МАТЕМАТИЧЕСКАЯ ОБРАБОТКА ГЕОДЕЗИЧЕСКОЙ ИНФОРМАЦИИ

УДК 528.011

ПОЛУЧЕНИЕ КОРРЕЛЯЦИОННЫХ ХАРАКТЕРИСТИК НА ОСНОВЕ УГЛА МЕЖДУ ПОДПРОСТРАНСТВАМИ

канд. техн. наук, доц. А.М. ДЕГТЯРЕВ, Е.В. ДЕГТЯРЕВА
(Полоцкий государственный университет)

Обоснована возможность получения известных корреляционных характеристик через угол между подпространствами. Показано, что такой подход является более общим по сравнению с традиционным, так как наряду с получением известных характеристик позволяет получить ряд новых, значительно расширяющих возможности анализа и требующих всестороннего изучения.

Корреляционный анализ результатов эксперимента получил достаточно широкое распространение в геодезической теории [1] и практике [2, 3]. Традиционно используют следующие коэффициенты корреляции:

- парный r_{ij} – показывает тесноту связи между i и j рядами, игнорируя влияние других;
- частный $r_{ij|k....}$ – показывает тесноту связи между i и j рядами при условии учета линейного влияния всех других рядов;
- множественный $r_{j|...}$ – показывает тесноту связи между i рядом и всеми остальными.

Наиболее удобный способ получения этих коэффициентов – через использование центрированной матрицы A^*, полученной из матрицы результатов эксперимента A (матрицы плана). Центрирование строится в виде отклонения от среднего по столбцам, каждого столбца из A. По центрированной матрице A^* вычисляется нормальная матрица $N = A^{*\mathrm{T}} \cdot A^*/n$ (матрица эмпирических моментов), и обратная к ней $Q = N^{-1}$. Тогда перечисленные выше коэффициенты корреляции в виде матриц (а не по одному) можно легко получить следующим образом:

- парные

$$R_1 = D \cdot N \cdot D. \tag{1}$$

Здесь $D = \begin{pmatrix} \frac{1}{\sqrt{n_{11}}} & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & \frac{1}{\sqrt{n_{nn}}} \end{pmatrix}$ – вспомогательная масштабирующая матрица из диагональных элементов матрицы N, n_{ii};

- частные

$$R_2 = -D' \cdot Q \cdot D', \tag{2}$$

т.е. совершенно эквивалентно по форме, как и парные, но по элементам обратной матрицы Q. Матрица D' также строится по элементам матрицы Q, аналогично D;

- множественные

$$R_3^2 = E - Q_d^{-1} \cdot N_d^{-1}. \tag{3}$$

Здесь Q_d, N_d – диагональные матрицы, выделенные из Q и N соответственно. Рассматривая внимательно парный коэффициент корреляции между рядами x и y

$$r_{xy} = \frac{\left[(x-\bar{x})\cdot(y-\bar{y})\right]}{\sqrt{\left[(x-\bar{x})^2\right]}\cdot\sqrt{\left[(y-\bar{y})^2\right]}} = \frac{\left[v_x \cdot v_y\right]}{\sqrt{\left[v_x^2\right]}\cdot\sqrt{\left[v_y^2\right]}} = \frac{\left(v_x, v_y\right)}{\left\|v_x\right\|\cdot\left\|v_y\right\|}, \tag{4}$$

несложно заметить известный факт, что коэффициент корреляции r_{xy} есть косинус угла между центрированными векторами v_x и v_y. В частности из рис. 1 следует, что $cos\,\beta = \dfrac{S}{\|v_x\|}$, откуда величина проекции есть

$$S = \frac{(v_x, v_y)}{\|v_y\|}. \tag{5}$$

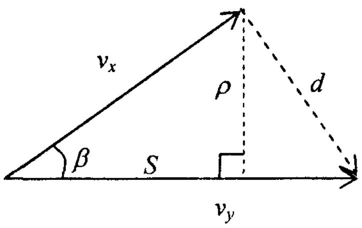

Рис. 1. Графическое представление парного коэффициента корреляции

С другой стороны, исходя из трактовки, что $r_{xy} = \cos\beta$ и рис. 1, коэффициент корреляции можно получить, используя теорему косинусов для треугольника с длинами сторон $\|v_x\|$, $\|v_y\|$, $\|d\| = \|v_y - v_x\|$:

$$r_{xy} = cos\,\beta = \frac{\|v_x\|^2 + \|v_y\|^2 - \|d\|^2}{2\|v_x\| \cdot \|v_y\|}. \tag{6}$$

Формула (6) может использоваться для многих случаев геодезической практики наряду с традиционной формулой (4). Достаточно важна для приложений и формула (5). Из рис. 1, на основании формул тригонометрии возможны и другие представления коэффициента корреляции, например, с применением теоремы о проекциях.

Очевидно, что векторы v_x и v_y можно трактовать как внутренние геометрические объекты или подпространства размерности $(n\times1)$ в общем пространстве $(n\times n)$. Исходя из этого появляется возможность трактовать меру статистической связи в виде коэффициента корреляции как косинус угла между более общими подпространствами, чем векторы.

Рассмотрим угол между двумя подпространствами в виде плоскостей (рис. 2). Для этого на пересечении AB их сечения возьмем произвольную точку C. Отложим от нее перпендикуляры CD и CE на одну и другую плоскости.

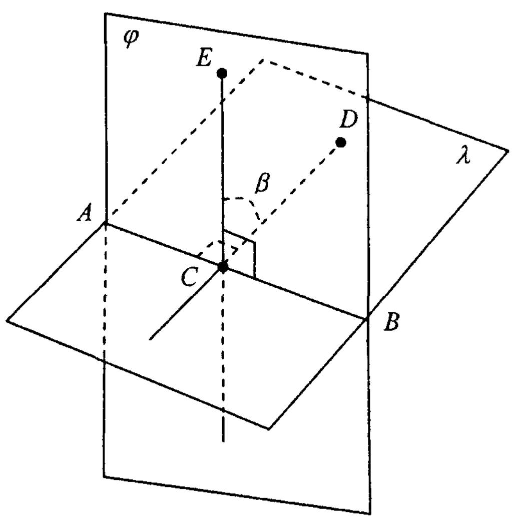

Рис. 2. Определение угла между плоскостями

Таким образом, угол β между плоскостями будет угол ECD, который является наименьшим из всех углов, образованных линией пересечения и перпендикулярами. Вдобавок ко всему, и это важно, угол β есть угол между осями ортогональных базисов ECB для плоскости φ и DCB для плоскости λ с одной общей осью CB, причем, опять наименьший. Базисы такого рода образуют базис для нового графического объекта $\varphi \bigcap \lambda$.

По аналогии с плоскостями, угол между любыми подпространствами может трактоваться как наименьший угол между осями в объединенном базисе для нового объекта $\Phi \bigcap \Lambda$. Процедуру его получения можно свести к следующим шагам:

1) центрирование векторов, описывающих процесс путем вычитания из элементов вектора среднего по этому вектору;
2) выделение в подпространствах Φ и Λ ортогональных базисов B_Φ и B_Λ;
3) объединение этих базисов в одну структуру по общим элементам $B = B_\Phi \bigcap B_\Lambda$;
4) выделение всех новых, полученных углов между осями разных базисов, кроме прямых;
5) нахождение минимального угла из всех углов, полученных в пункте 3.

Последовательность операций очень легко программируется в любом математическом пакете, имеющем в своем составе стандартную процедуру сингулярного разложения, например в среде MATLAB.

Из линейной алгебры известно, что выделить ортогональные базисы можно используя, например, сингулярное разложение, а объединение структур с общими частями произвести путем свертки по общим частям или перемножением транспонированной и прямой матриц задающих подпространства Φ и Λ $B_\Phi \bigcap B_\Lambda = B_\Phi^{\mathrm{T}} \cdot B_\Lambda$ [4].

Для случая анализа связи двух векторов v_x и v_y ортогональные базисы для них есть векторы их направляющих косинусов:

$$B_x = \frac{v_x}{\|v_x\|},$$

$$B_y = \frac{v_y}{\|v_y\|}.$$

Объединение в общую структуру дает формулу

$$B = B_x^{\mathrm{T}} \cdot B_y = \frac{v_x^{\mathrm{T}} \cdot v_y}{\|v_x\| \cdot \|v_y\|},$$

идентичную (4), т.е. сразу косинус угла между векторами v_x и v_y (так как он один), или коэффициент корреляции r_{xy}.

Рассмотрим все возможные комбинации векторов для определения тесноты связи между ними, посредством угла между подпространствами, которые эти комбинации образуют. Пусть процесс описан матрицей $A_{n \times k}$, состоящей из k вектор-столбцов состояний a_i из n элементов. Обозначим процедуру получения угла по описанной выше последовательности между подпространством, состоящим из векторов $\{a_{i_1}, a_{i_2}, \ldots\}$ и подпространством векторов $\{a_{j_1}, a_{j_2}, \ldots\}$, например как ***subspace*** ($\{a_{i_1}, a_{i_2}, \ldots\}, \{a_{j_1}, a_{j_2}, \ldots\}$). В первую очередь рассмотрим простейшие комбинации, когда одно из подпространств есть просто вектор a_i. Самая простая из них, когда рассматривается угол между двумя векторами a_i и a_j: ***subspace***(a_i, a_j). Косинус этого угла должен быть обычным парным коэффициентом корреляции r_{ij}.

Косинус угла между одним вектором a_i и группой других векторов, образующих подпространство $\{a_{i_1}, a_{i_2}, \ldots\}$, должен по определению представлять множественный коэффициент корреляции $r_{i|\{i_1, i_2, \ldots\}}$.

При определении множественного коэффициента корреляции нельзя чтобы вектор a_i принадлежал другому подпространству $\{a_{i_1}, a_{i_2}, \ldots\}$, иначе, естественно, угол будет равен нулю. Это обстоятельство может служить проверкой принадлежности вектора подпространству.

Следующая группа комбинаций подразумевает наличие общих частей в подпространствах. Если в одном и другом подпространстве не общие только два вектора a_i и a_j, то можно ожидать, что другие, общие, «компенсируют» друг друга. Косинус угла между такими подпространствами, по определению, будет частным коэффициентом корреляции $r_{ij|\ldots}$. Например, для процесса, определяемого тремя векторами $\{a_1, a_2, a_3\}$, это будет *cos*(***subspace***($\{a_1, a_2\}, \{a_2, a_3\}$)). Здесь общий вектор a_3, а не общие a_1 и a_2.

Определенная таким образом процедура должна давать частный коэффициент корреляции между векторами a_1 и a_2 при исключении («компенсации») влияния вектора a_3. Последовательность записи векторов в подпространстве не имеет значения.

Этим набором исчерпываются традиционные величины, характеризующие меру тесноты связи, известные как разного рода коэффициенты корреляции. Предложенный подход на основе понятия угла между подпространствами позволяет не только объединить их определение в едином алгоритме, но, что более важно, расширить, т.е. ввести некоторые новые меры тесноты связи. К ним можно отнести:

- меры на основе угла между двумя группами не пересекающихся (не имеющих общих) векторов;
- меры на основе угла между частично пересекающимися группами (т.е. имеющими несколько общих) векторов.

Так как в обоих случаях участвуют группы, то целесообразно такой класс мер тесноты связи назвать *групповым.* А так как эти меры не являются в привычном для нас смысле какими-либо коэффициентами корреляции, назовем их просто *коэффициентами связанности.*

Первую введенную меру можно трактовать как тесноту связи между подпространством $\{a_{i_1}, a_{i_2}, \ldots\}$ и подпространством, образованным всеми оставшимися векторами (или группой из них), описывающими процесс. Первое и второе подпространства не имеют общих векторов. Исходя из вышесказанного, новую введенную меру можно назвать *групповой множественный коэффициент связанности.*

Вторая мера в первом и втором подпространствах имеет общие векторы и группы не общих, чем похожа на частный коэффициент корреляции. Таким образом, её можно трактовать как тесноту связи между подпространством $\{a_{i_1}, a_{i_2}, \ldots\}$ и подпространством $\{a_{j_1}, a_{j_2}, \ldots\}$ при исключении влияния общих векторов, образующих также подпространство $\{a_{k_1}, a_{k_2}, \ldots\}$. При этом, что достаточно важно, мера может вычисляться при подпространстве пересечения состоящего из 1 вектора и до $n-2$ общих векторов. Такая процедура была или не возможна, или очень затруднена (в зависимости от соотношения подпространств общих и не общих векторов) при традиционном подходе.

Полученную меру можно назвать *групповой частный коэффициент связанности.*

Введение новых мер позволяет очень сильно расширить возможности анализа внутренних связей между структурами, описывающими исследуемые процессы в геодезии, экологии, других науках о Земле.

Проведем численные исследования предложенного подхода. Для этого используем результаты экологического мониторинга на территории санитарно-защитной зоны города Новополоцка по следующим параметрам:

- уровень грунтовых вод;
- характеристика солнечной активности в числах Вольфа;
- количество осадков за вегетационный период и за год;
- температура за вегетационный период; радиальный прирост поздней и ранней древесины ели.

Наблюдения брались за период 1975 – 2000 годов включительно. Вычислим парные, частные и множественные коэффициенты корреляции на основе традиционных формул (1), (2) и (3), а затем проверим их на основе получения углов между соответствующими подпространствами. В заключение получим значения некоторых групповых коэффициентов связанности.

Корреляционная матрица из (1) по данным таблицы есть

$$R_1 = \begin{vmatrix} 1.0000 & -0.3651 & -0.5391 & -0.5940 & 0.0024 & 0.1514 & 0.1876 \\ -0.3651 & 1.0000 & 0.1686 & 0.1646 & 0.0066 & 0.1030 & -0.1591 \\ -0.5391 & 0.1686 & 1.0000 & 0.8719 & -0.4174 & -0.3469 & -0.1659 \\ -0.5940 & 0.1646 & 0.8719 & 1.0000 & 0.2005 & -0.5127 & -0.2899 \\ 0.0024 & 0.0066 & -0.4174 & -0.2005 & 1.0000 & -0.0558 & -0.1055 \\ 0.1514 & 0.1030 & -0.3469 & -0.5127 & 0.0558 & 1.0000 & 0.8218 \\ 0.1876 & -0.1591 & -0.1659 & -0.2899 & -0.1055 & 0.8218 & 1.0000 \end{vmatrix}$$

матрица частных коэффициентов корреляции по (2):

$$R_2 = \begin{vmatrix} 1.0000 & -0.1236 & -0.1030 & -0.3366 & -0.2154 & -0.3052 & 0.2532 \\ -0.1236 & 1.0000 & 0.0175 & 0.1573 & 0.0553 & 0.4882 & -0.4809 \\ -0.1030 & 0.0175 & 1.0000 & 0.7452 & -0.4931 & 0.0551 & 0.0097 \\ -0.3366 & 0.1573 & 0.7452 & 1.0000 & 0.1833 & -0.4774 & 0.3109 \\ -0.2154 & 0.0553 & -0.4931 & 0.1833 & 1.0000 & -0.1148 & 0.0572 \\ -0.3052 & 0.4882 & 0.0551 & -0.4774 & -0.1148 & 1.0000 & 0.8757 \\ 0.2532 & -0.4809 & 0.0097 & 0.3109 & 0.0572 & 0.8757 & 1.0000 \end{vmatrix}$$

диагональная матрица из (3) множественных коэффициентов корреляции:

$$R_3 = \begin{vmatrix} 0.7068 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5954 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9104 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9262 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5744 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.9230 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.8938 \end{vmatrix}$$

Результаты экологического мониторинга по 7 составляющим

Уровень грунтовых вод	Числа Вольфа	Осадки за вегетационный период, мм	Осадки за год, мм	Температура за вегетационный период, °C	Поздняя древесина	Ранняя древесина
104,0	34	229	660	15,9	4,6	5,2
127,4	20	254	524	13,1	4,7	4,7
98,5	35	434	734	13,7	4,6	4,7
86,9	122	489	781	12,9	5	5
109,7	230	410	741	15,0	3,7	3,8
76,7	265	416	635	13,6	5,1	4,1
93,0	238	391	711	15,3	4,8	4,6
87,0	211	352	735	14,0	3,6	3,7
92,0	138	230	637	15,5	4,5	3,9
110,0	86	384	685	14,3	4,5	4,2
106,0	30	390	674	14,5	4,2	4,9
101,0	22	353	729	14,5	2,8	3,3
99,0	39	531	824	13,5	3,4	4,2
90,5	124	424	744	16,0	3,2	3,2
95,5	181	446	795	15,3	2,6	3,2
95,4	237	498	913	13,4	2,7	2,6
79,8	243	407	743	14,5	3,4	3,6
117,3	175	313	609	15,5	3,8	3,5
108,2	108	404	767	13,4	2,3	2,8
90,2	57	345	713	14,9	2	1,9
96,1	36	400	767	15,4	2,4	2,1
114,1	15	310	582	14,4	3,6	4,3
103,0	26	339	659	14,7	4,2	5,7
73,2	92	602	1014	14,7	3,3	4,6
110,7	138	183	539	16,1	5,2	4,8
109,0	205	374	717	13,8	4,4	5

Вычислим парные коэффициенты корреляции, используя понятие угла между подпространствами по процедуре $\boldsymbol{subspace}(\{a_i\},\{a_j\})$, полученной на основе определенной выше последовательности операций. В качестве подпространств $\{a_i\}$ и $\{a_j\}$ необходимо взять центрированные столбцы $a_i - \bar{a}_i$ исходных данных из табл. 1. В результате имеем:

$\boldsymbol{subspace}(\{a_1\},\{a_2\}) = 0.3651$, $\boldsymbol{subspace}(\{a_1\},\{a_3\}) = 0.5391, \ldots,$

$\boldsymbol{subspace}(\{a_4\},\{a_5\}) = 0.2005$, $\boldsymbol{subspace}(\{a_6\},\{a_7\}) = 0.8218,$

что численно совпадает с полученными традиционно, но имеются некоторые проблемы со знаком. На основе предложенного подхода возможно вычислить групповые коэффициенты связанности, например, между подпространствами $\{a_1\, a_2\},\{a_3\, a_4\}$ как $\boldsymbol{subspace}(\{a_1\, a_2\},\{a_3\, a_4\}) = 0.0418$ и другими неперекрывающимися подпространствами, чего нельзя сделать обычными методами, а польза от них для всестороннего анализа процесса очевидна.

Для вычисления частных коэффициентов корреляции по приведенным выше правилам в подпространствах должны быть только a_i и a_j (не общие вектора), а те, влияния которых требуется исключить, в качестве общих:

subspace$(\{a_1\ a_3\ a_4\ a_5\ a_6\ a_7\},\{a_2\ a_3\ a_4\ a_5\ a_6\ a_7\}) = 0.1236$, …,

subspace$(\{a_1\ a_2\ a_3\ a_4\ a_5\ a_7\},\{a_1\ a_2\ a_3\ a_4\ a_5\ a_6\}) = 0.8757$,

что численно совпадает с полученными выше в виде матрицы. Но предложенный метод позволяет быстро регулировать число векторов, влияние которых требуется исключить:

subspace$(\{a_1\ a_3\ a_4\ a_5\},\{a_2\ a_3\ a_4\ a_5\}) = 0.3270$ – исключено влияние только a_3, a_4 и a_5;

subspace$(\{a_1\ a_3\},\{a_2\ a_3\}) = 0.3303$ – исключено влияние только a_3 и так далее, что также повышает качество и детальность анализа.

При определении множественного коэффициента корреляции в качестве подпространств берутся один вектор a_i и все другие, влияние которых на первый оценивается в совокупности:

subspace$(\{a_1\},\{a_2\ a_3\ a_4\ a_5\ a_6\ a_7\}) = 0.7068$ (см. матрицу R_3), но легко можно регулировать количество векторов влияния ***subspace***$(\{a_1\},\{a_2\ a_3\ a_4\ a_5\}) = 0.6690$, а также брать несколько векторов в первом подпространстве (см. пример для парного коэффициента корреляции).

Таким образом, можно сделать вывод, что оценка тесноты связей через разного рода коэффициенты связанности, получаемые на основе угла между подпространствами, позволяет значительно как количественно, так и качественно расширить возможности анализа структуры исследуемых процессов. Возможно в рамках единого подхода получить все известные меры тесноты связи между частями процесса, описываемого векторами, а также ряд других, ранее не известных, названных групповыми коэффициентами связанности.

ЛИТЕРАТУРА

1. Кемниц Ю.В. Математическая обработка зависимых результатов измерений.-М.: Недра, 1970.- 190 с.
2. Лукьянов В.Ф. К вопросу определения и предвычисления оценок коэффициента корреляции // Известия вузов. Геодезия и аэрофотосъемка. - 1969. - № 5. - С. 41 - 44.
3. Сундаков Я.А. Геодезические работы при возведении крупных промышленных сооружений и высотных зданий. - М.: Недра, 1980. - 343 с.
4. Воеводин В.В., Кузнецов Ю.А. Матрицы и вычисления. - М.: Наука, 1984. - 320 с.