | 46,9 | 45,0 | 58,1 | 38,6 |
| 3 | 33,3 | 43,6 | 35,2 | 42,4 | 37,9 |
| 4 | 39,0 | 33,8 | 46,1 | 67,7 | 51,1 |
| 5 | 37,6 | 52,5 | 39,6 | 42,1 | 42,8 |

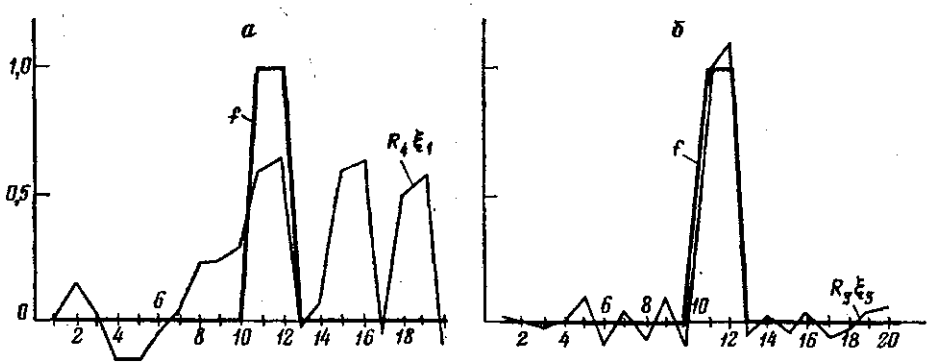


Рис. 1. Исходный сигнал $f$ и результат редукции в случае неверной $R_4\xi_1$ (*а*) и верной модели $R_3\xi_3$ (*б*) при малом шуме

минация моделей была в некотором смысле наилучшей. В этом параграфе будет показано, что тестирование позволяет различить любые две модели $[A_1, \Sigma]$ и $[A_2, \Sigma]$; $A_1\neq A_2$.

Задача различения двух моделей по наблюдению результата измерения тестового сигнала $f$ по сути эквивалентна задаче различения гипотез $\{H_1=H_1(f):\xi\sim\mathcal{N}(A_1f, \Sigma)\}$ и $\{H_2=H_2(f):\xi\sim\mathcal{N}(A_2f, \Sigma)\}$. Пусть $\varphi$ есть наиболее мощный критерий проверки $H_1$ при альтернативе $H_2$, $\alpha_k=\alpha_k(f)$ — вероятность ошибочно отвергнуть гипотезу $H_k$, $k=1, 2$, на основании $\varphi$. Будем говорить, что по измерению сигнала $f_0$ модели $[A_1, \Sigma]$ и $[A_2, \Sigma]$ различаются наилучшим образом, если $\alpha_1(f_0)+\alpha_2(f_0)$ минимально. Поскольку на практике амплитуда тестовых сигналов, как правило, ограничена, наложим на выбор $f$ условие: $\|f\|\leqslant1$.

Теорема 2. При условии $\|f\|\leqslant 1$ $f_0$ есть единичный собственный вектор оператора $(A_2-A_1)^*\Sigma^{-1}(A_2-A_1)$, соответствующий максимальному собственному значению.

*Доказательство.* По лемме Неймана—Пирсона [3] критическое

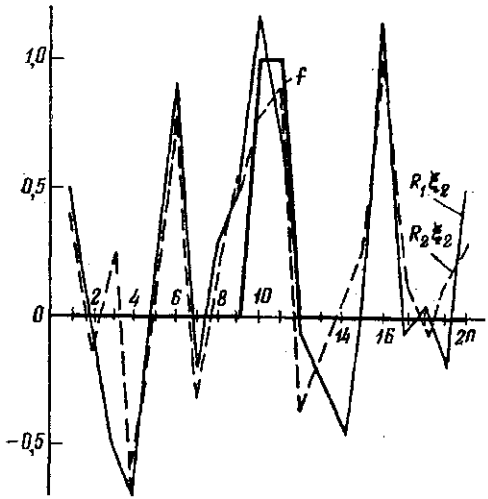


Рис. 2. Исходный сигнал $f$ и результат редукции в случае верной $R_2\xi_2$ и неверной модели $R_1\xi_2$ при большом шуме

множество $\varphi$ имеет вид $Z^C=\left\{x\in\mathcal{R}_n:\ \frac{n_2(x)}{n_1(x)}>C\right\}$, $n_k(x)$ — плотности распределения $\mathcal{N}(A_kf,\ \Sigma)$, $k=1,\ 2$. Ошибки критерия:

$$\alpha_1^C(f)=\int\limits_{Z^C} n_1(x)\,dx=\Phi(s^C),$$

$$\alpha_2^C(f)=\int\limits_{\mathcal{R}_n\setminus Z^C} n_2(x)\,dx=1-$$

$$-\Phi(s^C-\|\Sigma^{-1/2}(A_2-A_1)f\|^2),$$

где
$$s^C=\frac{(C-(\Sigma^{-1/2}A_1f,\ \Sigma^{-1/2}(A_2-A_1)f))}{\|\Sigma^{-1/2}(A_2-A_1)f\|},$$

$$\Phi(t)=\frac{1}{\sqrt{2\pi}}\int\limits_t^\infty e^{-\mu^2/2}\,d\mu$$

— функция ошибок. Условие $C_0(f)\sim\min\{\alpha_1{}^C(f)+\alpha_2{}^C(f)\,|\,C\in\mathcal{R}_1\}$ при каждом $f$ гарантирует минимальность суммарной ошибки критерия:

$$\varphi_0(x)=\begin{cases}1, & \text{если}\quad x\in Z^{C_0(f)},\\ 0, & \text{если}\quad x\overline{\in} Z^{C_0(f)}.\end{cases}\tag{7}$$

Нетрудно видеть, что

$$C_0(f)=\frac{1}{2}(\Sigma^{-1/2}(A_1+A_2)f,\ \Sigma^{-1/2}(A_2-A_1)f).\tag{8}$$

При этом ошибки критерия зависят только от $f$: $\alpha_k^{C_0(f)}(f)=\alpha_k(f)$, $k=1,2$ и $\min\{\alpha_1(f)+\alpha_2(f)\,|\,f\in\mathcal{R}_m:\|f\|\leqslant 1\}$ достигается на том же векторе $f=f_0$, что и $\max\{\|\Sigma^{-1/2}(A_2-A_1)f\|\,|\,f\in\mathcal{R}_m,\ \|f\|\leqslant 1\}$. Отсюда $f_0$ есть единичный собственный вектор $(A_2-A_1)^*\Sigma^{-1}(A_2-A_1)$, соответствующий максимальному собственному значению.

Итак, пусть $\xi$ есть результат измерения $f_0$. Формулы (7)—(8) при $f=f_0$ задают критерий различия моделей: если $\varphi_0(\xi)=0$, то по измерению $\xi$ выбирается $[A_1,\ \Sigma]$, если $\varphi_0(\xi)=1$, то модель $[A_2,\ \Sigma]$. Суммарная вероятность ошибочного решения при этом минимальна.

**Замечание.** Если верна гипотеза $H_1$, то надежность как вероятность ошибочно отвергнуть $H_1$ равна

$$\alpha_1(\xi) = \int\limits_{\eta(\nu, A_2, A_1)} n_0(\mu)\, d\mu, \text{ где } \eta(\nu, A_2, A_1) = \frac{(\Sigma^{-1/2}\nu,\ \Sigma^{-1/2}(A_2 - A_1) f)}{\|\Sigma^{-1/2}(A_2 - A_1) f\|}.$$

Здесь $n_0(\mu)$ — плотность одномерного распределения $\mathcal{N}(0, 1)$. Если верна $H_2$, то надежность равна

$$\alpha_2(\xi) = \int\limits_{\eta'}^{\infty} n_0(\mu)\, d\mu, \text{ где } \eta' = \eta(\nu, A_2, A_1) + \|\Sigma^{-1/2}(A_2 - A_1) f\|.$$

Поскольку при любых $A_2$ и $A_1$ $\eta(\nu, A_2, A_1)$ распределены одинаково, то при фиксированной надежности $\alpha_1(\xi)$ надежность $\alpha_2(\xi)$ минимальна, если максимальна величина $\|\Sigma^{-1/2}(A_2 - A_1) f\|$. Таким образом, $f = f_0$ является таким тестовым сигналом, который позволяет различать $[A_1, \Sigma]$ и $[A_2, \Sigma]$ по надежности в указанном выше смысле оптимальным образом.

В заключение отметим, что если модель выбрана в соответствии с измерением, то расчет погрешности редукции становится весьма сложным, поскольку в этом случае модель является случайной.

СПИСОК ЛИТЕРАТУРЫ

[1] Пытьев Ю. П. // Матем. сб. 1982. **118(160)**, № 1. С. 19. [2] Пытьев Ю. П.//Вестн. Моск. ун-та. Физ. Астрон. 1986. **27**, № 3. С. 14. [3] Боровков А. А. Математическая статистика. М., 1984.

Поступила в редакцию
23.04.87

ВЕСТН. МОСК. УН-ТА. СЕР. 3, ФИЗИКА. АСТРОНОМИЯ. 1988. Т. 29, № 5

УДК 514.7

## РЕКУРРЕНТНЫЕ СООТНОШЕНИЯ ДЛЯ КОВАРИАНТНЫХ ПРОИЗВОДНЫХ БИЛОКАЛЬНЫХ ФУНКЦИЙ НА РИМАНОВОМ МНОГООБРАЗИИ

**А. Я. Родионов, А. Ю. Таранов**

*(НИИЯФ)*

**Получены рекуррентные соотношения для кратных ковариантных производных билокальных функций на римановом многообразии. Рассмотрены, в частности, геодезический интервал и оператор параллельного переноса; вычислены производные детерминанта Ван Флека до 5-го порядка включительно.**

**Введение.** Ковариантная трактовка теории поля в тех случаях, когда приходится иметь дело с неплоскими римановыми многообразиями либо в качестве координатного пространства (теория на фоне нетривиальной метрики), либо в качестве конфигурационного (например, обобщенная σ-модель), является весьма желательной. При исследовании в ковариантном формализме билокальных величин, таких как функции Грина или вариации действия [1—4], появляется необходимость учитывать в той или иной мере глобальные свойства соответствующего многообразия. Влияние глобальной геометрии на билокальные объекты, аргументы которых разнесены не очень далеко, можно попытаться учесть приближенно. В качестве примера такого подхода можно привести вычисление расходящихся частей диаграмм в методе фонового поля (см., напр., [1—3]).

В настоящей работе приведены рекуррентные формулы для крат-