

Черевко Євген Володимирович

*к.ф.-м.н., доц., доцент кафедри кібербезпеки,
Національний університет «Одеська юридична академія»*

Чепурна Олена Євгенівна

*к.ф.-м.н., доц., доцент кафедри кібербезпеки,
Національний університет «Одеська юридична академія»*

Слатвінська Валерія Миколаївна

*PhD в галузі «Право», асистент кафедри кібербезпеки,
Національний університет «Одеська юридична академія»*

ІНФОРМАЦІЙНА ГЕОМЕТРІЯ: КІБЕРБЕЗПЕКА ТА ЗАХИСТ ІНФОРМАЦІЇ

DOI <https://doi.org/10.36059/978-966-397-537-5-5>

Актуальність теми. Із зростанням складності цифрових систем та широким використанням високовимірних моделей обробки інформації актуалізується проблема формалізованого підходу до аналізу безпеки інформаційних процесів. Більшість прикладних систем захисту базується на евристичних або сигнатурних методах, які демонструють обмежену ефективність за умов змінних розподілів даних, нових типів атак або нестабільного інформаційного середовища. У цьому контексті виникає необхідність пошуку альтернативних методів аналізу інформації таких, наприклад, як інформаційна геометрія, яка виступає в якості засобу побудови внутрішньо узгоджених моделей інформаційного простору.

Згідно з аналітичними звітами NIST, кібератаки, такі як розподілені відмови в обслуговуванні, цільові вторгнення та атаки на ланцюги постачання, становлять значну загрозу для інформаційних систем [1].

Інформаційна геометрія як міждисциплінарна галузь формалізує поняття статистичного многовиду, наділеного метричною структурою Фішера та узгодженою з нею афінною зв'язністю. Застосування таких структур дозволяє описувати розподіли ймовірностей у термінах диференціальної геометрії, аналізувати кривину параметричних

просторів, а також визначати напрямки локальних змін систем, індукованих зовнішнім впливом. У статистичних задачах, пов'язаних з виявленням аномалій, оцінкою ризиків, оптимізацією байєсівських процедур та моделюванням стійкості до атак, інформаційна геометрія надає інструменти для кількісного аналізу з урахуванням як локальних, так і глобальних властивостей розподілів.

Попри теоретичну опрацьованість тематики в класичних роботах Амарі, Нагаоки та Ватанабе, прикладне застосування інформаційної геометрії в галузі кібербезпеки залишається недостатньо розвиненим. Існують поодинокі дослідження, що демонструють ефективність використання геометричних метрик для виявлення змін у мережевому трафіку, оцінки відстаней між гіпотезами в задачах класифікації, а також формалізації криптографічних моделей у термінах ріманових многовидів. Однак системна інтеграція цих методів у задачі захисту інформації потребує розробки нових математичних конструкцій, адаптованих до структурного і стохастичного характеру кіберзагроз.

В Україні законодавство, зокрема Закон «Про основні засади забезпечення кібербезпеки України» (2017), визначає кібербезпеку як необхідну умову захисту національних інтересів у цифровому просторі [2]. У цьому контексті інформаційна геометрія, як математична дисципліна, що поєднує диференціальну геометрію, теорію інформації та статистичний аналіз, пропонує перспективні інструменти для моделювання інформаційних процесів, аналізу кіберзагроз і розробки стійких систем захисту.

Огляд сучасного стану досліджень. Теоретичні основи інформаційної геометрії сформовано в працях Амарі, де вперше систематизовано концепцію статистичного многовиду та інформаційних дивергенцій, таких як дивергенція Кулбека-Ляйблера [3]. У контексті кібербезпеки ці методи застосовуються для аналізу інформаційних потоків і виявлення аномалій. Наприклад, Картер і Стрейле досліджували використання метрик Фішера для ідентифікації відхилень у мережевому трафіку, що дозволяє виявляти кібератаки з високою точністю [4]. Черевко Є., Чепурна О. та Кулешова Є. розвинули ці ідеї, запропонувавши геометричні методи для аналізу складних даних [5].

У криптографії інформаційна геометрія знаходить застосування в аналізі квантових систем, як показано в роботі Нільсена та Чуанга,

де геометричні підходи використовуються для оцінки стійкості квантових протоколів [6]. Сучасні дослідження, такі як робота Лі та співавторів, демонструють потенціал інформаційної геометрії для розробки постквантових криптографічних алгоритмів [7].

Аналіз великих даних є ще одним важливим напрямом. Ватанабе запропонував геометричні методи для оптимізації байєсівських алгоритмів, які застосовуються в задачах кібербезпеки [8]. Чен та співавтори показали, як інформаційна геометрія сприяє виявленню прихованих закономірностей у великих масивах даних, що підвищує ефективність систем раннього попередження атак [9].

В Україні Коваленко та співавтори наголошують на необхідності математичних методів, зокрема інформаційної геометрії, для захисту критичної інфраструктури [10], тоді як Петренко досліджує геометричне моделювання кібератак у умовах невизначеності [11].

Інтеграція інформаційної геометрії з алгоритмами штучного інтелекту, як показано в роботі Сміта та співавторів, відкриває нові можливості для адаптивного захисту IoT-систем [12]. Проте залишаються виклики, пов'язані з обчислювальною складністю геометричних алгоритмів і відсутністю стандартизованих підходів до їх реалізації в системах кібербезпеки.

Обґрунтування вибору тематики. Інформаційна геометрія вибрана як основа дослідження через її здатність представляти інформаційні процеси у вигляді геометричних структур, що забезпечує точний аналіз складних даних і розробку адаптивних методів захисту. Традиційні підходи, такі як статистичний аналіз або сигнатурні методи, часто є недостатньо ефективними для протидії новим типам кіберзагроз, зокрема атакам нульового дня чи цілеспрямованим вторгненням. Геометричні методи дозволяють моделювати розподіли ймовірностей як точки на многовидах, вимірюючи інформаційні відстані, що є особливо цінним для аналізу великих даних, криптографії та оцінки ризиків. В Україні актуальність дослідження посилюється потребою в захисті критичної інфраструктури та протидії гібридним загрозам у цифровому просторі.

Постановка задачі та дослідницька проблема. Основна проблема полягає в обмеженій інтеграції методів інформаційної геометрії в системи кібербезпеки, що знижує їхню здатність протистояти

складним кіберзагрозам. Відсутність систематизованих підходів до використання геометричних методів для аналізу інформаційних потоків, моделювання ризиків і розробки криптографічних систем створює потребу в нових теоретичних і практичних розробках. Задача дослідження полягає в розробці математичних моделей і алгоритмів на основі інформаційної геометрії для підвищення ефективності захисту інформації, зокрема в задачах аналізу даних, виявлення аномалій і створення стійких криптографічних систем.

Об'єкт і предмет дослідження. Об'єктом дослідження є процеси забезпечення кібербезпеки в інформаційних системах, включаючи аналіз даних, виявлення кіберзагроз і захист інформації. Предметом дослідження є методи інформаційної геометрії, зокрема статистичні многовиди, дивергенції та метрики, застосовані до моделювання інформаційних процесів і розробки систем кібербезпеки.

Мета дослідження. Метою даного дослідження є формалізація підходів інформаційної геометрії у задачах моделювання та аналізу кібербезпеки.

Завдання дослідження.

1. Систематизувати математичні основи інформаційної геометрії, включаючи теорію статистичного многовиду та інформаційних дивергенцій, для їх застосування в кібербезпеці.
2. Розробити геометричні методи аналізу інформаційних потоків щоб в перспективі виявляти аномалії у мережевому трафіку.
3. Оцінити потенціал інтеграції інформаційної геометрії з алгоритмами штучного інтелекту.
4. Проаналізувати обчислювальні обмеження геометричних методів і розробити способи їх оптимізації.

1. Основи Ріманової геометрії

1.1. Тензори

Нехай задано тривимірний векторний простір V , в якому ми оберемо ортонормований базис e_1, e_2, e_3 . Розглянемо довільний вектор $\vec{a}$. Його можна розкласти за векторами базису:

$$\vec{a} = a^1 e_1 + a^2 e_2 + a^3 e_3 = \sum_{i=1}^3 a^i e_i.$$

Координатні індекси для векторів, з причин, що пояснимо нижче – будемо писати згори справа від головного символу. Зокрема, це дозволяє застосовувати так зване *правило Ейнштейна*. Воно полягає в тому, що поява однакового індексу згори і внизу означає підсумовування за цим індексом в межах його змінення. Тобто розкладання вектору $\vec{a}$ за базисом можна записати так:

$$\vec{a} = a^i e_i, (i = 1, 2, 3).$$

Тепер розглянемо скалярний добуток двох векторів $\vec{a} = a^i e_i$ та $\vec{b} = b^j e_j$:

$$\vec{a} \cdot \vec{b} = a^1 b^1 + a^2 b^2 + a^3 b^3 = \sum_{i=1}^3 a^i b^i.$$

Зауважимо зокрема, що в цьому випадку ми не можемо застосовувати правило Ейнштейна, оскільки обидва індекси є верхні. Втім, можна використати одиничну матрицю, так звану *дельту Кронекера*:

$$\delta_{ij} = \begin{cases} 1, \text{ якщо } i = j \\ 0, \text{ якщо } i \neq j. \end{cases}$$

Тоді скалярний добуток векторів $\vec{a} = a^i e_i$ та $\vec{b} = b^j e_j$ може бути записаний так:

$$\vec{a} \cdot \vec{b} = a^i \delta_{ij} b^j \quad (i, j = 1, 2, 3).$$

Тепер виконаємо лінійне перетворення нашого базису, використовуючи невироджену матрицю $C = (C_i^j)$, $\det C \neq 0$:

$$e'_1 = C_1^1 e_1 + C_1^2 e_2 + C_1^3 e_3;$$

$$e'_2 = C_2^1 e_1 + C_2^2 e_2 + C_2^3 e_3;$$

$$e'_3 = C_3^1 e_1 + C_3^2 e_2 + C_3^3 e_3.$$

Або коротше:

$$e'_i = C_i^j e_j \quad (i, j = 1, 2, 3).$$

Надалі ми не будемо кожного разу виписувати діапазон зміни індексів, якщо останній буде очевидним з контексту. Також, зауважимо, що можна з міркувань зручності перейменовувати індекс за яким ведеться підсумовування, наприклад $\vec{a} = a^i e_i = a^k e_k$. Про такий індекс іноді кажуть що він «німий».

В новому базисі розкладання вектору $\vec{a}$ виглядатиме таким чином:

$$\vec{a} = a^{i'} e'_i.$$

Але вектор $\vec{a}$ (вектор $\vec{b}$ також) лишився тим самим, тобто

$$a^i e_i = a^{i'} e'_i = a^{i'} C_i^j e_j.$$

Це означає що відповідно мають змінитися і координати його розкладання.

Перейменовуючи німі індекси маємо:

$$a^i e_i = a^k e_k = a^{i'} C_i^j e_j = a^{i'} C_i^k e_k,$$

або

$$a^k e_k = a^{i'} C_i^k e_k.$$

Враховуючи лінійну незалежність векторів базису, маємо:

$$a^k = a^{i'} C_i^k.$$

Фактично це є скороченим записом матричного добутку:

$$\begin{pmatrix} a^1 \\ a^2 \\ a^3 \end{pmatrix} = \begin{pmatrix} C_1^1 & C_2^1 & C_3^1 \\ C_1^2 & C_2^2 & C_3^2 \\ C_1^3 & C_2^3 & C_3^3 \end{pmatrix} \begin{pmatrix} a^{1'} \\ a^{2'} \\ a^{3'} \end{pmatrix}.$$

Зауважимо, що на відміну від матричної форми, у скороченій формі запису, так званій *тензорній* – ми можемо міняти порядок множників, не боячись втрати сенсу:

$$a^k = C_i^k a^{i'}.$$

З цього випливає, що

$$\begin{pmatrix} a^{1'} \\ a^{2'} \\ a^{3'} \end{pmatrix} = \begin{pmatrix} C_1^1 & C_2^1 & C_3^1 \\ C_1^2 & C_2^2 & C_3^2 \\ C_1^3 & C_2^3 & C_3^3 \end{pmatrix}^{-1} \begin{pmatrix} a^1 \\ a^2 \\ a^3 \end{pmatrix},$$

де C^{-1} є оберненою матрицею до C .

$$a^{i'} = \left(C_k^{i'} \right)^{-1} a^k.$$

Тобто ми бачимо, що координати векторів при зміні базису змінюються *контраваріантно* до зміни векторів базису. Такі координати ми будемо називати *контраваріантними*, i відповідні верхні індекси також називатиме *контраваріантними*. Нижні індекси мають назву *коваріантних*. Надалі індекси що відповідають координатам в новому базисі будемо позначати як штриховані. Матрицю заміни базису позначатимемо:

$$C = C_{k'}^i.$$

Зворотню до неї, відповідно:

$$C^{-1} = C_k^{i'}.$$

Зрозуміло, що

$$C_{k'}^i C_j^{k'} = \delta_j^i, C_k^{i'} C_j^k = \delta_{j'}^{i'}.$$

Тут δ_j^i – вже знайомий нам *символ Кронекера*:

$$\delta_j^i = \begin{cases} 1, & \text{якщо } i = j \\ 0, & \text{якщо } i \neq j. \end{cases}$$

Пояснимо, чому в попередньому випадку, коли ми записували через дельту скалярний добуток, то обидва індекси були представлені як коваріантні. Справа в тому, що при заміні базису, скалярний добуток двох векторів

$\vec{a} \cdot \vec{b} = a^i \delta_{ij} b^j$ має лишатися незмінним, навіть у випадку, коли новий базис не буде ортонормованим. Введемо позначення для нашого початкового базису:

$$g_{ij} = \delta_{ij},$$

тоді:

$$\vec{a} \cdot \vec{b} = a^i g_{ij} b^j.$$

В новому базисі скалярний добуток має виглядати аналогічним чином:

$$\vec{a} \cdot \vec{b} = a^{i'} g_{i'j'} b^{j'},$$

але матриця g_{ij} , взагалі кажучи, вже не буде одиничною. Продовжуємо:

$$\vec{a} \cdot \vec{b} = a^{i'} g'_{i'j'} b^{j'} = a^k C_k^{i'} g'_{i'j'} C_l^{j'} b^l = a^k g_{kl} b^l.$$

Отже ми бачимо, що

$$C_k^{i'} g'_{i'j'} C_l^{j'} = g_{kl},$$

або

$$g'_{i'j'} = C_{i'}^k g_{kl} C_{j'}^l.$$

Об'єкт позначається як g_{kl} називають двічі коваріантним тензором. Якщо такий об'єкт визначає скалярний добуток, то він має назву *метричного тензору*, або просто *метрики*. У випадку ортонормованого базису:

$$g_{ij} = \delta_{ij}.$$

Взагалі кажучи, враховуючи властивості скалярного добутку g_{ij} є невід'ємною симетричною матрицею, що визначає додатньо означену квадратичну форму, тобто таку, що для будь-якого ненульового вектора $\vec{a}$:

$$a^i g_{ij} a^j > 0.$$

Узагальнюючи поняття вектору, метричного тензору для n -вимірного простору, можна ввести більш загальне поняття p -контраваріантного та q -коваріантного тензора, або тензору типу (p, q) .

Означення. Тензором типу (p, q) ми називаємо геометричний об'єкт, що визначається в кожному координатному базисі лінійного простору сукупністю чисел:

$$T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p}$$

(кожен з індексів $i_1 i_2 \dots i_p; j_1 j_2 \dots j_q$ приймає незалежно один від одного усі значення від 1 до n) і змінюється при переході від одного базису $e_1, e_2, \dots, e_n$ до іншого $e_1', e_2', \dots, e_n'$ ($e_i' = C_i^{j'} e_j$) відповідно до наступного закону:

$$T_{j_1' j_2' \dots j_q'}^{i_1' i_2' \dots i_p'} = T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p} C_{j_1}^{i_1'} C_{j_2}^{i_2'} \dots C_{j_1}^{i_p'} C_{j_1'}^{j_1} C_{j_2'}^{j_2} \dots C_{j_q'}^{j_q}. \quad (1.1)$$

Фактично, якщо, наприклад, g_{ij} є квадратною матрицею розмірності $(n \times n)$, то тензор $T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p}$ є $(p+q)$ -мірною «гіперкубічною» матрицею розмірності $(\underbrace{n \times n \times \dots \times n}_{p+q \text{ разів}})$.

Варто додати, що ми можемо для матриці g_{ij} , оскільки вона не є виродженою, знайти обернену. Її позначають як g^{ij} . Очевидно, що

$$g_{ik}g^{kj} = \delta_i^j.$$

За допомогою цих двох матриць можна робити операції «опускання» та «піднімання» індексів. Наприклад, взявши контраваріантний вектор a^i , опустивши індекс ми отримуємо *коваріантний вектор*, або просто *ковектор*:

$$a^k g_{kj} = a_j.$$

Ковектор можна витлумачити як лінійну функцію на множині векторів. Наприклад, якщо нам заданий ковектор f_i , то це означає, що на множині векторів n – вимірного простору задано лінійну функцію, що побудовано наступним чином:

$$f(x) = (f, x) = f_i x^i = f_1 x^1 + f_2 x^2 + \dots + f_n x^n.$$

Піднімання індексу робиться за допомогою матриці g^{ij} . Наприклад для ковектору f_i ми можемо побудувати відповідний вектор:

$$f_i g^{ij} = f^j.$$

Якщо ми називаємо об'єкт «вектором», то найчастіше маємо на увазі саме контраваріантний вектор. Зрозуміло, що у випадку ортонормованого базису, контраваріантні і коваріантні компоненти одного того ж самого об'єкту співпадатимуть.

Також, надалі нам може стати в пригоді операція так званої *згортки* тензору. Згортка перетворює типу (p, q) у тензор типу $(p-1, q-1)$. Наприклад,

$$S_{j_1 j_3 \dots j_q}^{i_2 i_3 \dots i_p} = T_{j_1 j_3 \dots j_q}^{a i_2 i_3 \dots i_p} = \sum_{a=1}^n T_{j_1 j_3 \dots j_q}^{a i_2 i_3 \dots i_p}.$$

Як ми бачимо, новий тензор $S_{j_1 j_3 \dots j_q}^{i_2 i_3 \dots i_p}$ отримано шляхом згортки за першим контраваріантним та другим коваріантним індексами. Взагалі кажучи, операцію згортки можна здійснювати по будь-якій парі контраваріантного та і коваріантного індексів. Головне, щоб в парі індексів – один був коваріантним а другий контраваріантним. Якщо цю вимогу порушувати – то отриманий внаслідок згортки об'єкт вже не буде тензором, оскільки закон перетворення координат тензору (1.1) вже не буде виконуватись.

1.2. Гладкі многовиди

Перш за все зазначимо, що вважатимемо усі функції тут і надалі диференційованими потрібну кількість разів, або навіть аналітичними, тобто такими, чий степеневі ряди збігаються до саме цих функцій, принаймні локально.

Для подальшого засвоєння цілком достатньо розуміти, що об'єктом дослідження ріманової геометрії є n -вимірний простір, тобто множина, на якій положення кожної точки задається набором з n чисел. Цей параграф призначений лише для обґрунтування можливості введення системи координат на цьому просторі, тому при першому читанні його можна пропустити. Додамо, що в подальшому викладення терміни «простір» і «многовид» використовуватимуться як синоніми.

Як пише французький математик Елі Картан [16], дуже важко дати загальне означення поняття многовиду. Але воно є фактично центральним поняттям геометрії, тому зараз ми спробуємо надати деяке уявлення про нього, так щоб не обтяжувати викладення аксіоматикою. Як зауважують Амарі і Нагаока, многовид M – це множина на якій задано координатну систему. Елементами цієї множини є точки. Наявність координатної системи на многовиді M означає взаємно-однозначне відображення M (або його підмножини) на R^n , яке дозволяє кожній точці p співставити впорядкований набір n дійсних чисел – координат точки (рис. 1).

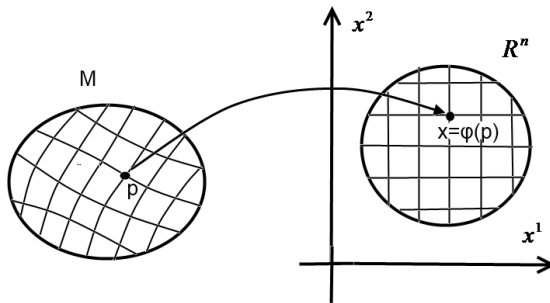


Рисунок 1 – Локальна координатна система на многовиді M

Нехай M є многовидом і $\varphi: M \rightarrow R^n$ є координатною системою на M . Тоді, φ відображує кожную точку $p \in R^n$ на n дійсних чисел [13]:

$$\varphi(p) = [x^1(p), \dots, x^n(p)] = [x^1, \dots, x^n].$$

Вони є координатами точки p . Кожне x^i може розглядатися як функція $p \rightarrow x^i(p)$, що відображує точку p на її i -ту координату. Ми будемо називати функції $x^i: M \rightarrow R$ *координатними функціями*. Відображення φ має назву *карти многовиду*. Але на жаль на завжди вдається обійтися однією картою, так щоб воно було взаємно однозначним. Наприклад, візьмемо систему географічних координат земної поверхні, а саме: θ – географічна широта, ψ – географічна довгота то раптом виявиться, що точка полюса, наприклад північного описується як координатами $\theta = \frac{\pi}{2}$; $\psi = 0$, так і наприклад, $\theta = \frac{\pi}{2}$; $\psi = \frac{\pi}{4}$. Тобто, точка полюса відображується у цілий відрізок площини R^2 : $\theta = \frac{\pi}{2}$; $\psi \in [0; 2\pi]$. Тому множину M покривають відкритими множинами U_a , тобто

$$M = \bigcup_a U_a.$$

На кожній множині U_a ми задаємо функцію φ_a , так щоб відображення $\varphi_a: U_a \rightarrow R^n$ було взаємно однозначним. Разом вони утворюють карту (U_a, φ_a) . Множина усіх карт має назву *атласу* многовиду. Як ми бачимо, в загальному випадку ми маємо справу не з однією картою, а *множиною*, що складається з багатьох карт. Власне, тому M і має назву *многовиду*.

Також, атлас многовиду має відповідати ще одній вимозі, а саме: якщо $U_\alpha \cap U_\beta \neq \emptyset$, то існує гладке відображення:

$$\varphi_\beta \circ \varphi_\alpha^{-1}: \varphi_\alpha(U_\alpha \cap U_\beta) \rightarrow \varphi_\beta(U_\alpha \cap U_\beta)$$

однієї області R^n на іншу. Фактично, остання формула описує перехід від однієї карти многовиду до іншої. Зауважимо, що надалі ми будемо працювати *локально*, тобто розглядати геометричні властивості на відкритій множині $U \subset M$, у межах, в яких

відображення $\varphi: M \rightarrow R^n$ на відкриту множину $\varphi(U)$ є неперервним. Зрозуміло, що такою множиною із обраною системою координат $(U_\alpha, \varphi_\alpha)$ може бути одна з карт атласа многовиду M .

Далі розглянемо функції на многовиді. Нехай $f: M \rightarrow R$ є функцією, що є визначеною на многовиді. Якщо ми вже обрали деяку координатну систему $\varphi = [x^1, \dots, x^n]$ для M , то ця функція має бути представленою як функція координат, тобто якщо $[x^1, \dots, x^n] \in$ координатами точки p , то $f(p) = \bar{f}(x^1, \dots, x^n)$, де $\bar{f} = f \circ \varphi^{-1}$. Таким чином, $\bar{f}$ є такою, що приймає дійсні значення, а областю визначення є якась відкрита підмножина R^n . Якщо $\bar{f}(x^1, \dots, x^n)$ є диференційовною в кожній точці $\varphi(U)$, то частинна похідна $\frac{\partial}{\partial x^i} \bar{f}(x^1, \dots, x^n)$ буде також функцією на $\varphi(U)$. Таким чином, ми можемо визначити частинні похідні для f :

$$\frac{\partial f}{\partial x^i} \stackrel{\text{def}}{=} \frac{\partial \bar{f}}{\partial x^i} \circ \varphi: M \rightarrow R.$$

Аналогічно можна визначити похідні вищих порядків для функцій заданих на многовиді M .

1.3. Дотичні вектори і дотичні простори

Нехай $[x^1, \dots, x^n]$ – деяка система координат на многовиді M . Розглянемо i -ту координатну криву, що утворюється при фіксації усіх координат x^j крім x^i , тобто $i \neq j$, і проходить крізь точку p (рис. 2).

Позначимо дотичний вектор до координатної кривої x^i як e_i . Таким чином ми отримуємо дотичний простір T_p , породжений векторами $e_1, e_2, \dots, e_n$. І якщо точка $p' \in$ «дуже близькою» до p , $[x^i]$ і $[x^i + dx^i]$ є відповідно координатами p та p' , то з'єднуючи ці точки отримуємо вектор $\overrightarrow{pp'} = dx^i e_i$. Він є елементом отриманого дотичного простору T_p .

Тепер розглянемо заміну системи координат на M :

$$x^i = x^i(x^{1'}, x^{2'} \dots x^{n'}), (i = 1, 2, \dots, n).$$

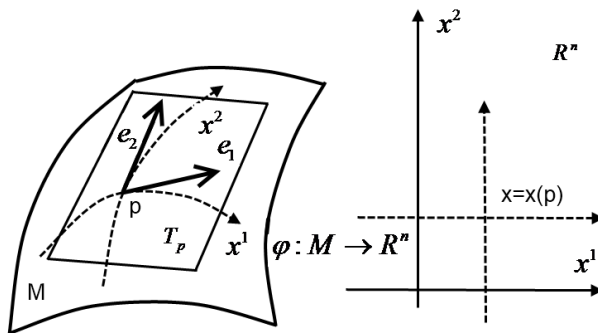


Рисунок 2 – Відображення дотичного до многовиду M простору на $\mathbb{R}^n$

Тоді лінійні перетворення векторів базису будуть здійснюватися за допомогою матриці Якобі:

$$\left(\frac{\partial x^i}{\partial x^{i'}} \right) = \begin{pmatrix} \frac{\partial x^1}{\partial x^{1'}} & \dots & \frac{\partial x^1}{\partial x^{n'}} \\ \vdots & \ddots & \vdots \\ \frac{\partial x^n}{\partial x^{1'}} & \dots & \frac{\partial x^n}{\partial x^{n'}} \end{pmatrix}.$$

Тобто, ці вектори базису e_i будуть перетворюватися за формулою:

$$e_{i'} = \frac{\partial x^i}{\partial x^{i'}} e_i.$$

Очевидно, що матриця $\left(\frac{\partial x^i}{\partial x^{i'}} \right)$ є невинродженою, тобто $\det \left(\frac{\partial x^i}{\partial x^{i'}} \right) \neq 0$.

Обернена до неї матриця виглядатиме так:

$$\left(\frac{\partial x^{i'}}{\partial x^i} \right) = \begin{pmatrix} \frac{\partial x^{1'}}{\partial x^1} & \dots & \frac{\partial x^{1'}}{\partial x^n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x^{n'}}{\partial x^1} & \dots & \frac{\partial x^{n'}}{\partial x^n} \end{pmatrix}.$$

Легко бачити, що

$$\frac{\partial x^k}{\partial x^{i'}} \frac{\partial x^{i'}}{\partial x^j} = \delta_j^k, \quad \frac{\partial x^{k'}}{\partial x^i} \frac{\partial x^i}{\partial x^{j'}} = \delta_{j'}^{k'}.$$

Якщо в дотичному лінійному просторі T_p заданий вектор $\vec{a} = a^i e_i$, то його координати змінюватимуться контраваріантно:

$$a^{i'} = \frac{\partial x^{i'}}{\partial x^i} a^i.$$

Врахуємо, що оскільки наш многовид M є гладким, то дотичний простір існує в кожній точці що лежить принаймні в межах деякої карти многовиду. Також, в цих межах може бути задане векторне поле $a^i(x)$, де $x = (x^1, x^2 \dots x^n)$. Тоді

$$a^{i'}(x') = \frac{\partial x^{i'}}{\partial x^i} a^i(x),$$

де $x' = (x^{1'}, x^{2'} \dots x^{n'})$.

Також, у кожній точці для відповідного дотичного простору задамо метричний тензор:

$$g_{ij}(x) = e_i(x) \cdot e_j(x).$$

Його компоненти є скалярними добутками базисних векторів дотичного простору у точці x .

Для цього двічі коваріантного тензору маємо:

$$g_{i'j'}(x') = \frac{\partial x^i}{\partial x^{i'}} g_{ij}(x) \frac{\partial x^j}{\partial x^{j'}}.$$

Узагальнюючи маємо таке означення.

Означення 2. Тензорним полем (або просто тензором) типу (p, q) ми називаємо геометричний об'єкт, що визначається в кожному координатному базисі лінійного простору сукупністю чисел

$$T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p}(x)$$

(кожен з індексів $i_1 i_2 \dots i_p; j_1 j_2 \dots j_q$ приймає незалежно один від одного усі значення від 1 до n) і змінюється при переході від однієї

системи координат $x^1, x^2 \dots x^n$ до іншої $x^{1'}, x^{2'} \dots x^{n'}$ (за формулами $x^i = x^i(x^{1'}, x^{2'} \dots x^{n'})$) відповідно до наступного закону:

$$T_{j_1' j_2' \dots j_g'}^{i_1' i_2' \dots i_p'}(x') = T_{j_1 j_2 \dots j_g}^{i_1 i_2 \dots i_p}(x) \frac{\partial x^{i_1'}}{\partial x^{j_1}} \frac{\partial x^{i_2'}}{\partial x^{j_2}} \dots \frac{\partial x^{i_p'}}{\partial x^{j_p}} \frac{\partial x^{j_1}}{\partial x^{j_1'}} \frac{\partial x^{j_2}}{\partial x^{j_2'}} \dots \frac{\partial x^{j_g}}{\partial x^{j_g'}}. \quad (1.2)$$

1.4. Коваріантне диференціювання та зв'язність

Нехай на многовиді M задано векторне поле $\vec{a}(x)$. Це означає, що в кожній точці воно може бути представленим як $\vec{a}(x) = a^i(x) e_i(x)$. Отже, як координати вектора так і вектори базису дотичного простору є функціями координат $x = (x^1, x^2 \dots x^n)$, заданих на многовиді.

Знайдемо частинну похідну поля $\vec{a}(x)$ за змінною x^k :

$$\frac{\partial \vec{a}(x)}{\partial x^k} = \frac{\partial a^i(x)}{\partial x^k} e_i(x) + a^i(x) \frac{\partial e_i(x)}{\partial x^k}.$$

Центральним пунктом цих міркувань є те, що у загальному випадку криволінійних координат вектори базису $e_1(x), e_2(x), \dots, e_n(x)$ не є незмінними: від точки до точки вони повертаються і змінюють свої абсолютні величини. Очевидно $\frac{\partial e_i(x)}{\partial x^k}$ є вектором і він може бути розкладений за векторами базису:

$$\frac{\partial e_i(x)}{\partial x^k} = \Gamma_{ik}^j(x) e_j(x).$$

Таким чином

$$\frac{\partial \vec{a}(x)}{\partial x^k} = \frac{\partial a^i(x)}{\partial x^k} e_i(x) + a^i(x) \Gamma_{ik}^j(x) e_j(x).$$

Перейменовуючи німі індекси, маємо:

$$\frac{\partial \vec{a}(x)}{\partial x^k} = \frac{\partial a^i(x)}{\partial x^k} e_i(x) + a^j(x) \Gamma_{jk}^i(x) e_i(x) = \left(\frac{\partial a^i(x)}{\partial x^k} + a^j(x) \Gamma_{jk}^i(x) \right) e_i(x).$$

Зауважимо, що коли ми говоримо про вектор, скажімо $\vec{a}(x)$, то дуже часто посилаємось на нього, як сукупність його контраваріантних координат $a^i(x)$, оскільки вважаємо систему локальних координат a з нею і базиси дотичного простору в кожній точці многовиду M напередвизначеними. Таке диференціювання координат вектору ми називаємо *коваріантним диференціюванням* вектору, і позначаємо таким чином:

$$\nabla_k a^i(x) = \frac{\partial a^i(x)}{\partial x^k} + a^j(x) \Gamma_{jk}^i(x).$$

Наведений знак коваріантного диференціювання має назву «набла». Часто для компактного запису замість набла використовують знак коми або крапки з комою:

$$a^i_{,k}(x) = \frac{\partial a^i(x)}{\partial x^k} + a^j(x) \Gamma_{jk}^i(x).$$

Тривимірна матриця $\Gamma_{jk}^i(x)$ має назву *об'єкта афінної зв'язності*. Об'єкт афінної зв'язності не є тензором, оскільки при переході від системи координат $x^1, x^2 \dots x^n$ до нової $x^1, x^2 \dots x^n$ відбувається за формулами:

$$\Gamma_{j'k'}^{i'}(x') = \Gamma_{jk}^i(x) \frac{\partial x^j}{\partial x^{j'}} \frac{\partial x^k}{\partial x^{k'}} \frac{\partial x^{i'}}{\partial x^i} + \frac{\partial^2 x^i}{\partial x^{j'} \partial x^{k'}} \frac{\partial x^i}{\partial x^i}. \quad (1.3)$$

Такий закон перетворення об'єкту афінної зв'язності пов'язаний з тим, що коваріантна похідна вектору має перетворюватися при переході до іншої системи координат як тензор, тобто за формулою (1.2):

$$\nabla_{k'} a^{i'}(x') = \nabla_k a^i(x) \frac{\partial x^k}{\partial x^{k'}} \frac{\partial x^{i'}}{\partial x^i}.$$

Враховуючи формулу (1.2) можна також показати, що для того, щоб коваріантна похідна ковектору b_i також була тензором, потрібно щоб вона була обчислена за формулою:

$$\nabla_k b_i(x) = \frac{\partial b_i(x)}{\partial x^k} - b_j(x) \Gamma_{ik}^j(x).$$

Узагальнюючи, можна отримати формулу знаходження коваріантної похідної довільного тензора:

$$\begin{aligned} \nabla_k T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p} &= T_{j_1 j_2 \dots j_q, k}^{i_1 i_2 \dots i_p} = \frac{\partial T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p}}{\partial x^k} + T_{j_1 j_2 \dots j_q}^{\alpha i_2 \dots i_p} \Gamma_{\alpha k}^{i_1} + T_{j_1 j_2 \dots j_q}^{i_1 \alpha \dots i_p} \Gamma_{\alpha k}^{i_2} + \dots \\ &+ T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots \alpha} \Gamma_{\alpha k}^{i_p} - T_{\alpha j_2 \dots j_q}^{i_1 i_2 \dots i_p} \Gamma_{j_1 k}^{\alpha} - T_{j_1 \alpha \dots j_q}^{i_1 i_2 \dots i_p} \Gamma_{j_2 k}^{\alpha} - \dots - T_{j_1 j_2 \dots \alpha}^{i_1 i_2 \dots i_p} \Gamma_{j_q k}^{\alpha}. \end{aligned} \quad (1.4)$$

Тут і надалі в усіх випадках, коли не відбувається заміна координат – ми не будемо вказувати x як аргумент тензорних функцій і об'єкта афінної зв'язності, тобто писатимемо $T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p}$ замість $T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p}(x)$, Γ_{ik}^j замість $\Gamma_{ik}^j(x)$ тощо.

Ми називаємо тензор коваріантно сталим, якщо

$$\nabla_k T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p} = 0.$$

Поняття коваріантної похідної дозволяє розвинути так званий апарат абсолютного диференціювання.

Абсолютним диференціалом тензора ми називаємо вираз: $DT_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p} = \nabla_k T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p} dx^k$.

Ми кажемо що тензор $T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p}$, заданий в кожній точці деякого шляху *паралельно переноситься уздовж цього шляху*, якщо його абсолютний диференціал при будь-якому нескінченно малому зміщенні уздовж цього шляху дорівнює нулю.

Зокрема при паралельному перенесенні вектора ξ^i за напрямом $\{dx^k\}$ маємо:

$$D\xi^i = \nabla_k \xi^i dx^k = \frac{\partial \xi^i}{\partial x^k} dx^k + \Gamma_{jk}^i \xi^j dx^k = d\xi^i + \Gamma_{jk}^i \xi^j dx^k = 0.$$

Отже, отримуємо формулу паралельного перенесення вектора:

$$d\xi^i = -\Gamma_{jk}^i \xi^j dx^k.$$

Запис $\nabla_X Y$ означає коваріантну похідну вектору $Y = Y^i e_i$ за напрямом вектору $X = X^i e_i$. Якщо розписати докладно в координатах, отримуємо:

$$\nabla_X Y = (X^k \nabla_k Y^i) e_i = X^k \left(\frac{\partial Y^i}{\partial x^k} + \Gamma_{jk}^i Y^j \right) e_i.$$

Варто відзначити ще таку деталь, наприклад, запис

$$\nabla_X Y = 0$$

може бути витлумачений по-різному, залежно від контексту. Якщо X є довільним векторним полем, то це означатиме що $\nabla_k Y^i = 0$, тобто коваріантні похідні по усім змінним є нульовими, отже певне поле Y є коваріантно сталим. Якщо не тільки Y , але і X є також певним векторним полем, то відповідний запис в координатах означитиме $X^k \nabla_k Y^i = 0$, і певне поле Y є коваріантно сталим лише уздовж інтегральних кривих певного поля X .

Якщо виконується рівність

$$\nabla_k Y^i = \rho_k Y^i$$

(ρ_k – деякий ковектор), то векторне поле Y називають *рекурентним*. У безкоординатній формі це виглядатиме так:

$$\nabla_X Y = \rho(X)Y,$$

тут X є довільним векторним полем, тобто $\forall X (\nabla_X Y = \rho(X)Y)$.

1.5. Геодезичні лінії

Нехай на многовиді M задано криву $\gamma(t)$ як відображення $\gamma: \bullet \rightarrow M$, де $\bullet \subset \mathbb{R}$, тобто якийсь відрізок $a \leq t \leq b$. Оскільки ми обрали систему координат, то криву можна описати системою параметричних рівнянь.

$$x^i = x^i(t), (i=1, 2, \dots, n).$$

Тоді дотичний до такої кривої вектор виглядатиме так:

$$\dot{x}^i = \frac{dx^i}{dt}, (i=1, 2, \dots, n).$$

Тензорне поле $T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p}$ називають *рекурентним уздовж кривої $\gamma(t)$* , якщо

$$\nabla_{\dot{x}} T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p} = \dot{x}^k \nabla_k T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p} = f(t) T_{j_1 j_2 \dots j_q}^{i_1 i_2 \dots i_p},$$

де $f(t)$ – деякий скаляр. Зокрема, для вектору ξ^i рекурентність уздовж кривої $\gamma(t)$ виглядатиме так:

$$\nabla_{\dot{x}} \xi^i = \dot{x}^k \nabla_k \xi^i = \frac{\partial \xi^i}{\partial x^k} \dot{x}^k + \Gamma_{jk}^i \xi^j \dot{x}^k = f(t) \xi^i.$$

Оскільки $\frac{\partial \xi^i}{\partial x^k} \dot{x}^k = \frac{\partial \xi^i}{\partial x^k} \frac{dx^k}{dt} = \frac{d\xi^i}{dt}$ (ланцюгове правило диференціювання), то вимогу рекурентності уздовж кривої запишемо у вигляді:

$$\frac{d\xi^i}{dt} + \Gamma_{jk}^i \xi^j \dot{x}^k = f(t) \xi^i.$$

За допомогою властивості рекурентності ми можемо ввести нове, дуже важливе в геометрії, поняття.

Крива $\gamma(t)$, що задана рівняннями $x^i = x^i(t)$, $(i=1,2,\dots,n)$ має назву *геодезичної многовиди* M у зв'язності Γ_{jk}^i , якщо дотичний до цієї кривої вектор є рекурентним уздовж неї:

$$\frac{d^2 x^i}{dt^2} + \Gamma_{jk}^i \frac{dx^j}{dt} \frac{dx^k}{dt} = f(t) \frac{dx^i}{dt}. \quad (1.5)$$

Зауважимо, що можна обрати таку параметризацію $s = s(t)$ кривої γ , що рівняння геодезичної прийме вигляд:

$$\frac{d^2 x^i}{ds^2} + \Gamma_{jk}^i \frac{dx^j}{ds} \frac{dx^k}{ds} = 0, \quad (1.6)$$

Тобто дотичний вектор тепер переноситься паралельно. Такий параметр має назву *канонічного*.

Поняття геодезичної є дуже важливим. Справа в тому, що простори з якими працює ріманова геометрія не містять прямих ліній в тому сенсі до яких ми звикли у евклідовій геометрії. Геодезичні лінії є фактично є аналогами прямих у неевклідовому просторі, «найпрямішими» серед кривих.

1.6. Плоска зв'язність, тензор кривини

Взагалі кажучи, об'єкт афінної зв'язності Γ_{jk}^i є досить довільною структурою, головною вимогою є лише перетворення його при заміні координат згідно з формулою (1.3). Також, у загальному випадку не

є обов'язковим, щоб зв'язність була симетричною, тобто можливі випадки, коли

$$\Gamma_{jk}^i \neq \Gamma_{kj}^i.$$

В такому випадку кажуть, що зв'язність має ненульове *кручення*.

Об'єкт, який описує кручення має вигляд:

$$T_{jk}^i = \Gamma_{jk}^i - \Gamma_{kj}^i$$

і крім того є тензором. Це – так званий *тензор кручення*. Кол зв'язність є симетричною, то тензор кручення дорівнює нулю.

Дослідимо випадок, коли можна перейти до такої системи координат $y^1, y^2 \dots y^n$, в якій

$$\dot{\Gamma}_{jk}^i(y) \equiv 0. \quad (1.7)$$

Запишемо (1.3) у вигляді:

$$\dot{\Gamma}_{jk}^i(x) = \Gamma_{\alpha\beta}^\gamma(y) \frac{\partial y^\alpha}{\partial x^j} \frac{\partial y^\beta}{\partial x^k} \frac{\partial x^i}{\partial y^\gamma} + \frac{\partial^2 y^\gamma}{\partial x^j \partial x^k} \frac{\partial x^i}{\partial y^\gamma}.$$

«Німі» індекси, тобто ті, за якими відбувається згортка позначено грецькими літерами. Враховуючи (1.7) отримуємо звідси

$$\Gamma_{jk}^i(x) = \frac{\partial^2 y^\gamma}{\partial x^j \partial x^k} \frac{\partial x^i}{\partial y^\gamma}.$$

Остаточно це можна записати у вигляді системи диференціальних рівнянь в частинних похідних відносно n невідомих функцій:

$$y^i = y^i(x^1, x^2 \dots x^n),$$

$$\frac{\partial^2 y^i}{\partial x^j \partial x^k} = \Gamma_{jk}^m(x) \frac{\partial y^i}{\partial x^m} \quad (i=1, 2, \dots n). \quad (1.8)$$

Така система рівнянь, розв'язаних відносно старших похідних називають *системою типу Коші*. Якщо вона має розв'язок, то частинні похідні його мають згідно за теоремою Шварца задовільняти рівності (рівність мішаних похідних незалежно від порядку диференціювання):

$$\frac{\partial}{\partial x^l} \left(\frac{\partial^2 y^i}{\partial x^j \partial x^k} \right) - \frac{\partial}{\partial x^k} \left(\frac{\partial^2 y^i}{\partial x^j \partial x^l} \right) = 0, \quad (i=1, 2, \dots n). \quad (1.9)$$

Зазвичай, (1.9) називають умовами інтегровності (1.8)

Враховуючи (1.8), з (1.9) отримуємо:

$$\begin{aligned} \frac{\partial}{\partial x^l} \left(\Gamma_{jk}^m(x) \frac{\partial y^i}{\partial x^m} \right) - \frac{\partial}{\partial x^k} \left(\Gamma_{jl}^m(x) \frac{\partial y^i}{\partial x^m} \right) &= 0, \\ \frac{\partial y^i}{\partial x^m} \frac{\partial \Gamma_{jk}^m(x)}{\partial x^l} + \Gamma_{jk}^m(x) \frac{\partial^2 y^i}{\partial x^l \partial x^m} - \frac{\partial y^i}{\partial x^m} \frac{\partial \Gamma_{jl}^m(x)}{\partial x^k} + \Gamma_{jl}^m(x) \frac{\partial^2 y^i}{\partial x^k \partial x^m} &= 0, \\ \frac{\partial y^i}{\partial x^m} \frac{\partial \Gamma_{jk}^m(x)}{\partial x^l} + \Gamma_{jk}^m(x) \Gamma_{lm}^n(x) \frac{\partial y^i}{\partial x^n} - \frac{\partial y^i}{\partial x^m} \frac{\partial \Gamma_{jl}^m(x)}{\partial x^k} + \Gamma_{jl}^m(x) \Gamma_{km}^n(x) \frac{\partial y^i}{\partial x^n} &= 0. \end{aligned}$$

Перейменовуючи «німі» індекси, маємо:

$$\frac{\partial y^i}{\partial x^m} \frac{\partial \Gamma_{jk}^m(x)}{\partial x^l} + \Gamma_{jk}^n(x) \Gamma_{ln}^m(x) \frac{\partial y^i}{\partial x^m} - \frac{\partial y^i}{\partial x^m} \frac{\partial \Gamma_{jl}^m(x)}{\partial x^k} + \Gamma_{jl}^n(x) \Gamma_{kn}^m(x) \frac{\partial y^i}{\partial x^m} = 0.$$

Виносимо спільний множник за дужки:

$$\frac{\partial y^i}{\partial x^m} \left(\frac{\partial \Gamma_{jk}^m(x)}{\partial x^l} + \Gamma_{jk}^n(x) \Gamma_{ln}^m(x) - \frac{\partial \Gamma_{jl}^m(x)}{\partial x^k} - \Gamma_{jl}^n(x) \Gamma_{kn}^m(x) \right) = 0.$$

Остаточно (1.9) можна записати у вигляді:

$$\frac{\partial y^i}{\partial x^m} R_{jlk}^m(x) = 0, (i = 1, 2, \dots, n), \quad (1.10)$$

де

$$R_{jlk}^m(x) = \frac{\partial \Gamma_{jk}^m(x)}{\partial x^l} + \Gamma_{jk}^n(x) \Gamma_{ln}^m(x) - \frac{\partial \Gamma_{jl}^m(x)}{\partial x^k} - \Gamma_{jl}^n(x) \Gamma_{kn}^m(x). \quad (1.11)$$

Символом $R_{jlk}^m(x)$ ми позначаємо один раз контраваріантний, 3 рази коваріантний тензор, що має назву *тензора кривини*, а також *тензора Рімана*, або *символа Рімана другого роду*.

Оскільки матриця Якобі $\left(\frac{\partial y^i}{\partial x^m} \right)$ є невинродженою, тобто $\det \left(\frac{\partial y^i}{\partial x^m} \right) \neq 0$, то умова інтегровності (10) буде виконуватися, якщо тензор кривини буде тотожно дорівнювати нулю:

$$R_{jlk}^m(x) \equiv 0.$$

Зв'язність, коефіцієнти можна перетворенням координат зробити тотожно рівними нулю ($\Gamma_{jk}^i(y) \equiv 0$) має назву *плоскої*. Ми бачимо, що плоскою може бути лише така зв'язність, тензор кривини якої тотожно дорівнює нулю. Додамо також, що з рівняння (1.8) є очевидним, що плоскою зв'язністю може бути лише симетрична зв'язність, тобто для якої $\Gamma_{jk}^i = \Gamma_{kj}^i$.

1.7. Ріманова метрика і зв'язність Леві – Чивіта

Нехай заданий многовид M із симетричною зв'язністю $\Gamma_{jk}^i(x)$ та метричним тензором $g_{ij}(x)$.

Як ми вже бачили, зв'язність на многовиді визначає зокрема як мають змінюватися координати тензора (зокрема вектору) при його паралельному перенесенні, враховуючи як змінюється від однієї точки до іншої дотичний простір, а з ним і базис цього дотичного простору. Метричний тензор задає формулу для обчислення скалярного добутку у дотичному просторі. Дослідимо, як можна узгодити зв'язність та метрику.

Розглянемо задану на многовиді M деяку криву $\gamma(t)$ (не обов'язково геодезичну) що описується системою параметричних рівнянь.

$$x^i = x^i(t), (i=1, 2, \dots, n).$$

Нехай уздовж цієї кривої паралельно переносяться два вектори: $\vec{a}(x) = a^i(x)e_i(x)$ та $\vec{b}(x) = b^i(x)e_i(x)$.

Це означає що їх координати уздовж кривої $\gamma(t)$ задовільнятимуть рівнянням:

$$\begin{aligned} \nabla_{\dot{x}} a^i &= \dot{x}^k \nabla_k a^i = \frac{\partial a^i}{\partial x^k} \dot{x}^k + \Gamma_{jk}^i a^j \dot{x}^k = 0, \\ \nabla_{\dot{x}} b^i &= \dot{x}^k \nabla_k b^i = \frac{\partial b^i}{\partial x^k} \dot{x}^k + \Gamma_{jk}^i b^j \dot{x}^k = 0. \end{aligned} \quad (1.12)$$

Тут $\dot{x}^i = \frac{dx^i}{dt}$ – дотичний до кривої $\gamma(t)$ вектор.

Скалярний добуток цих векторів обчислюємо за формулою:

$$\vec{a} \cdot \vec{b} = a^i g_{ij} b^j.$$

Варто очікувати, що оскільки вектори переносяться паралельно уздовж кривої, то їх скалярний добуток має бути незмінний:

$$\frac{d(\vec{a}(x(t)) \cdot \vec{b}(x(t)))}{dt} = 0.$$

Тому маємо:

$$\frac{d(\vec{a}(x(t)) \cdot \vec{b}(x(t)))}{dt} = \nabla_x a^i g_{ij} b^j + a^i \nabla_x g_{ij} b^j + a^i g_{ij} \nabla_x b^j = 0.$$

Оскільки вектори переносяться паралельно, внаслідок (1.12) в останньому рівнянні перший та третій доданки дорівнюють нулю. Отже

$$a^i \nabla_x g_{ij} b^j = 0.$$

Оскільки вектори a^i та b^i – довільні¹, то

$$\nabla_x g_{ij} = \frac{dx^k}{dt} \nabla_k g_{ij} = 0.$$

Крива $\gamma(t)$ є також довільною, тому маємо:

$$\nabla_k g_{ij} \equiv 0. \quad (1.13)$$

Коли виконується умова (1.13), то кажуть що зв'язність є узгодженою з метрикою.

Згідно із загальною формулою (1.4) для коваріантного диференціювання довільного тензора, (1.13) можна записати таким чином:

$$\frac{\partial g_{ij}}{\partial x^k} - g_{aj} \Gamma_{ik}^a - g_{ia} \Gamma_{jk}^a = 0. \quad (1.14)$$

Опускаючи індекси у коефіцієнтах зв'язності за формулою:

$$\Gamma_{ik,j} = g_{aj} \Gamma_{ik}^a, \quad (1.15)$$

¹ Строго кажучи, насправді довільними є початкові значення координат векторів $a^i(x(t_0))$ та $b^i(x(t_0))$ для t_0 . Значення координат векторів на шляху паралельного перенесення $\gamma(t)$ однозначно визначаються з рівнянь (1.12). Однозначність випливає з припущення локальної аналітичності функцій з якими ми оперуємо, зокрема $x^i(t)$ та $\Gamma_{jk}^i(x)$.

З (1.14) маємо:

$$\frac{\partial g_{ij}}{\partial x^k} = \Gamma_{ik,j} + \Gamma_{jk,i}. \quad (1.16)$$

Перейменуємо індекси у (1.15):

$$\frac{\partial g_{ik}}{\partial x^j} = \Gamma_{ij,k} + \Gamma_{kj,i}, \quad (1.17)$$

$$\frac{\partial g_{kj}}{\partial x^i} = \Gamma_{ki,j} + \Gamma_{ji,k}. \quad (1.18)$$

Додаючи та віднімаючи за схемою (1.17)+(1.17)–(1.18), отримуємо:

$$\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{kj}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k} = \Gamma_{ij,k} + \Gamma_{kj,i} + \Gamma_{ki,j} + \Gamma_{ji,k} - \Gamma_{ik,j} - \Gamma_{jk,i}.$$

Враховуючи симетричність символу Крістофеля першого роду за першими двома нижніми індексами, маємо

$$\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{kj}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k} = \Gamma_{ij,k} + \Gamma_{ji,k} = 2\Gamma_{ij,k}.$$

Остаточно

$$\Gamma_{ij,k} = \frac{1}{2} \left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{kj}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k} \right). \quad (1.19)$$

Отримана таким чином величина (1.19) має назву *символа Крістофеля першого роду*.

Враховуючи (1.15) піднімаємо у (1.19) індекс k :

$$\Gamma_{ij}^k = \frac{g^{k\alpha}}{2} \left(\frac{\partial g_{i\alpha}}{\partial x^j} + \frac{\partial g_{aj}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^\alpha} \right). \quad (1.20)$$

Величину, отриману за допомогою формули (1.20) називають *символом Крістофеля другого роду*. Можна легко показати, що обидві величини, визначені в (1.19) і (1.20) не є тензорами.

Така симетрична зв'язність, що є узгодженою з метрикою називають *зв'язністю Леві – Чівіта*. Серед симетричних зв'язностей лише вона одна є узгодженою з метрикою, тобто є коваріантно сталою, тому має місце тотожність (1.14).

Але маючи зв'язність Леві – Чівіта ми можемо побудувати і інші, неузгоджені з даною нам метрикою зв'язності.

Формула проста:

$$\tilde{\Gamma}_{ij}^k = \Gamma_{ij}^k + T_{ij}^k. \quad (1.21)$$

Тут T_{ij}^k – довільний 1 раз контраваріантний і 2-коваріантний тензор, Γ_{ij}^k – символ Крістофеля другого роду. Якщо ми хочемо, щоб нова зв'язність $\tilde{\Gamma}_{ij}^k$ була симетричною, то треба вимагати, щоб T_{ij}^k був симетричним по коваріантним індексам. Зауважимо також, що $\tilde{\Gamma}_{ij}^k$ ми не можемо називати символом Крістофеля другого роду, оскільки ці коефіцієнти не є узгодженими формулою (1.20) з даною нам метрикою $g_{ij}(x)$. Поняття символу Крістофеля першого роду для $\tilde{\Gamma}_{ij}^k$ – взагалі позбавлене сенсу, оскільки нам для того, щоб опустити верхній індекс потрібна нова метрика $\tilde{g}_{ij}(x)$, з якою узгоджена нова зв'язність. Тим не менш, її можливо отримати, якщо розв'язати систему лінійних рівнянь (1.14), де нам відомі лише коефіцієнти нової зв'язності $\tilde{\Gamma}_{ij}^k$.

$$\frac{\partial \tilde{g}_{ij}}{\partial x^k} = \tilde{g}_{aj} \tilde{\Gamma}_{ik}^a + \tilde{g}_{ia} \tilde{\Gamma}_{jk}^a. \quad (1.22)$$

Лише отримавши розв'язок системи рівнянь (1.22), ми можемо називати $\tilde{\Gamma}_{ij}^k(x)$ символом Крістофеля другого роду, і за допомогою знайденої метрики $\tilde{g}_{ij}(x)$ отримати відповідний символ Крістофеля першого роду. Втім, може виявитися, що система (1.22) не має розв'язку. Тоді кажуть, що зв'язність $\tilde{\Gamma}_{ij}^k(x)$ не є метризовною.

Але повернемося до зв'язності Леві – Чівіта. Якщо вона є плоскою, то це означає, що існує система координат, в якій

$$\Gamma_{ij}^k(x) \equiv 0. \quad (1.23)$$

З рівняння (1.14) випливає, що

$$\frac{\partial g_{ij}}{\partial x^k} = 0. \quad (1.24)$$

Умови інтегрованості для (1.24) задовільняються тотожно. Отже, рівняння (1.24) є цілком інтегровними, що забезпечує локально в досліджуваній області многовиду задати евклідову метрику:

$$g_{ij} = \delta_{ij} = \begin{cases} 1, & \text{якщо } i = j, \\ 0, & \text{якщо } i \neq j. \end{cases}$$

1.7. Геодезичні у зв'язності Леві – Чивіта

Згадаємо відому із загального курсу задачу. Нехай U тривимірному евклідовому просторі R^3 задана крива параметричним рівнянням:

$$\begin{cases} x = x(t), \\ y = y(t), \\ z = z(t). \end{cases}$$

Потрібно знайти довжину відрізка кривої, що з'єднує точки $(x(t_1), y(t_1), z(t_1))$ та $(x(t_2), y(t_2), z(t_2))$.

Ми користувались формулою:

$$l = \int_{t_1}^{t_2} \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} dt.$$

Підкореневий вираз є скалярним квадратом дотичного до кривої вектора:

$$\vec{r}(\dot{x}, \dot{y}, \dot{z}),$$

$$\text{де } \dot{x} = \frac{dx(t)}{dt}, \quad \dot{y} = \frac{dy(t)}{dt}, \quad \dot{z} = \frac{dz(t)}{dt}:$$

$$\vec{r} \cdot \vec{r} = \dot{x}^2 + \dot{y}^2 + \dot{z}^2.$$

Скалярний квадрат обчислено для евклідової метрики $g_{ij} = \delta_{ij}$.

Якщо деяку криву $\gamma(t)$ задано системою параметричних рівнянь.

$$x^i = x^i(t), (i=1, 2, \dots, n).$$

на неевклідовому многовиді M з метрикою g_{ij} , то скалярний квадрат дотичного вектору $\vec{\dot{x}} \cdot \vec{\dot{x}} = g_{ij} \dot{x}^i \dot{x}^j$.

Внаслідок цього, для обчислення довжини відрізка кривої на неевклідовому многовиді M матимемо:

$$l = \int_{t_1}^{t_2} \sqrt{g_{ij} \dot{x}^i \dot{x}^j} dt. \quad (1.25)$$

Виникає питання: якими мають бути функції $x^i = x^i(t)$, щоб інтеграл (1.25) мав найменше значення з можливих. Тобто, потрібно знайти криву, що з'єднуючі задані точки матиме найменшу довжину.

Якщо ми проаналізуємо ситуацію для довільної метрики g_{ij} , то можна довести, що значення інтегралу (1.25) буде мати екстремальне значення саме у випадку, коли криві є геодезичними, а саме, відповідають рівнянню (1.6).

Автопаралельні та цілком геодезичні підмноговиди

Розглянемо многовид M_m розмірності m що вкладений як поверхня у многовид N_n . Нехай у M_m задана координатна система $x^1, x^2 \dots x^m$, а у N_n координатна система $y^1, y^2 \dots y^n$ [14]. Рівняння такої поверхні задаються так:

[illegible]

Тоді є i -ю координатою у просторі, дотичному до N_n вектора, що є дотичним до координатної лінії x^a на поверхні M_m . Зрозуміло, що вектори $\frac{\partial \mathbf{y}}{\partial x^1}, \frac{\partial \mathbf{y}}{\partial x^2}, \dots, \frac{\partial \mathbf{y}}{\partial x^m}$ мають утворювати базис дотичного простору² до поверхні M_m у довільній точці цієї поверхні $\mathbf{y}_0(x_0^1, x_0^2, \dots, x_0^m)$. Тому ці вектори мають бути лінійно незалежними. Звідси випливає, що ранг матриці Якобі

$$\left(\frac{\partial y^i}{\partial x^\alpha}\right) = \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \dots & \frac{\partial y^1}{\partial x^m} \\ \vdots & \ddots & \vdots \\ \frac{\partial y^n}{\partial x^1} & \dots & \frac{\partial y^n}{\partial x^m} \end{pmatrix}.$$

² Такий дотичний простір часто позначається як $T_{x_0}(M_m)$. Очевидно, що $T_{x_0}(M_m)$ є підпростором до $T_{y(x_0)}(N_n)$ – простору, дотичного до N_n в точці $y_0(x_0^1, x_0^2, \dots, x_0^m)$.

має дорівнювати m . Для подальшого викладення домовимосся позначати:

$$B_{\alpha}^i = \frac{\partial y^i}{\partial x^{\alpha}}. \quad (1.27)$$

Якщо на N_n задано метрику, g_{ij} то на M_m буде індуковано метрику $\dot{g}_{\alpha\beta}$:

$$\dot{g}_{\alpha\beta} = B_{\alpha}^i g_{ij} B_{\beta}^j, \quad (1.28)$$

І відповідно зв'язність Леві – Чівіта:

$$\dot{\Gamma}_{\alpha\beta}^{\gamma}(x) = B_i^{\gamma} \left(\Gamma_{jk}^i(y(x)) B_{\alpha}^j B_{\beta}^k + \frac{B_{\alpha}^i}{\partial x^{\beta}} \right). \quad (1.29)$$

Тут B_i^{γ} обчислюється за допомогою матриці $\dot{g}^{\alpha\gamma}$, що є оберненою до матриці індукованої метрики (1.28):

$$B_i^{\gamma} = B_{\alpha}^j g_{ij} \dot{g}^{\alpha\gamma}.$$

Об'єкт (1.27) є так званим «мішаним» тензором. Це пов'язано з тим, що верхній індекс у B_{α}^i відноситься до структури, пов'язної з N_n , нижній відповідно до структури M_m . Тому при коваріантному диференціюванні треба враховувати як зв'язність охоплюючого простору N_n так і індуковану зв'язність (1.29) вкладеного простору M_m .

$$\dot{\nabla}_{\gamma} B_{\alpha}^i = \frac{\partial B_{\alpha}^i}{\partial x^{\gamma}} + B_{\alpha}^j B_{\gamma}^k \Gamma_{jk}^i - B_{\beta}^i \dot{\Gamma}_{\alpha\gamma}^{\beta}. \quad (1.30)$$

Коваріантна похідна $\dot{\nabla}$ мішаного тензора має назву *коваріантної похідної Ван дер Вардена-Бортолотті*.

Коваріантно диференціюючи (1.28) використовуючи (1.30) можна впевнитися, що

$$\dot{\nabla}_{\gamma} B_{\alpha}^i g_{ij} B_{\beta}^j = 0. \quad (1.31)$$

З (1.31) випливає, що є ортогональним до всіх векторів, що утворюють дотичний простір $T_{x_0}(M_m)$.

Тому ми можемо в кожній точці M_m побудувати $n-m$ взаємно перпендикулярних одиничних векторів:

$$C_{|m+1}^i, C_{|m+2}^i, \dots, C_{|n}^i$$

так щоб разом з векторами $B_1^i, B_2^i, \dots, B_m^i$ утворювали базис всього охоплюючого дотичного простору $T_{y(x_0)}(N_n)$.

Враховуючи (1.31), можна записати розкладання по векторах утвореного базису:

$$\overset{\cdot}{\nabla}_{\gamma} B_{\alpha}^i = \sum_{\sigma=m+1}^n H_{\gamma|\sigma} C_{|\sigma}^i. \quad (1.32)$$

Коефіцієнти розкладання $H_{\gamma|\sigma}$ вектора $\overset{\cdot}{\nabla}_{\gamma} B_{\alpha}^i$ по векторах $C_{|\sigma}^i$, $\sigma = m+1, m+2, \dots, n$ є *другими фундаментальними тензорами* по відношенню до векторів $C_{|\sigma}^i$ [14]. Першим *фундаментальним тензором* вважається індукована на многовиді метрикою $\overset{\cdot}{g}_{\alpha\beta}$.

Якщо ми коваріантно продиференціюємо нормальний вектор, то виявляється, що його похідна має вигляд:

$$\overset{\cdot}{\nabla}_{\gamma} C_{|\sigma}^i = - \sum_{\alpha=1}^m H_{\gamma|\sigma}^{\alpha} B_{\alpha}^i + \sum_{\tau=m+1}^n L_{\gamma|\sigma\tau} C_{|\tau}^i. \quad (1.33)$$

Формула (1.32) має назву *формули Гауса* а (1.33) *формули Вейнгартена*. Також вони мають спільну назву – *дериваційні рівняння M_m* вкладеного у N_n .

Щоб виписати умови інтегрованості дериваційних рівнянь, нам знадобиться ще одне поняття. В (1.11) ми ввели тензор Рімана другого роду. *Тензор Рімана першого роду* можна отримати, опускаючи у тензора Рімана другого роду контраваріантний індекс за допомогою метрики:

$$R_{ijkl} \stackrel{\text{def}}{=} g_{ih} R_{jlk}^h. \quad (1.34)$$

Зауважимо, що тензор Рімана другого роду можна отримати для будь-якої афінної зв'язності, наявність метрики не є обов'язковою. Тензор Рімана першого роду (1.34) можна отримати для метричного простору, причому зв'язність, для якої розраховано R_{jlk}^h має бути узгоджена з метрикою g_{ij} .

Якщо дослідити умови інтегрованості (1.32) та (1.33), то ми отримаємо такі рівняння:

$$\overset{\cdot}{R}_{\alpha\beta\gamma\delta} = \sum_{\sigma=m+1}^n \left(H_{\alpha\gamma|\sigma} H_{\beta\delta|\sigma} - H_{\alpha\gamma|\sigma} H_{\beta\gamma|\sigma} \right) + R_{ijkl} B_{\alpha}^i B_{\beta}^j B_{\gamma}^k B_{\delta}^l, \quad (1.35)$$

$$H_{\alpha\beta,\gamma|\sigma} - H_{\alpha\gamma,\beta|\sigma} = \sum_{\tau=m+1}^n \left(L_{\gamma|\tau\sigma} H_{\alpha\beta|\tau} - L_{\beta|\tau\sigma} H_{\alpha\gamma|\tau} \right) + R_{ijkl} B_{\alpha}^i B_{\beta}^j B_{\gamma}^k C_{|\sigma}^l, \quad (1.36)$$

$$\sigma = m+1, m+2, \dots, n$$

Рівняння (1.35) називають *формулою Гауса*, а (1.36) – *формулою Кодацці*.

Важливим частинним випадком підмноговидів є поняття автопаралельних та цілком геодезичних многовидів.

Многовид M_m вкладений у N_n називають *геодезичним у точці* $p \in M_m$, якщо будь-яка геодезична многовиду M_m , що проходить через точку p є також геодезичною для простору N_n . Якщо M_m є геодезичним у будь-якій своїй точці, то тоді його називають *цілком геодезичним підмноговидом* простору N_n [3; 12]. Підмноговид $M_m \subset N_n$ називають *автопаралельним*, якщо для кожного вектора, дотичного $T_{x_0}(M_m)$ і кожної кривої γ , що виходить з x_0 , паралельний перенос цього вектору вздовж γ (відносно афінної зв'язності охоплюючого простору N_n) призводить до того, що перенесений вектор лишається дотичним до M_m . Наприклад, сфера – не є автопаралельним у нашому тривимірному просторі. Якщо ми візьмемо будь-яку точку на меридіані глобуса, уявімо собі дотичний до меридіана вектор і почнемо його паралельно переносити уздовж меридіану – то побачимо, що він у процесі перенесення перестає бути дотичним як до меридіана, так і до кола в цілому (рис. 3).

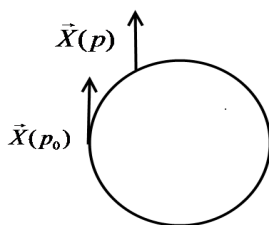


Рисунок 3 – Паралельне перенесення вектора відносно афінної зв'язності охоплюючого простору $\mathbb{R}^3$ уздовж меридіана сфери – підмноговида тривимірного простору

До речі, саме меридіани є геодезичними лініями сфери. А от серед паралелей – геодезичною є лише екватор.

Існує таке твердження: кожний автопаралельний підмноговид M_m , вкладений у N_n буде в N_n цілком геодезичним, якщо афінна зв'язність в N_n є симетричною. І навпаки, симетричність афінної зв'язності N_n

гарантує, що будь-який цілком геодезичний його підмноговид M_m буде автопаралельним.

Оскільки ми будемо в наших задачах стикатися лише із симетричною зв'язністю, то для нас автопаралельний та цілком геодезичні многовиди будуть синонімічними поняттями.

Дуже важливою ознакою того що многовид M_m , вкладений у N_n буде цілком геодезичним є той факт, що коефіцієнти другої квадратичної форми тотожно дорівнюють нулю:

$$H_{\gamma\alpha|\sigma} \equiv 0, \gamma, \alpha = 1, 2, \dots, m; \sigma = m+1, m+2, \dots, n. \quad (1.37)$$

І навпаки, якщо підмноговид є цілком геодезичним (автопаралельним), то має виконуватися (1.35).

У такому випадку дериваційні формули приймають вигляд:

$$\begin{aligned} \dot{\nabla}_{\gamma} B_{\alpha}^i &= 0; \\ \dot{\nabla}_{\gamma} C_{|\sigma}^i &= \sum_{\tau=m+1}^n L_{\gamma|\sigma\tau} C_{|\tau}^i. \end{aligned}$$

Стосовно формул Гауса та Кодацці, то вони відповідно виглядатимуть для цілком геодезичних підмноговидів так:

$$\begin{aligned} \dot{R}_{\alpha\beta\gamma\delta} &= R_{ijkl} B_{\alpha}^i B_{\beta}^j B_{\gamma}^k B_{\delta}^l; \\ R_{ijkl} B_{\alpha}^i B_{\beta}^j B_{\gamma}^k C_{|\sigma}^l &= 0. \end{aligned}$$

З викладеного видно, що перший основний тензор $\dot{g}_{\alpha\beta}$ цілком визначає внутрішню геометрію вкладеного многовиду, другий основний тензор $H_{\gamma\alpha|\sigma}$ – показує як саме підмноговид M_m є розташованим у охоплюючому просторі N_n .

Наостанок зауважимо ще один дуже важливий момент. Ми дуже часто казали і будемо казати, що підмноговид M_m є **вкладений** у N_n . Термінологічно – це не зовсім точно, оскільки весь наведений вище математичний апарат працює лише локально а не в цілому.

Термін **вкладання** – гарантує що мнотгеновид M_m взаємно-однозначно відобразиться на деяку поверхню у N_n , без розривів, самоперетинів, склеювання. Тобто, як кажуть, із повним збереженням топології. При локальному підході – ми можемо гарантувати взаємно-однозначність лише в межах деякої локальної карти

многовида. Для такого випадку існує термін: *занурення*, або *імерсія*. Але оскільки дослідження топології статистичних многовидів лежить поза межами даної роботи, то поняття «вкладання» і «занурення» ми будемо використовувати як синонімічні.

2. Основи інформаційної геометрії

2.1. Метод максимальної правдоподібності

(*Maximum likelihood estimation, MLE*)

Нехай у нас є результати n спостережень деякої випадкової величини $x_1, x_2, \dots, x_n$, нам відома загальна формула її розподілу $p(x, \xi)$, але параметри $\xi = (\xi^1, \xi^2, \dots, \xi^m)$ цього розподілу невідомі.

Зауваження 2.1. Уточнимо, що без втрати загальності ми можемо вважати випадкову величину *одновимірною*. Для k -вимірної величини – кожному результату випробування відповідає k -вимірний вектор, а інтегральна функція розподілу виражається k -кратним інтегралом.

Підставляємо результати у функцію розподілу і отримані n виразів перемножуємо:

$$L(x_1, x_2, \dots, x_n; \xi) = p(x_1, \xi) p(x_2, \xi) \dots p(x_n, \xi).$$

Отримана функція має назву *функції правдоподібності* [3].

Зручніше користуватися *логарифмічною функцією правдоподібності*:

$$l(x_1, x_2, \dots, x_n; \xi) = \ln L(x_1, x_2, \dots, x_n; \xi) = \sum_{i=1}^n \ln p(x_i, \xi).$$

Вона є функцією від m змінних $\xi^1, \xi^2, \dots, \xi^m$.

Сутність методу максимальної правдоподібності полягає в тому, що значення невідомих параметрів мають бути такими, що при даних емпіричних результатах функція правдоподібності набувала б найбільшого значення, тобто ми маємо задачу знаходження безумовного екстремуму функції.

Для цього знайдемо частині похідні:

$$\frac{\partial l}{\partial \xi^j} = \sum_{i=1}^n \frac{1}{p(x_i, \xi)} \frac{\partial p(x_i, \xi)}{\partial \xi^j}.$$

Отриманий вектор градієнта має назву *функція внеску* (*score function*).

Для визначення оцінок параметрів прирівнюємо функцію внеску до нуля:

$$\frac{\partial l}{\partial \xi^j} = \sum_{i=1}^n \frac{1}{p(x_i, \xi)} \frac{\partial p(x_i, \xi)}{\partial \xi^j} = 0 \quad (i = 1, 2, \dots, m). \quad (2.1)$$

Ми отримали систему m рівнянь з m невідомими. Взагалі кажучи, система є нелінійною, і тому її розв'язування може бути пов'язаним з обчислювальними труднощами, знайдена критична точка $\hat{\xi} = (\hat{\xi}^1, \hat{\xi}^2, \dots, \hat{\xi}^m)$ може бути точкою лише локального екстремуму, тощо. В цьому полягає один з недоліків методу.

Але в будь-якому випадку, якщо вдалося знайти критичну точку, її потрібно перевірити на те, що вона є точкою хоча б локального максимуму.

Для цього складаємо матрицю Гессе:

$$H(\hat{\xi}) = \begin{pmatrix} \frac{\partial^2 l}{\partial \xi^{12}} & \dots & \frac{\partial^2 l}{\partial \xi^1 \partial \xi^k} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \xi^k \partial \xi^1} & \dots & \frac{\partial^2 l}{\partial \xi^{k2}} \end{pmatrix}. \quad (2.2)$$

Якщо матриця (2.2) відповідатиме негативно визначеній квадратичній формі,

$$\sum_{i,j} \frac{\partial^2 l}{\partial \xi^i \partial \xi^j} d\xi^i d\xi^j < 0,$$

це гарантує, що в критичній точці $\hat{\xi} = (\hat{\xi}^1, \hat{\xi}^2, \dots, \hat{\xi}^m)$ функція правдоподібності має локальний максимум. Таким чином, ми знайшли оцінки параметрів розподілу. Такі оцінки є *слухиними* (*конзистентними*), але на жаль не завжди є незміщеними. Якщо для параметра набору параметрів $\hat{\xi}$ існує ефективна оцінка, то система (2.1) матиме один і тільки один розв'язок.

Якщо виконується остання нерівність, то зрозуміло, що виконується протилежна:

$$-\sum_{i,j} \frac{\partial^2 l}{\partial \xi^i \partial \xi^j} d\xi^i d\xi^j > 0.$$

А це означає що матриця $\left(-\frac{\partial^2 l}{\partial \xi^i \partial \xi^j}\right)$ задає позитивно визначену квадратичну форму, і тому може відігравати роль метрики на деякому многовиді. Розглянемо такий многовид. Що стосується метода максимальної правдоподібності, то ми до нього повернемось у розділі 4.3.

2.2. Статистичні многовиди

Розглянемо регулярну статистичну модель

$$M = \{p(x, \xi)\}.$$

Вона представляє собою множину функцій що є щільностями розподілу деякої випадкової величини x , параметризовану m -вимірним вектором параметрів $\xi = (\xi^1, \xi^2, \dots, \xi^m)$. Таку модель можна розглядати як m -вимірний многовид, причому $\xi = (\xi^1, \xi^2, \dots, \xi^m)$ є локальною системою координат, що задана на цьому многовиді.

В даному випадку *функцією внеску (score function)* ми також називаємо m -вимірний вектор $s(x, \xi)$ але цього разу компоненти його визначаються як частинні похідні натурального логарифму щільності розподілу $l_\xi = \ln p(x, \xi)$:

$$s_i(x, \xi) = \partial_i l_\xi = \frac{\partial_i p(x, \xi)}{p(x, \xi)}, \quad i = 1, \dots, m. \quad (2.3)$$

Тут ми позначили $\partial_i = \frac{\partial}{\partial \xi^i}$.

Очевидно, що виконується така рівність, що свідчить про нульове математичне сподівання функції внеску (аналог формули (2.1)):

$$E_\xi [s_i(x, \xi)] = 0, \quad i = 1, \dots, m. \quad (2.4)$$

Тут величина E_ξ – математичне сподівання $p(x, \xi)$.

Фактично маємо [15; 3]:

$$\begin{aligned} E_{\xi} [s_i(x, \xi)] &= \int \partial_i l_{\xi}(x, \xi) p(x, \xi) dx = \int \partial_i (\ln(p(x, \xi))) p(x, \xi) dx = \\ &= \int \frac{\partial_i p(x, \xi)}{p(x, \xi)} p(x, \xi) dx = \int \partial_i p(x, \xi) dx = \partial_i \int p(x, \xi) dx = \partial_i 1 = 0. \end{aligned} \quad (2.4')$$

Зауваження 2.2. У формулах (2.4') мається на увазі визначений інтеграл, хоча границь інтегрування не вказано. Інтегрування відбувається по всій області що є множиною усіх можливих значень випадкової величин, або як кажуть у математичному аналізі, по всьому «носію» функції $p(x, \xi)$ стосовно аргументу x . Носієм функції $f(x)$ ми називаємо множину значень аргументу, для яких $f(x) \neq 0$. Також, якщо випадкова величина є k -вимірною, то у (2.4) матиметься на увазі k -кратний інтеграл по k -вимірній області. Але оскільки обчислення у (2.4') мають суто загальний характер – можна задовільнитися спрощеною символікою.

Компонент $s_i(x, \xi) = \partial_i l_{\xi}$ можна розглядати як дотичний вектор до многовиду M в напрямі координатної кривої ξ^i (Рис. 4). Звідси базисний вектор:

$$e_i = \frac{\partial}{\partial \xi^i}$$

є представленим внеском $s_i(x, \xi)$. Очевидно, він є випадковою змінною. Тому ми маємо випадкове представлення базисних векторів $e_i = \partial_i l_{\xi}$.

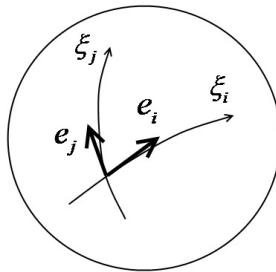


Рисунок 4 – Вектори базису дотичного простору до статистичного многовиду

Далі ми можемо визначити скалярний добуток двох базисних векторів:

$$\begin{aligned} e_i, e_j &= E_{\xi} [\partial_i l_{\xi} \partial_j l_{\xi}] = \int \frac{\partial \ln(p(x, \xi))}{\partial \xi^i} \frac{\partial \ln(p(x, \xi))}{\partial \xi^j} p(x, \xi) dx = \\ &= \int \frac{\partial p(x, \xi)}{\partial \xi^i} \frac{\partial p(x, \xi)}{\partial \xi^j} \frac{1}{p(x, \xi)} dx. \end{aligned} \quad (2.5)$$

Таким чином ми отримуємо матрицю $G(\xi) = g_{ij}$ яка також має назву *інформаційна матриця Фішера*:

$$g_{ij} \stackrel{\text{def}}{=} e_i, e_j = E [\partial_i l_{\xi} \partial_j l_{\xi}]. \quad (2.5')$$

Розглянемо $E [\partial_j \partial_i l_{\xi}]$:

$$\begin{aligned} E [\partial_j \partial_i l_{\xi}] &= \int \frac{\partial}{\partial \xi^j} \left(\frac{\partial \ln(p(x, \xi))}{\partial \xi^i} \right) p(x, \xi) dx = \\ &= \int \frac{\partial}{\partial \xi^j} \left(\frac{\partial_i p(x, \xi)}{p(x, \xi)} \right) p(x, \xi) dx = \\ &= \int \frac{\partial_j \partial_i p(x, \xi) p(x, \xi) - \partial_i p(x, \xi) \partial_j p(x, \xi)}{(p(x, \xi))^2} p(x, \xi) dx = \\ &= \int \frac{\partial_j \partial_i p(x, \xi) p(x, \xi) - \partial_i p(x, \xi) \partial_j p(x, \xi)}{p(x, \xi)} dx = \\ &= \int \frac{\partial_j \partial_i p(x, \xi) p(x, \xi)}{p(x, \xi)} dx - \int \frac{\partial_i p(x, \xi) \partial_j p(x, \xi)}{p(x, \xi)} dx = \\ &= \int \partial_j \partial_i p(x, \xi) dx - E [\partial_i l_{\xi} \partial_j l_{\xi}]. \end{aligned}$$

Внаслідок (2.4'):

$$\int \partial_j \partial_i p(x, \xi) dx = \partial_j \int \partial_i p(x, \xi) dx = 0.$$

Отже остаточно:

$$E [\partial_j \partial_i l_{\xi}] = -E [\partial_i l_{\xi} \partial_j l_{\xi}].$$

Отримаємо:

$$g_{ij} = -E[\partial_j \partial_i l_\xi]. \quad (2.6)$$

Також можна довести, що для g_{ij} існує і таке представлення:

$$g_{ij} = 4 \int \partial_i \sqrt{p(x, \xi)} \partial_j \sqrt{p(x, \xi)} dx. \quad (2.7)$$

Очевидно що матриця $G(\xi) = g_{ij}$ симетричною і для будь-якого m -вимірного вектору $\mathbf{a} = (a^1, a^2, \dots, a^m)^T$ (символ T означає транспонування) маємо:

$$\mathbf{a} G(\xi) \mathbf{a}^T = g_{ij} a^i a^j = \int \left(\sum_{i=1}^m a^i \partial_i l_\xi(x, \xi) \right)^2 p(x, \xi) dx \geq 0,$$

тобто квадратична форма, що задається матрицею $G(\xi)$ є позитивно напіввизначеною. Але оскільки ми вважаємо вектори $\{\partial_1 l_\xi, \partial_2 l_\xi, \dots, \partial_m l_\xi\}$ лінійно-незалежними, то побудована нами за допомогою них матриця задаватиме позитивно визначену квадратичну форму. Отже, g_{ij} задає метрику на многовиді M .

Очевидно, оскільки на многовиді задана метрика, то і існуватиме і узгоджена з нею зв'язність. Її компоненти можуть бути визначені за формулою (1.20).

$${}^0 \Gamma_{ij}^k = \frac{g^{k\alpha}}{2} \left(\frac{\partial g_{i\alpha}}{\partial \xi^j} + \frac{\partial g_{aj}}{\partial \xi^i} - \frac{\partial g_{ij}}{\partial \xi^\alpha} \right), \quad (2.8)$$

де $g^{k\alpha}$ означає матрицю, обернену до метричного тензору $g_{i\alpha}$. Нулем, що стоїть зліва від гамми, позначається об'єкт зв'язності Леві – Чівіті.

Статистичним многовидом ми називаємо трійку (M, g, T) , де M є зв'язним та гладкий многовид, g – ріманова метрика і нарешті T є гладке коваріантне цілком симетричне тензорне поле валентності три, що має назву *асиметрії (skewness)* [3; 15].

Асиметрія визначається за формулою:

$$T_{ijk}^{\text{def}} = E[\partial_i l_\xi \partial_j l_\xi \partial_k l_\xi], \quad (2.9)$$

або більш докладно:

$$T_{ijk} = \int \frac{\partial \ln(p(x, \xi))}{\partial \xi^i} \frac{\partial \ln(p(x, \xi))}{\partial \xi^j} \frac{\partial \ln(p(x, \xi))}{\partial \xi^k} p(x, \xi) dx. \quad (2.9')$$

Використовуючи тензор (2.9) ми можемо задати так звану α -зв'язність $\nabla^\pm$ і її коефіцієнти визначаються таким чином:

$${}^\alpha\Gamma_{ij,k} \stackrel{\text{def}}{=} {}^0\Gamma_{ij,k} - \frac{\alpha}{2} T_{ijk}. \quad (2.10)$$

Загалом, α є деяким дійсним числом. Але для статистичних досліджень вважають $\alpha = \pm 1$. Це буде обговорено нижче.

Приймаючи до уваги (2.5), (2.6), (2.8)–(2.10) в кожній точці $\xi = (\xi^1, \xi^2, \dots, \xi^m)$ ми отримуємо:

$$\begin{aligned} {}^\alpha\Gamma_{ij,k} &= \int \left(\partial_j \left(\frac{\partial_i p(x, \xi)}{p(x, \xi)} \right) + \frac{1-\alpha}{2} \frac{\partial_i p(x, \xi)}{p(x, \xi)} \frac{\partial_j p(x, \xi)}{p(x, \xi)} \right) \times \\ &\quad \times \frac{\partial_k p(x, \xi)}{p(x, \xi)} p(x, \xi) dx = \\ &= \int \left(\frac{\partial_j \partial_i p(x, \xi) \partial_k p(x, \xi)}{(p(x, \xi))^2} - \frac{\partial_i p(x, \xi)}{p(x, \xi)} \frac{\partial_j p(x, \xi)}{p(x, \xi)} \frac{\partial_k p(x, \xi)}{p(x, \xi)} + \right. \\ &\quad \left. + \frac{1-\alpha}{2} \frac{\partial_i p(x, \xi)}{p(x, \xi)} \frac{\partial_j p(x, \xi)}{p(x, \xi)} \frac{\partial_k p(x, \xi)}{p(x, \xi)} \right) p(x, \xi) dx. \end{aligned}$$

Остаточно маємо:

$$\begin{aligned} {}^\alpha\Gamma_{ij,k} &= \\ &= \int \left(\frac{\partial_j \partial_i p(x, \xi) \partial_k p(x, \xi)}{p(x, \xi)} - \frac{1+\alpha}{2} \frac{\partial_i p(x, \xi) \partial_j p(x, \xi) \partial_k p(x, \xi)}{(p(x, \xi))^2} \right) dx. \end{aligned} \quad (2.11)$$

або

$${}^\alpha\Gamma_{ij,k} = E \left[\left(\partial_j \partial_i l_\xi(x, \xi) + \frac{1-\alpha}{2} \partial_i l_\xi(x, \xi) \partial_j l_\xi(x, \xi) \right) \partial_k l_{\xi(x, \xi)} \right]. \quad (2.12)$$

Також, диференціюючи (2.5) по ξ^k маємо:

$$\begin{aligned} \frac{\partial g_{ij}}{\partial \xi^k} &= E \left[(\partial_k \partial_i l_\xi) (\partial_j l_\xi) \right] + E \left[(\partial_i l_\xi) (\partial_k \partial_j l_\xi) \right] + \\ &+ E \left[(\partial_i l_\xi) (\partial_j l_\xi) (\partial_k l_\xi) \right] = {}^0\Gamma_{ki,j} + {}^0\Gamma_{kj,i}. \end{aligned} \quad (2.13)$$

Розглядаючи (2.12) та (2.13), порівнюючи з (2.8), ми приходимо до висновку, що коли $\alpha = 0$, ми отримуємо 0-зв'язність ∇^0 , яка є рімановою зв'язністю, узгодженою з метрикою Фішера.

В закінчення цього розділу варто зазначити, що матриця, обернена до інформаційної матриці Фішера використовується для оцінки середньої квадратичної помилки незміщеної оцінки а α -зв'язність $\nabla^\pm$ застосовується для обчислення дивергенції між різними статистичними розподілами.

2.3. Приклади статистичних многовидів

Розглянемо деякі приклади [3]. Позначимо $\mathcal{X}$ множину значень, яку може приймати випадкова величина.

Відповідно, $\mathcal{P}(\mathcal{X})$ – множину можливих розподілів $p(x)$ випадкової величини, літерою Ξ позначимо простір параметрів розподілу.

Формально це виглядатиме так:

$$\mathcal{P}(\mathcal{X}) = \left\{ p: \mathcal{X} \rightarrow \mathbb{R} \mid \forall x \in \mathcal{X} p(x) \geq 0, \int p(x) dx = 1 \right\}.$$

1. Нормальний розподіл. $\mathcal{X} = \mathbb{R}$, розмірність множини параметрів $\dim \Xi = 2$,

$$\Xi = \{[\mu, \sigma] \mid -\infty < \mu < \infty, 0 < \sigma < \infty\};$$

$$p(x, \xi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

2. Багатовимірний нормальний розподіл. $\mathcal{X} = \mathbb{R}^k$, розмірність множини параметрів $\dim \Xi = k + \frac{k(k+1)}{2}$,

$$\Xi = \{[\mu, \Sigma] \mid \mu \in \mathbb{R}^k, \Sigma \in \mathbb{R}^{k \times k}\},$$

$$p(x, \xi) = \frac{1}{(\sqrt{2\pi})^k \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right).$$

Зазначимо, що матриця Σ утворює позитивно визначену квадратичну форму.

3. Розподіл Пуассона. $\mathcal{X} = \{0, 1, 2, \dots\}$, розмірність множини параметрів $\dim \Xi = 1$,

$$\Xi = \{\xi | \xi \geq 0\},$$

$$p(x, \xi) = e^{-\xi} \frac{\xi^x}{x!}.$$

4. Розподіл дискретної випадкової величини із скінченною кількістю можливих значень. $\mathcal{X} = \{x_0, x_1, x_2, \dots, x_n\}$, розмірність множини параметрів $\dim \Xi = n$

$$\Xi = \left\{ [\xi^0, \xi^1, \dots, \xi^n] \mid \forall i \xi^i \geq 0, \sum_{i=1}^n \xi^i < 1 \right\},$$

$$p(x, \xi) = \begin{cases} \xi^i & (1 \leq i \leq n) \\ 1 - \sum_{i=1}^n \xi^i & (i = 0). \end{cases}$$

3. Геометричні властивості статистичних многовидів

3.1. Статистичні многовиди і деякі дифеоморфізми

Нехай афінна зв'язність ∇ допускає взаємно однозначне диференційоване відображення f на іншу афінну зв'язність $\bar{\nabla}$ таке, що обернене відображення є також диференційованим. Локально обидві зв'язності є симетричними та віднесені до спільної координатної системи $\xi = (\xi^1, \xi^2, \dots, \xi^m)$.

Тензор деформації відносно відображення f виглядає так:

$$P_{ij}^h(\xi) = \bar{\Gamma}_{ij}^h(\xi) - \Gamma_{ij}^h(\xi), \quad (3.1)$$

де $\Gamma_{ij}^h(\xi)$ та $\bar{\Gamma}_{ij}^h(\xi)$ – компоненти афінних зв'язностей відповідно ∇ та $\bar{\nabla}$ відповідно.

Дифеоморфізм $f: \nabla \rightarrow \bar{\nabla}$ називається *геодезичним* відображенням, якщо будь-яка геодезична лінія відносно зв'язності ∇ відображається під дією f на геодезичну відносно $\bar{\nabla}$.

Відомо, що для того, щоб відображення f простору з афінною зв'язністю ∇ на інший простір з афінною зв'язністю $\bar{\nabla}$ було

геодезичним, необхідно та достатньо, щоб у спільній системі координат тензор деформації можна було представити у вигляді:

$$P_{ij}^h(\xi) = \psi_i(\xi)\delta_j^h + \psi_j(\xi)\delta_i^h, \quad (3.2)$$

де δ_i^h є вже знайомим символом Кронекера ψ_i – деякий ковектор. Геодезичне відображення називають *нетривіальним*, якщо $\psi \neq 0$.

Виникає питання, чи є можливим, щоб усі геодезичні відносно ріманової зв'язності ∇^0 співпадали з геодезичними α -зв'язності ∇^α якщо $\alpha \neq 0$. Якщо ми маємо таку відповідність, то

$$P_{ijk} = -\frac{\alpha}{2} T_{ijk}. \quad (3.3)$$

Прийнявши до уваги (3.2) і опускаючи верхній індекс за допомогою g_{hk} , маємо:

$$-\frac{\alpha}{2} T_{ijk} = \psi_i g_{jk} + \psi_j g_{ik}. \quad (3.4)$$

Міняючи у (3.4) місцями два останні індекси, ми отримуємо:

$$-\frac{\alpha}{2} T_{ikj} = \psi_i g_{kj} + \psi_k g_{ij}.$$

Оскільки тензор $r T_{ijk}$ є цілком симетричним, можна записати:

$$\psi_i g_{jk} + \psi_j g_{ik} = \psi_i g_{kj} + \psi_k g_{ij},$$

або

$$\psi_j g_{ik} = \psi_k g_{ij}. \quad (3.5)$$

Згортаючи (3.5) з $\psi^j = \psi_h g^{hj}$, ми отримуємо:

$$\|\psi\|^2 g_{ik} = \psi_k \psi_i. \quad (3.6)$$

Тут $\|\psi\|^2 = \psi^i \psi^j g_{ij} = \psi^j \psi_j$. Але відомо, що $\text{rang}(g_{ij}) \geq 2$. (3.5) містить протиріччя. Звідси ми отримуємо теорему.

Теорема 3.1. *Статистичний многовид (M, g, T) не дозволяє нетривіальних геодезичних відображень між рімановою зв'язністю ∇^0 та будь-якою α -зв'язністю ∇^α якщо $\alpha \neq 0$.*

Тепер розглянемо випадок зв'язності Вейля, тобто існування такої зв'язності Вейля по відношенню до метричного тензору g_{ij} ,

яка співпадатиме з α -зв'язністю $\nabla^\pm$ якщо $\alpha \neq 0$. В цьому випадку (3.1) та (3.3) також виконуються, але замість (3.2) ми маємо:

$$P_{ij}^h(\xi) = \varphi_i(\xi)\delta_j^h + \varphi_j(\xi)\delta_i^h - \varphi^h g_{ij},$$

де φ_i – деякий ковектор. Відповідно, замість (3.4) запишемо:

$$-\frac{\alpha}{2}T_{ijk} = \varphi_i g_{jk} + \varphi_j g_{ik} - \varphi_k g_{ij}.$$

Застосовуючи техніку подібну до того, як ми робили вище, отримуємо:

$$2\varphi_j g_{ik} = 2\varphi_k g_{ij}. \quad (3.7)$$

Згорнемо (3.7) з $\psi^j = \psi_h g^{hj}$ та поділимо на 2. Це нам дає

$$\|\varphi\|^2 g_{ik} = \varphi_k \varphi_i,$$

де $\|\varphi\|^2 = \varphi^i \varphi^j g_{ij} = \varphi^j \varphi_j$. Як і вище, ми отримали протиріччя, і тому можемо сформулювати теорему.

Теорема 3.2. *Статистичний многовид (M, g, T) не дозволяє зв'язності Вейля по відношенню до метричного тензору g_{ij} , яка співпадатиме з α -зв'язністю ∇^α якщо $\alpha \neq 0$.*

3.2. Експоненціальні і мішані сімейства як структури Гессе

Експоненціальне сімейство розподілів ймовірності може бути представленим записом щільності у так званій канонічній формі:

$$p(x, \theta) = \exp \left(C(x) + \sum_{i=1}^n \theta^i F_i(x) - \psi(\theta) \right).$$

Для взаємно-однозначної параметризації є необхідним та достатнім, щоб $n+1$ функцій $\{F_1, \dots, F_n, 1\}$, де 1 означає сталу функцію, що тотожно дорівнює одиниці. Параметри θ^i мають назву *канонічних*. Очевидно, $\psi(\theta)$ не є довільною функцією. Вона однозначно визначається формулою:

$$\psi(\theta) = \ln \int \exp \left(C(x) + \sum_{i=1}^n \theta^i F_i(x) \right) dx.$$

Відповідно до формул (2.5) та (2.6) для експоненціального сімейства:

$$g_{ij} = \frac{\partial^2 \psi(\theta)}{\partial \theta^i \partial \theta^j}. \quad (3.8)$$

З (2.8), (2.12) та (3.8) випливає, що

$${}^0\Gamma_{ij,k} = \frac{1}{2} \frac{\partial^3 \psi(\theta)}{\partial \theta^i \partial \theta^j \partial \theta^k}.$$

Такі структури мають назву *структур Гессе* [3; 17]. Існує така теорема [17].

Теорема 3.3. Нехай (M, g) є рімановим многовидом. Тоді наступні умови є еквівалентними.

1. g_{ij} є метрика Гессе, тобто локально існує потенціальна функція $\psi(\theta)$ така, що $g_{ij} = \partial_i \partial_j \psi(\theta)$.

$$2. \quad \frac{\partial g_{ij}}{\partial \theta^k} = \frac{\partial g_{ik}}{\partial \theta^j}.$$

3. Символ Кристофеля першого роду є цілком симетричним: $\Gamma_{ij,k} = \Gamma_{ik,j}$.

Це з очевидністю випливає леми Пуанкаре.

Також з (2.9), випливає, що

$$T_{ijk} = \frac{\partial^3 \psi(\theta)}{\partial \theta^i \partial \theta^j \partial \theta^k}. \quad (3.9)$$

Беручи до уваги (2.10), маємо

$${}^\alpha\Gamma_{ij,k} = \frac{1-\alpha}{2} \frac{\partial^3 \psi(\theta)}{\partial \theta^i \partial \theta^j \partial \theta^k}. \quad (3.10)$$

Ми бачимо, що ${}^1\Gamma_{ij,k} = 0$, тобто $[\theta^i]$ – афінна система координат і наш статистичний многовид є 1-плоский. Використовуючи (1.11), обчислимо ріманів тензор кривини:

$${}^\alpha R_{ijk}^h = \frac{1-\alpha^2}{2} \left(g^{ht} T_{tsk} g^{sp} T_{ijp} - g^{ht} T_{tsj} g^{sp} T_{ikp} \right). \quad (3.11)$$

Тут g^{ht} -матриця, обернена до g_{ij} , що розрахована за формулою (3.8), T_{ijk} отриманий згідно (3.9). Ми можемо бачити, що тензор кривини стає

нульовим не тільки, якщо $\alpha = 1$, але і у випадку коли $\alpha = -1$. Це означає, що також існує координатна система в якій ${}^{-1}\Gamma_{ij,k}$ тотожно дорівнює нулю. Щоб отримати таку координатну систему ми здійснимо так зване перетворення Лежандра:

$$\eta_i = \frac{\partial \psi(\theta)}{\partial \theta^i}.$$

Воно дає нам дуальну опуклу функцію:

$$\varphi(\eta) = \theta^i(\eta) \eta_i - \psi(\theta(\eta))$$

де

$$\theta^i = \frac{\partial \varphi(\eta)}{\partial \eta_i}$$

є оберненим перетворенням.

Ми отримуємо, що [3]

$${}^{\alpha}\Gamma^{ij,k} = \frac{1+\alpha}{2} \frac{\partial^3 \varphi(\eta)}{\partial \eta_i \partial \eta_j \partial \eta_k}.$$

Можна побачити, що коефіцієнти зв'язності обертаються в нуль при $\alpha = -1$.

Зауваження 3.1. Порівнюючи формули (3.9) та (3.10) можна записати:

$${}^{\alpha}\Gamma_{ij,k} = \frac{1-\alpha}{2} T_{ijk}.$$

Тим не менш, ми цього намагалися уникати. Справа в тому, що воно є вірним лише для даної системи координат. Якщо ми перейдемо до іншої, то взагалі кажучи, ця рівність перестане бути рівною, оскільки об'єкт зв'язності ${}^{\alpha}\Gamma_{ij,k}$ перетворюється за законом (1.3), а тензор T_{ijk} за формулами (1.2).

Тепер розглянемо n -вимірну модель, виражену у термінах лінійно незалежних функцій $\{C, F_1, \dots, F_n\}$ як

$$p(x, \theta) = C(x) + \sum_{i=1}^n \theta^i F_i(x). \quad (3.12)$$

Формула (3.12) визначає так зване *мішане сімейство* з *мішаними параметрами* θ^i . Для мішаного сімейства, застосувавши (2.3) маємо

$$\partial_i l_\theta(x, \theta) = \frac{F_i(x)}{p(x, \theta)}, \quad \partial_i \partial_j l_\theta(x, \theta) = -\frac{F_i(x) F_j(x)}{(p(x, \theta))^2}.$$

Ми бачимо, що

$$\partial_i \partial_j l_\theta(x, \theta) + \partial_i l_\theta(x, \theta) \partial_j l_\theta(x, \theta) = 0. \quad (3.13)$$

З (3.12) та (2.11) випливає, що ${}^{-1}\Gamma_{ij,k} = 0$.

Для довільного α маємо

$${}^\alpha\Gamma_{ij,k}(\theta) = -\frac{1+\alpha}{2} E[\partial_i l_\theta(x, \theta) \partial_j l_\theta(x, \theta) \partial_k l_\theta(x, \theta)]. \quad (3.14)$$

З (3.14) випливає, що при $\alpha=0$ ми отримуємо цілком симетричний символ Крістофеля першого роду для метрики Фішера. Згідно із теоремою 3.3 це означає, що на многовиді існує потенціальна функція $\psi(\theta)$, така що $g_{ij} = \partial_i \partial_j \psi(\theta)$. Таким чином, ми довели теорему.

Теорема 3.4. *Інформаційна матриця Фішера обчислена для моделі мішаного сімейства є метрикою Гессе.*

Розглянемо довільний статистичний многовид (M, g, T) . Нехай для деякого значення α усі компоненти ${}^\alpha\Gamma_{ij,k}$ перетворюються в нуль. Згідно з (2.10).

$${}^0\Gamma_{ij,k} = \frac{\alpha}{2} T_{ijk}. \quad (3.15)$$

Ми знаємо, що згідно з означенням асиметрії (skewness) T_{ijk} є цілком симетричний. Тоді внаслідок (3.15) символ Крістофеля першого роду ${}^0\Gamma_{ij,k}$ є цілком симетричним. Ми маємо теорему.

Теорема 3.5. *Нехай статистичний многовид (M, g, T) і для деякого значення компоненти α -зв'язності тотожно обертаються в нуль. Тоді його інформаційна матриця Фішера g є метрикою Гессе.*

Стосовно структур Гессе варто додати ще одне важливе міркування. Справа в тому, що «гессевість» многовида не є інваріантною властивістю для перетворень координат, виключаючи лінійні [3; 17]. Деякий многовид може бути многовидом Гессе для однієї координатної системи, і не бути многовидом Гессе для іншої. Розглянемо випадок, коли метрика многовиду, взагалі кажучи, не є метрикою

Гессе по відношенню до старої координатної системи $(x_1, \dots, x_n)$, але допускає таку трансформацію координат, що в новій координатній системі $(y_1, \dots, y_n)$, многовид має метрику Гессе.

Відомо, що символ Крістофеля першого роду у новій координатній системі є представленим в такому вигляді [14]:

$${}^0\bar{\Gamma}_{ij,k}(y) = {}^0\Gamma_{\alpha\beta,\gamma}(x) \frac{\partial x^\alpha}{\partial y^i} \frac{\partial x^\beta}{\partial y^j} \frac{\partial x^\gamma}{\partial y^k} + \frac{\partial^2 x^\alpha}{\partial y^i \partial y^j} g_{\alpha\beta} \frac{\partial x^\beta}{\partial y^k}. \quad (3.16)$$

Згідно з теоремою 3.3 рівняння:

$${}^0\bar{\Gamma}_{ij,k}(y) = {}^0\bar{\Gamma}_{ik,j}(y) \quad (3.17)$$

має виконуватись тотожно в нових координатах. Беручі до уваги (3.16) та (3.17) отримуємо:

$$\begin{aligned} & {}^0\Gamma_{\alpha\beta,\gamma}(x) \frac{\partial x^\alpha}{\partial y^i} \frac{\partial x^\beta}{\partial y^j} \frac{\partial x^\gamma}{\partial y^k} - {}^0\Gamma_{\alpha\beta,\gamma}(x) \frac{\partial x^\alpha}{\partial y^i} \frac{\partial x^\beta}{\partial y^k} \frac{\partial x^\gamma}{\partial y^j} + \\ & + \frac{\partial^2 x^\alpha}{\partial y^i \partial y^j} g_{\alpha\beta} \frac{\partial x^\beta}{\partial y^k} - \frac{\partial^2 x^\alpha}{\partial y^i \partial y^k} g_{\alpha\beta} \frac{\partial x^\beta}{\partial y^j} = 0. \end{aligned} \quad (3.18)$$

Очевидно, якщо система диференціальних рівнянь (3.18) має розв'язок відносно функцій $x^i = x^i(y_1, \dots, y_n)$, $i = \overline{1, n}$, тоді розглянутий многовид локально допускає можливість введення метрики Гессе.

3.3. Приклади експоненціальних сімейств

1. Нормальний розподіл.

$$C(x) = 0, \quad F_1(x) = x, \quad F_2(x) = x^2, \quad \theta^1 = \frac{\mu}{\sigma^2}, \quad \theta^2 = -\frac{1}{2\sigma^2},$$

$$\psi(\theta) = \frac{\mu^2}{2\sigma^2} + \ln(\sigma\sqrt{2\pi}) = -\frac{(\theta^1)^2}{4\theta^2} + \frac{1}{2} \ln\left(-\frac{\pi}{\theta^2}\right).$$

2. Багатовимірний нормальний розподіл. Тут можна покласти

$$C(x) = 0, \quad F_i(x) = x_i, \quad F_{ij}(x) = x_i x_j, \quad (i \leq j),$$

$$\theta^i = \sum_j (\Sigma^{-1})^{ij} \mu_j, \quad \theta^{ii} = -\frac{1}{2} (\Sigma^{-1})^{ii}, \quad \theta^{ij} = -(\Sigma^{-1})^{ij},$$

і також

$$F_A(x) = x, \quad F_B(x) = xx^T, \quad \theta^A = \Sigma^{-1}\mu, \quad \theta^B = -\frac{1}{2}\Sigma^{-1}.$$

В результаті маємо

$$\begin{aligned} p(x, \theta) &= \exp \left(\sum_{1 \leq i \leq k} \theta^i F_i(x) + \sum_{1 \leq i \leq j \leq k} \theta^{ij} F_{ij}(x) - \psi(\theta) \right) = \\ &= \exp \left((\theta^A)^T F_A(x) + \text{tr}(\theta^B F_B(x)) - \psi(\theta) \right), \end{aligned}$$

де

$$\begin{aligned} \psi(\theta) &= \frac{1}{2}\mu^T \Sigma^{-1}\mu + \frac{1}{2}\ln(\det(2\pi\Sigma)) = \\ &= -\frac{1}{4}(\theta^A)^T (\theta^B)^{-1} \theta^A + \frac{1}{2}\ln(-\pi(\theta^B)^{-1}). \end{aligned}$$

3. Розподіл Пуассона.

$$\begin{aligned} C(x) &= -\ln(x!), \quad F(x) = x, \quad \theta = \ln \xi, \\ \psi(\theta) &= \xi = e^\xi. \end{aligned}$$

4. Розподіл дискретної випадкової величини із скінченною кількістю можливих значень.

$$\begin{aligned} C(x) &= 0, \quad F_i(x) = \begin{cases} 1(x = x_i) \\ 0(x \neq x_i). \end{cases} \\ \theta^i &= \ln \frac{p(x_i)}{p(x_0)} = \ln \frac{p(x_i)}{1 - \sum_{i=1}^n \xi^i}, \quad (i = 1, 2, \dots, n), \\ \psi(\theta) &= -\ln(p(x_0)) = -\ln \left(1 - \sum_{i=1}^n \xi^i \right) = \ln \left(1 + \sum_{i=1}^n \exp(\theta^i) \right). \end{aligned}$$

3.4. e -зв'язність та m -зв'язність на статистичному многовиді

Порівняємо вирази для коефіцієнтів α -зв'язності експоненціального та мішаного сімейств, тобто вирази (3.10) та (3.14) відповідно:

Експоненціальне сімейство:

$${}^\alpha \Gamma_{ij,k} = \frac{1-\alpha}{2} \frac{\partial^3 \psi(\theta)}{\partial \theta^i \partial \theta^j \partial \theta^k}.$$

Мішане сімейство:

$${}^{\alpha}\Gamma_{ij,k}(\theta) = -\frac{1+\alpha}{2} E\left[\partial_i l_{\theta}(x, \theta) \partial_j l_{\theta}(x, \theta) \partial_k l_{\theta}(x, \theta)\right].$$

Ми бачимо, що у випадку експоненціального сімейства коефіцієнти зв'язності стають нульовими, якщо $\alpha=1$, а для мішаного сімейства це ж саме відбувається коли $\alpha=-1$. У зв'язку з цим, систему координат, для якої ${}^1\Gamma_{ij,k}=0$ ми називаємо *e-координатною системою*, а в якій таке «обнулення» відбувається при $\alpha=-1$, тобто ${}^{-1}\Gamma_{ij,k}=0$ – називаємо відповідно *t-координатною системою*.

4. Статистичний висновок і геометрія

4.1. Достатня статистика і нерівність Рао – Крамера

Перш ніж застосовувати апарат інформаційної геометрії розглянемо деякі фундаментальні поняття статистики.

Нехай задана регулярна статистична модель:

$$M = \{p(x, \xi)\}, \quad (4.1)$$

і також відображення $F: \mathcal{X} \rightarrow \mathcal{Y}$, що перетворює значення випадкової величини $X \in \mathcal{X}$ у $Y = F(X)$. Тоді даний розподіл $p(x, \xi)$ перетворюється у $q(y, \xi)$, що задає поведінку іншої випадкової величини, а саме Y . Ми кажемо, що $F(x)$ є **достатньою статистикою** для моделі M якщо $p(x, \xi)$ можна представити наступним чином:

$$p(x, \xi) = q(F(x), \xi) r(x). \quad (4.2)$$

Якщо ми запишемо (4.2) у вигляді $r(x) = \frac{p(x, \xi)}{q(F(x), \xi)}$,

то стає зрозумілим, що $F(x)$ приймає такі значення, що є узгодженими з певним набором значень параметрів ξ . Це означає, що та частина розподілу $p(x, \xi)$, що залежить від ξ цілком зосереджена в розподілі $q(y, \xi)$ величини $Y = F(X)$ і тому для того, щоб оцінити невідомі параметри ξ (невідомий розподіл $p(x, \xi)$) достатньо знати лише значення Y [3]. Отже, $F(x)$ можна застосовувати для оцінювання значень параметрів ξ . Рівність (4.2) ще

має назву теореми про факторізацію. З неї ми маємо дуже просте і водночас потужне знаряддя для статистичного дослідження. Працює воно наступним чином.

Нехай задана модель (4.1) і запропоновано функцію $y = F(x)$, що ми розраховуємо представити розподіл $p(x, \xi)$ у вигляді добутку:

$$p(x, \xi) = q(F(x), \xi) r(x).$$

Тоді нам потрібно розглянути індуковану модель:

$$M_F \stackrel{\text{def}}{=} \{q(F(x), \xi)\}. \quad (4.3)$$

І порівняти її з даною, тобто (4.1).

Для цього ми маємо знайти за формулою (2.5) метрики Фішера $G(\xi) = (g_{ij}(\xi))$ для моделі (4.1) і для індукованої моделі (4.3), яку позначимо $G^F(\xi) = (g^F_{ij}(\xi))$. З цього приводу існує теорема [3].

Теорема 4.1 Інформаційна матриця Фішера $G^F(\xi) = (g^F_{ij}(\xi))$ обчислена для індукованої моделі $M_F = \{q(F(x), \xi)\}$ задовільняє нерівність

$$G^F(\xi) \leq G(\xi),$$

в тому сенсі, що відповідна квадратична форма для матриці $\Delta G(\xi) = G(\xi) - G^F(\xi)$ є позитивно напіввизначена. Необхідна та достатня умова для того, щоб статистика $F(x)$ була достатньою є рівність $G(\xi) = G^F(\xi)$.

Кажуть, що різниця:

$$\Delta G(\xi) = G(\xi) - G^F(\xi) \quad (4.4)$$

визначає *втрату інформації (information loss)*.

Узагальнюючи (4.4) ми отримуємо ще три такі властивості інформаційної матриці Фішера, що є дуже широко застосовними у статистичних дослідженнях.

1. Нехай поряд з моделлю (4.1) матриця Фішера якої є $G(\xi)$ задано умовний розподіл $\kappa = \{\kappa(y|x) \geq 0; x \in \mathcal{X}, y \in \mathcal{Y}\}$, такий, що для будь-якого x :

$$\int \kappa(y|x) dy = 1,$$

і $G_{\kappa}(\xi)$ є матрицею Фішера для індукованої моделі:

$$q(y, \xi) = \int \kappa(y|x) p(x, \xi) dx.$$

Тоді,

$$G(\xi) = G_{\kappa}(\xi) + \Delta G(\xi). \quad (4.5)$$

З (4.1) випливає, що $G(\xi) \geq G_{\kappa}(\xi)$.

2. Нехай задано дві незалежних моделі виду (4.1):

$$M_1 = \{p_1(x_1, \xi)\} \text{ та } M_2 = \{p_2(x_2, \xi)\},$$

що мають той самий простір елементарних подій (sample set) $\mathcal{X}$ і той самий простір параметрів Ξ . Тоді, матриця Фішера $G_{12}(\xi)$ для моделі-добутку, із щільністю розподілу $p_{12}(x_1, x_2, \xi) = p_1(x_1, \xi) p_2(x_2, \xi)$ матиме вигляд:

$$G_{12}(\xi) = G_1(\xi) + G_2(\xi), \quad (4.6)$$

де $G_1(\xi)$ та $G_2(\xi)$ є відповідно матрицями Фішера моделей M_1 та M_2 .

Властивість (4.6) має назву *адитивності*.

3. Нехай задано дві незалежних моделі виду (4.1):

$$M_1 = \{p_1(x_1, \xi)\} \text{ та } M_2 = \{p_2(x_2, \xi)\},$$

що мають той самий простір елементарних подій $\mathcal{X}$ і той самий простір параметрів Ξ і відповідно мають матриці Фішера $G_1(\xi)$ та $G_2(\xi)$.

Тоді для моделі із щільністю розподілу:

$$p_{\lambda}(x, \xi) = \lambda p_1(x, \xi) + (1 - \lambda) p_2(x, \xi), 1 \leq \lambda \leq 1$$

матриця Фішера задовільнятиме нерівності:

$$G_{\lambda}(\xi) \leq \lambda G_1(\xi) + (1 - \lambda) G_2(\xi). \quad (4.7)$$

Рівняння (4.7) відображує властивість **опуклості** матриці Фішера.

Тепер розглянемо відому нерівність *Рао – Крамера*. Нехай дані x випадково генеруються згідно з розподілом, що є невідомим, але вважатимемо, що цей розподіл належить n -вимірній моделі (4.1):

$$M = \{p(x, \xi) \mid \xi = (\xi^1, \xi^2, \dots, \xi^m) \in \Xi\}.$$

Нам потрібно оцінити невідомий параметр $\xi = (\xi^1, \xi^2, \dots, \xi^m)$ за допомогою функції $\hat{\xi}(x)$, аргументом якої є статистичні дані.

Відображення

$$\hat{\xi} = (\hat{\xi}^1, \hat{\xi}^2, \dots, \hat{\xi}^m): \mathcal{X} \rightarrow \mathbb{R}^m \quad (4.8)$$

має назву *оцінки (estimator)*. Ми кажемо, що (4.8) є *незміщеною оцінкою*, якщо математичне сподівання відносно простору елементарних подій відповідає рівності:

$$E_{\xi} [\hat{\xi}(x)] = \xi. \quad (4.9)$$

Середня квадратична похибка незміщеної оцінки $\hat{\xi}$ може бути виражено як коваріаційна матриця $V_{\xi} [\hat{\xi}] = (v_{\xi}^{ij})$ обчислена як [3]:

$$v_{\xi}^{ij} = E_{\xi} \left[\left(\hat{\xi}^i(x) - \xi^i \right) \left(\hat{\xi}^j(x) - \xi^j \right) \right]. \quad (4.10)$$

Теорема 4.2. (Рао – Крамера) Коваріаційна матриця $V_{\xi} [\hat{\xi}]$ для незміщеної оцінки $\hat{\xi}$ задовільняє нерівність:

$$V_{\xi} [\hat{\xi}] \geq G^{-1}(\xi). \quad (4.11)$$

В тому сенсі що $V_{\xi} [\hat{\xi}] - G^{-1}(\xi)$ визначає позитивно напіввизначену квадратичну форму.

$G^{-1}(\xi)$ є матрицею, оберненою до матриці Фішера $G(\xi)$.

Якщо для незміщеної оцінки $\hat{\xi}$ виконується рівність:

$$V_{\xi} [\hat{\xi}] = G^{-1}(\xi), \quad (4.12)$$

то кажуть, що $\hat{\xi}$ є *ефективною оцінкою*. Є очевидним, що ефективна оцінка є найкращою незміщеною оцінкою в тому сенсі, що її дисперсія є мінімальною серед дисперсій усіх незміщених оцінок. Обернене твердження, загалом, не є вірним. Справа у тому, що найкраща незміщена оцінка не завжди досягає межі Рао – Крамера, тобто $V_{\xi} [\hat{\xi}] = G^{-1}(\xi)$. Більше того, відомий випадок, коли зміщена оцінка виявилася такою, що має меншу середню квадратичну похибку, ніж ефективна оцінка.

Зазначимо, що може і не існувати ефективна оцінка моделі (4.1). Але слід наголосити, що завжди існує послідовність оцінок $\left\{ \hat{\xi}_N = \hat{\xi}_N(x_1, \bar{x}_2, \dots, x_N) \right\}_{N=1}^{\infty}$, що асимптотично досягається рівність (4.12) при $N \rightarrow \infty$. Така послідовність оцінок має назву *асимптотично ефективної оцінки*, або *ефективної оцінки першого порядку*. Таким чином, матриця $G^{-1}(\xi)$ відображує, наскільки близько асимптотично оптимальна оцінка попадає у окіл істинного значення параметра $\xi = (\xi^1, \xi^2, \dots, \xi^m)$. Інакше кажучи, чим менші значення компонент $G^{-1}(\xi)$, тим кращою є оцінка. Зрозуміло, що відповідно значення компонент $G(\xi)$ мають бути якнайбільшими.

4.2. Оцінювання базоване на незалежних дослідженнях

Розглянемо сімейство розподілів ймовірностей $M = \{p(x, \xi)\}$ що залежить від m параметрів $\xi = (\xi^1, \xi^2, \dots, \xi^m)$. Згідно прийнятним умовам регулярності³, ми розглядаємо M як m -вимірний многовид, в якому $\xi = (\xi^1, \xi^2, \dots, \xi^m)$ виступає в якості локальної системи координат. Нехай у нас є N незалежних спостережень $x_1, x_2, \dots, x_N$ випадкової величини x , яка в загальному випадку є n -вимірною, тобто кожне спостереження виступає у вигляді деякого n -вимірного вектора. Ця випадкова величина розподілена згідно щільністю розподілу $p(x, \xi)$. Позначимо результати наших спостережень $x^N = (x_1, x_2, \dots, x_N)$.

Мета статистичного висновку – вивести розподіл ймовірності, базуючись на даних, що є вибіркою N значень. *Оцінювання* – це такий тип задач, де метою є знайти оцінку $\hat{\xi} = (\hat{\xi}^1, \hat{\xi}^2, \dots, \hat{\xi}^m)$ для параметру (параметрів) $\xi = (\xi^1, \xi^2, \dots, \xi^m)$. З іншого боку, існує тип задач перевірки істинності гіпотез, де метою є визначення що саме треба зробити – чи прийняти «нульову» гіпотезу $H_0 : \xi = \xi_0$ всупереч альтернативної гіпотези $H_1 : \xi \neq \xi_0$, чи навпаки, відкинути.

³ Ми маємо на увазі, що усі функції, використані в розрахунках локально мають неперервні похідні достатньо високого порядку.

Очевидно, що розподіл ймовірностей випадкової величини $x^N = (x_1, x_2, \dots, x_N)$ може бути записаний з використанням функції розподілу результату одиничного випробування $p(x, \xi)$ як

$$p_N(x^N, \xi) = \prod_{t=1}^N p(x_t, \xi).$$

Нагадаємо, що у загальному випадку x_t є n -вимірним вектором $x_t = (x_{1t}, x_{2t}, \dots, x_{nt})$. Отже ми маємо:

$$\ln(p_N(x^N, \xi)) = \sum_{t=1}^N \ln(p(x_t, \xi)). \quad (4.13)$$

Розглядаючи x^N як випадкову змінну, ми отримуємо модель $M_N = \{p(x^N, \xi)\}$, яка подібно $M = M_1$ є многовидом з $\xi = (\xi^1, \xi^2, \dots, \xi^m)$ як координатної системи. Тому, з рівняння (4.13) і відповідних означень, ми знайдемо об'єкти геометричної структури на многовиді M_N :

$${}^N g_{ij}(\xi) = N g_{ij}(\xi), \quad (4.14)$$

$${}^{N(\alpha)} \bar{\Gamma}_{ij,k}(\xi) = {}^{(\alpha)} \bar{\Gamma}_{ij,k}(\xi). \quad (4.15)$$

Тут $g_{ij}(\xi)$ та ${}^{(\alpha)} \bar{\Gamma}_{ij,k}(\xi)$ – відповідно метрика Фішера та об'єкт α -зв'язності моделі (4.1) розраховані за формулами (2.5) та (2.11). Як ми бачимо, геометрична структура, яку ми ввели на M_N фактично є такою, як і на M , тільки відрізняється масштабуючим множником N .

Отже, для дослідження геометрії моделі M_N цілком достатньо використовувати геометрію моделі (4.1). Згідно із загальною теорією оцінювання, оцінка є визначеною як функція на N точках, які власне і складають вибірку $x^N = (x_1, x_2, \dots, x_N)$:

$$\hat{\xi}_N = \hat{\xi}(x^N) = \hat{\xi}(x_1, x_2, \dots, x_N). \quad (4.16)$$

Якщо ми розглядаємо x^N як випадкову змінну, то маємо розглядати $\hat{\xi}_N$ теж як випадкову змінну. Крім того, нам потрібна умова, якій має задовольняти $\hat{\xi}$, щоб гарантувати близькість оцінки $\hat{\xi}$ до реального значення параметра ξ , що входить до аналітичного виразу розподілу $p(x, \xi)$. Такою умовою є незміщеність. Грубо кажучи, вона полягає в тому, випадкова величина $\hat{\xi}$ є розподіленою навколо ξ . Цей факт

був вже записаний як (4.9). для випадку $N = 1$. Ми можемо розширити цю формулу для $N > 1$.

$$E_{\xi} [\hat{\xi}] = \xi. \quad (4.17)$$

Ми бачимо, що формули (4.9) та (4.17) фактично співпадають. Але якщо у випадку $N=1$ математичне сподівання функції $\hat{\xi}(x_1, x_2, \dots, x_n)$ n -вимірної випадкової величини (4.9) обчислюється як n -кратний інтеграл, то коли для $N > 1$ формула (4.17) виглядатиме як інтеграл кратості $n \cdot N$ в якому замість $p(x, \xi)$ фігуруватиме розподіл $p_N(x^N, \xi)$.

Середня квадратична похибка використовується як засіб вимірювання якості оцінки $\hat{\xi}$ і може бути виражено як коваріаційна матриця $V_{\xi} [\hat{\xi}] = (v_{\xi}^{ij})$ обчислена як [3]:

$$v_{\xi}^{ij} = E_{\xi} \left[\left(\hat{\xi}^i(x) - \xi^i \right) \left(\hat{\xi}^j(x) - \xi^j \right) \right]. \quad (4.18)$$

Як ми бачимо, (4.18) повністю співпадає з (4.10), але як у випадку (4.17) – інтегрування проводиться не згідно розподілу $p(x, \xi)$, а $p_N(x^N, \xi)$.

Варто зазначити, що як незміщеність, так і середня квадратична похибка не є геометрично інваріантними. Тобто вони суттєво залежать від вибору координатної системи. Нижня межа середньої квадратичної похибки незміщеної оцінки $\hat{\xi}$ дається нерівністю Рао – Крамера (Теорема 4.2), яке зараз приймає форму:

$$(v_{\xi}^{ij}) \geq \frac{1}{N} (g^{ij}(\xi)). \quad (4.19)$$

Тут $g^{ij}(\xi)$ означає матрицю, обернену до інформаційної матриці Фішера $(g_{ij}(\xi))$ моделі M . Множник $\frac{1}{N}$ на відміну від (4.11) виникає внаслідок того, що $(g_{ij}(\xi))$ розраховано для моделі з розподілом $p(x, \xi)$, а (v_{ξ}^{ij}) вже обчислено для $p_N(x^N, \xi)$.

4.3. Достатня статистика у статистичних дослідженнях

У секції 4.1 ми ввели поняття достатньої статистики слідуючи [3]. Однак для більш глибокого розуміння варто розібрати застосування поняття достатньої статистики на прикладах, пов'язаних

із статистичним дослідженням. Перш за все наголосимо, що статистичне дослідження – це в першу чергу отримання репрезентативної вибірки $X_1, X_2, \dots, X_N$. В контексті цієї секції *статистикою* називатимемо будь-яку функцію вигляду:

$$Y = F(X_1, X_2, \dots, X_N).$$

Зрозуміло, що дослідника цікавитимуть саме такі функції від значень, що прийняли елементи випадкової вибірки, що можна використовувати як оцінки значень невідомих йому параметрів. Питання полягає в тому, як знайти таку функцію, щоб здобувала з вибірки саме ту інформацію, що характеризує генеральну сукупність. В зв'язку з цим ми можемо дати таке евристичне означення *достатньої статистики* [19].

Ми кажемо, що $Y = F(X_1, X_2, \dots, X_N)$ є *достатньою статистикою*, якщо для оцінювання параметру ξ знання лише одного числового значення Y є цілком рівноцінним знанню всієї вибірки. Тобто, функція, яку ми називаємо *достатньою статистикою* екстрагує з вибірки з N значень – лише одне, але таке, яке характеризує вибірку в цілому.

Для знаходження такої функції застосовують вже згадану у секції 4.1 теорему про факторизацію. Але перш ніж її сформулювати у вигляді, застосовному у статистиці, зробимо деякі пояснення. Як ми вже писали, що розподіл ймовірностей випадкової величини $x^N = (x_1, x_2, \dots, x_N)$ може бути записаний з використанням функції розподілу результату одиничного випробування $p(x, \xi)$ як

$$p_N(x^N, \xi) = \prod_{t=1}^N p(x_t, \xi).$$

Тобто, в таких дослідженнях тип розподілу вважається вже відомим. Дослідник має знайти оцінки для невідомих параметрів.

Вибірка $X_1, X_2, \dots, X_N$ – фактично є результатом розігрування випадкових величин $x_1, x_2, \dots, x_N$, а функція

$$p(x_1, x_2, \dots, x_N | \xi) = p_N(x^N, \xi) = \prod_{t=1}^N p(x_t, \xi)$$

є їх спільним розподілом.

Таким чином, теорема про факторизацію виглядає так.

Теорема 4.3 Нехай $X_1, X_2, \dots, X_N$ утворює випадкову вибірку із спільним розподілом $p(x_1, x_2, \dots, x_N | \xi)$. Статистика $Y = F(X_1, X_2, \dots, X_N)$ є достатньою для параметра ξ , якщо і тільки якщо цей спільний розподіл може бути розкладеним на множники наступним чином:

$$p(x_1, x_2, \dots, x_N | \xi) = q(F(x_1, x_2, \dots, x_N), \xi) r(x_1, x_2, \dots, x_N), \quad (4.20)$$

де $q(F(x_1, x_2, \dots, x_N), \xi)$ та $r(x_1, x_2, \dots, x_N)$, є функціями, що приймають лише невід'ємні значення. Функція $r(x_1, x_2, \dots, x_N)$ залежить від усіх значень повного об'єму вибірки, але не залежить від оцінюваного невідомого параметру ξ . Функція $q(F(x_1, x_2, \dots, x_N), \xi)$ може залежати від параметру ξ , але може залежати від вибірки лише через значення функції $F(x_1, x_2, \dots, x_N)$.

Зауважимо, що $r(x_1, x_2, \dots, x_N)$ може залежати від інших параметрів, окрім ξ . Головне, щоб $r(x_1, x_2, \dots, x_N)$ не залежав саме від того параметру, для якого ми шукаємо достатню статистику. Додамо, що (4.20) – це є нічим іншим, як детальним представленням (4.2) у вигляді, що є застосовним у статистичних дослідженнях.

Розберемо деякі приклади

1. *Нормальний розподіл, невідоме математичне сподівання, дисперсія відома.* Отже, ми розглядаємо нормально розподілену генеральну сукупність, для якої треба оцінити математичне сподівання μ , дисперсію σ^2 вважаємо відомою. Спільний розподіл має вигляд:

$$\begin{aligned} p(x_1, x_2, \dots, x_N | \xi) &= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right) = \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma^2} \right) = \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 \right) \exp \left(-\frac{n\mu^2}{2\sigma^2} + \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i \right) = \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 \right) \exp \left(-\frac{n\mu^2}{2\sigma^2} + \frac{\mu}{\sigma^2} F(x_1, x_2, \dots, x_N) \right). \end{aligned}$$

Оскільки дисперсія σ^2 є відомою, можна вважати

$$r(x_1, x_2, \dots, x_N) = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 \right)$$

та

$$q(F(x_1, x_2, \dots, x_N), \xi) = \exp \left(-\frac{n\mu^2}{2\sigma^2} + \frac{\mu}{\sigma^2} F(x_1, x_2, \dots, x_N) \right),$$

де

$$F(x_1, x_2, \dots, x_N) = \sum_{i=1}^n x_i.$$

Згідно із теоремою про факторизацію (4.20) з цього випливає, що $\sum_{i=1}^n X_i$ є достатньою статистикою. З цього випливає, що середня вибірка

$$\bar{X} = \frac{1}{N} \sum_{i=1}^n X_i$$

також є достатньою статистикою.

2. Розподіл Пуассона, невідоме математичне сподівання [20].

Для цієї дискретно розподіленої випадкової величини потрібно знайти достатню статистику, виходячи з вибірки $X_1, X_2, \dots, X_N$.

Факторизація згідно із (4.20) дає:

$$p(x_1, x_2, \dots, x_N | \xi) = \prod_{t=1}^N p(x_t, \xi) = \prod_{t=1}^N e^{-\xi} \frac{\xi^{x_t}}{x_t!} = \frac{e^{-n\xi} \xi^{\sum_{t=1}^N x_t}}{\prod_{t=1}^N x_t!}.$$

де

$$q(F(x_1, x_2, \dots, x_N), \xi) = e^{-n\xi} \xi^{\sum_{t=1}^N x_t},$$

та

$$r(x_1, x_2, \dots, x_N) = \frac{1}{\prod_{t=1}^N x_t!}.$$

Тут також статистика

$$F(x_1, x_2, \dots, x_N) = \sum_{i=1}^n x_i$$

є достатньою.

3. *Однорідний розподіл [19].* X_i розподілена однорідно на проміжку $[0, b]$, параметр b невідомий. Тоді спільна щільність розподілу

$$p(x_1, x_2, \dots, x_N | \xi) = b^{-n} 1(x_i \leq b, i = 1, 2, \dots, N),$$

де $1(E)$ – так звана індикаторна функція. Вона дорівнює 1, якщо подія E відбулась, і дорівнює 0 в протилежному випадку. Ми бачимо, що $x_i \leq b, i = 1, 2, \dots, N$ тоді і тільки тоді, якщо $\max\{x_1, x_2, \dots, x_N\} \leq b$.

Таким чином отримуємо:

$$p(x_1, x_2, \dots, x_N | \xi) = b^{-n} 1(\max\{x_1, x_2, \dots, x_N\} \leq b).$$

Згідно з теоремою про факторизацію,

$$F(x_1, x_2, \dots, x_N) = \max\{x_1, x_2, \dots, x_N\}$$

і є ефективною статистикою.

4.4. Асимптотична теорія оцінювання

В асимптотичній теорії оцінювання засобом є представлення оцінки шляхом граничного переходу $N \rightarrow \infty$. В цьому випадку незміщеність не є настільки значущою, як тоді, коли N є фіксованим. Натомість ми зазвичай вимагаємо, щоб оцінка $\hat{\xi}_N = \hat{\xi}(x^N)$, або якщо бути точнішим, послідовність оцінок $\{\hat{\xi}_N; N = 1, 2, \dots\}$ була *слушною (конзистентною)* в тому сенсі, що для будь-якого параметру ξ його оцінка $\hat{\xi}(x^N)$ збігається за ймовірністю до ξ по мірі того, як $N \rightarrow \infty$, коли x^N є розподіленою за законом $p_N(x^N, \xi)$. Інакше кажучи, для будь-якого $\varepsilon > 0$

$$\lim_{N \rightarrow \infty} P(|\hat{\xi}_N - \xi| > \varepsilon) = 0,$$

тобто ймовірність того, що значення оцінки буде суттєво відрізнятися від справжнього значення параметру, прямує до нуля. Ймовірносний розподіл слушної оцінки $\hat{\xi}_N$ є усе більше і більше зконцентрованим навколо цього реального значення ξ разом з тим, що $N \rightarrow \infty$. Згідно з вимогами регулярності математичне сподівання випадкової величини, якою є $\hat{\xi}_N$, рівномірно збігається до ξ , отже:

$$\lim_{N \rightarrow \infty} E_{\xi} \left[\hat{\xi}_N \right] = \xi$$

та

$$\lim_{N \rightarrow \infty} \frac{\partial E_{\xi} \left[\hat{\xi}_N^i \right]}{\partial \xi^j} = \partial_j \xi^i = \delta_j^i.$$

Середня квадратична похибка такої асимптотично незміщеної оцінки задовільняє асимптотичну нерівність Рао – Крамера:

$$\lim_{N \rightarrow \infty} N \left(v_{\xi}^{ij} \left(\hat{\xi}_N^i \right) \right) \geq \left(g^{ij}(\xi) \right). \quad (4.21)$$

Слушна оцінка, що утворює знак рівності у (4.21) має назву асимптотично ефективної оцінки, або ефективної оцінки першого порядку. Така оцінка є оптимальною по відношенню до середньої квадратичної похибки з точністю порядку $\frac{1}{N}$. Виходячи з цього виникає питання, чи завжди існуватиме для довільної моделі асимптотично ефективна оцінка. На відміну від випадку ефективної оцінки для скінченного значення N , відповідь на це питання є позитивною і в якості прикладу можна навести метод максимальної правдоподібності (Maximum likelihood estimation, MLE), з яким ми познайомилися у розділі 2.1. Зараз з метою обґрунтування опишемо його більш формально, враховуючи викладений теоретичний апарат.

Маючи емпіричні дані x^N , розглянемо функцію $p_N(x^N, \xi)$ як функцію від $\xi = (\xi^1, \xi^2, \dots, \xi^m)$, і назвемо її *функцією правдоподібності*. Назвемо $\hat{\xi}_N$ *оцінкою максимальної правдоподібності* (maximum likelihood estimator, MLE), позначивши як $\hat{\xi}_{MLE}$, якщо вона задовільняє умові:

$$\max_{\xi} p_N(x^N, \xi) = p_N(x^N, \hat{\xi}_N), \quad (4.22)$$

тобто приймає значення для ξ , яке максимізує правдоподібність. Наступна теорема лежить в основі методу максимальної правдоподібності [3].

Теорема 4.3 Оцінка максимальної правдоподібності $\hat{\xi}_{MLE}$ є асимптотично ефективною. Це виражається наступною рівністю:

$$\lim_{N \rightarrow \infty} N v_{\xi}^{ij} \left(\hat{\xi}_{MLE} \right) = g^{ij}(\xi). \quad (4.23)$$

Фактично, є можливим довести більш сильне твердження, а саме, що оцінка $\hat{\xi}$, як випадкова величина має багатовимірний нормальний розподіл із математичним сподіванням ξ і коваріаційною матрицею $\frac{1}{N}(g^{ij}(\xi))$.

Середня квадратична похибка асимптотично ефективної оцінки $\hat{\xi}$ може бути записаною так:

$$v_{\xi}^{ij}(\hat{\xi}) = \frac{1}{N} g^{ij}(\xi) + O\left(\frac{1}{N^2}\right).$$

Аналіз доданку другого порядку є предметом *асимптотичної теорії вищого порядку*, і вона має розглядатись окремо.

5. Дуальні зв'язності

5.1. α -афінні многовиди і α -сімейства

В цій секції ми розширимо і узагальнемо деякі результати отримані стосовно α -зв'язності, отримані в розділах 2 та 3.

Ми трохи послабимо наші фундаментальні припущення, яких ми досі трималися. Досі ми розглядали статистичну модель $M = \{p(x, \xi)\}$, кожний елемент якої є розподілом ймовірностей $p(x, \xi) \in \mathcal{P}(\mathcal{X})$. В цій секції ми розглянемо випадки, коли

$$\int p(x, \xi) dx$$

може приймати довільні позитивні значення. Рівняння (2.5) та (2.11), що визначають відповідно метрику Фішера та α -зв'язність лишаються в силі, в той самий час (2.4) та (2.6) взагалі кажучи, не є такими, що обов'язково мають виконуватись.

Для кожного значення $\alpha \in \mathbb{R}$ задано:

$$L^{(\alpha)}(u) \stackrel{\text{def}}{=} \begin{cases} \frac{2}{1-\alpha} u^{\frac{1-\alpha}{2}} & (\alpha \neq 1) \\ \ln u & (\alpha = 1) \end{cases}, \quad (5.1)$$

$$l^{(\alpha)}(x, \xi) \stackrel{\text{def}}{=} L^{(\alpha)}(p(x, \xi)). \quad (5.2)$$

Зазначимо, що зокрема $l^{(\alpha)}(x, \xi) = l(x, \xi)$ що $l^{(-1)}(x, \xi) = p(x, \xi)$. Для дотичного вектора $X \in T_\xi(M)$ функцію від змінної x :

$$X^{(\alpha)}(x) \stackrel{\text{def}}{=} X l^{(\alpha)}(x, \xi) = X^i \frac{\partial l^{(\alpha)}(x, \xi)}{\partial \xi^i}. \quad (5.3)$$

ми вважатимемо його α -представленням. Відомі нам e -, m - та 0 -представлення, що ми згадували в секції 3.4 є нічим іншим, як випадками, що відповідають $\alpha = 1; -1; 0$. Оскільки:

$$\partial_i l^{(\alpha)} = p^{\frac{1-\alpha}{2}} \partial_i l \quad \text{та} \quad \partial_i \partial_j l^{(\alpha)} = p^{\frac{1-\alpha}{2}} \left(\partial_j \partial_i l + \frac{1-\alpha}{2} \partial_i l \partial_j l \right),$$

Ми можемо переписати (2.5) та (2.11) у вигляді:

$$g_{ij}(\xi) = \int \partial_i l^{(\alpha)}(x, \xi) \partial_j l^{(-\alpha)}(x, \xi) dx, \quad (5.4)$$

$${}^\alpha \Gamma_{ij,k}(\xi) = \int \partial_i \partial_j l^{(\alpha)}(x, \xi) \partial_k l^{(-\alpha)}(x, \xi) dx. \quad (5.5)$$

Ці рівняння можуть бути прочитаними в тому сенсі, що α -зв'язність на M індукована афінною структурою простору $\mathbb{R}^{\mathcal{X}}$ функцій на $\mathcal{X}$ через вкладання $\xi \rightarrow l^{(\alpha)}(x, \xi)$. Зауважимо, що у випадку коли $M = \{p(x, \xi)\}$ складається з розподілів ймовірності, відповідно для (2.4) та (2.6), ми маємо:

$$\int p^{\frac{1+\alpha}{2}}(x, \xi) \partial_i l^{(\alpha)}(x, \xi) dx = 0, \quad (5.6)$$

Зафіксуємо певне значення α . Якщо в деякій системі координат $[\theta^i]$:

$$\frac{\partial^2 l^{(\alpha)}(x, \theta)}{\partial \theta^i \partial \theta^j} = 0, \quad (5.7)$$

то з рівняння (5.5) ми бачимо, що $[\theta^i]$ є α -афінна система координат, отже $M = \{p(x, \xi)\}$ є α -плоским. Ми називатимемо такий M α -афінним *многовидом*. Умова (5.7) є еквівалентною існуванню таких функцій $\{F_1, \dots, F_n, C\}$, на $\mathcal{X}$, що:

$$l^{(\alpha)}(x, \xi) = C(x) + \sum_{i=1}^n \theta^i F_i(x). \quad (5.8)$$

5.2. Дуальність зв'язностей

Нехай M є мновидом із рімановою метрикою g_{ij} та афінними зв'язностями ∇ та ∇^* . В довільній точці p задано два дотичних до многовиду вектори $X(p), Y(p) \in T_p(M)$. Крім того, точка p і інша точка $q(p \neq q)$ з'єднані кривою $\gamma(t)$. Вектори X та Y паралельно переносяться уздовж кривої $\gamma(t)$. Це означає, що їх координати уздовж $\gamma(t)$ мають змінюватися згідно рівнянню (1.12). Але від класичного випадку ситуація відрізняється тим, що вектор X переноситься паралельно у зв'язності ∇ , а вектор Y – відповідно у зв'язності ∇^* .

Тобто,

$$\frac{\partial X^i}{\partial \xi^k} \dot{\gamma}^k + \Gamma_{jk}^i X^j(t) \dot{\gamma}^k(t) = 0, \quad (5.9)$$

$$\frac{\partial Y^i}{\partial \xi^k} \dot{\gamma}^k + \Gamma_{jk}^{*i} Y^j(t) \dot{\gamma}^k(t) = 0. \quad (5.10)$$

Тут $\dot{\gamma}^k(t)$ – дотичний до кривої $\gamma(t)$ вектор, а Γ_{jk}^i та Γ_{jk}^{*i} – коефіцієнти зв'язностей ∇ та ∇^* відповідно.

Вимагаємо, щоб скалярний добуток паралельно перенесених векторів зберігався, тобто

$$X^i(p) g_{ij} Y^j(p) = X^i(q) g_{ij} Y^j(q).$$

Це означає, що

$$\frac{d}{dt} (X^i(t) g_{ij} Y^j(t)) = 0, \quad (5.11)$$

або

$$\nabla_{\dot{\gamma}(t)} X^i(t) g_{ij} Y^j(t) + X^i(t) \tilde{\nabla}_{\dot{\gamma}(t)} g_{ij} Y^j(t) + X^i(t) g_{ij} \nabla_{\dot{\gamma}(t)}^* Y^j(t) = 0.$$

Для коваріантних похідних $\nabla_{\dot{\gamma}(t)} X^i$ та $\nabla_{\dot{\gamma}(t)}^* Y^i(t)$ зрозуміло, що вони мають вираховуватись за формулами (5.9) та (5.10), отже перший та останній доданок обертаються в нуль. Звідси випливає, що

$$X^i(t) \tilde{\nabla}_{\dot{\gamma}(t)} g_{ij} Y^j(t) = 0.$$

Поки що не зрозуміло, як має виглядати коваріантна похідна метричного тензору. Але оскільки крива $\gamma(t)$ є довільною, то очевидно що

$$\tilde{\nabla}_k g_{ij} = 0. \quad (5.12)$$

З іншого боку, мають задовільнятись (5.09) та (5.10).

Враховуючи (5.12), маємо

$$\tilde{\nabla}_k g_{ij} = \frac{\partial g_{ij}}{\partial x^k} - \Gamma_{ik}^l g_{lj} - \Gamma_{jk}^l g_{il} = 0.$$

Звідси випливає, що для будь-якого векторних полів (цього разу довільних)

$$\frac{\partial}{\partial \xi^k} (X^i g_{ij} Y^j) = \nabla_k X^i g_{ij} Y^j + X^i g_{ij} \nabla_k^* Y^j. \quad (5.13)$$

Пара зв'язностей, що відповідають рівнянню (5.13) мають назву *дуальних, або спряжених одна до одної по відношенню до метрики Фішера*.

Очевидно, що $(\nabla^*)^* = \nabla$. Також можна легко переконатись, що $\frac{\nabla + \nabla^*}{2}$ є метричною, тобто зв'язністю Леві – Чивіта. З іншого боку, навпаки, якщо нам відомо що ∇ та ∇' є симетричними зв'язностями і $\frac{\nabla + \nabla'}{2}$ є метричною, то можна стверджувати, що ∇' і ∇ є взаємно дуальними, тобто $\nabla' = \nabla^*$.

Проаналізувавши (5.4) та (5.5) з очевидністю отримуємо теорему.

Теорема 5.1. *Для будь-якої статистичної моделі α -зв'язність і $(-\alpha)$ -зв'язність є взаємно дуальними по відношенню до метрики Фішера.*

Для двох взаємно дуальних зв'язностей ∇ та ∇^* за формулами (1.11) можна розрахувати відповідні тензори кривини: R_{ijk}^h та R_{ijk}^{*h} . З очевидністю бачимо, що

$$g_{lh} R_{ijk}^h = -g_{lh} R_{ijk}^{*h}, \quad (5.14)$$

де g_{ij} -метрика Фішера. Враховуючи невиниродженість цієї матриці, отримуємо наступну теорему.

Теорема 5.2 Якщо тензор кривини однієї з двох взаємно дуальних зв'язностей ∇ та ∇^* дорівнює нулю, то нулю дорівнюватиме і тензор кривини іншої.

5.3. Дивергенції

Нехай M є многовидом і на ньому задана гладка функція $D(\cdot|\cdot): M \times M \rightarrow \mathbb{R}$. Позначати будемо $D(p(x, \xi)|p(x, \xi'))$, або $D(\xi|\xi')$ чи $D(p|p')$ в залежності від контекста задачі. Для будь-якої пари розподілів $p = p(x, \xi) \in M$ та $p' = p(x, \xi') \in M$ має виконуватись:

1. $D(p|p') \geq 0$,
 2. $D(p|p') = 0$, тоді і тільки тоді $p = p'$.
- (5.15)

Ця функція є у деякому сенсі мірою віддаленням однією точки від іншої і в цьому схожа на функцію відстані, але загалом, не задовольняє аксіомам симетричності та нерівності трикутника.

Введемо деякі позначення

$$\begin{aligned} D\left((\partial_i)_p | p'\right) &\stackrel{\text{def}}{=} \partial_i D(p|p'), \\ D\left((\partial_i)_p | (\partial_j)_{p'}\right) &\stackrel{\text{def}}{=} \partial_i \partial'_j D(p|p'), \\ D\left((\partial_i \partial_j)_p | (\partial_k)_{p'}\right) &\stackrel{\text{def}}{=} \partial_i \partial_j \partial'_k D(p|p'), \text{ тощо,} \end{aligned}$$

де вказані в правих частинах оператори $\partial_i = \frac{\partial}{\partial \xi^i}$ застосовуються

до відповідної змінної у $p = p(x, \xi)$, а $\partial'_i = \frac{\partial}{\partial \xi'^i}$ – до відповідної

змінної у $p' = p(x, \xi')$. Визначені таким чином операції можуть бути розширені до таких:

$$\begin{aligned} &D\left((X_1, \dots, X_l)_p | p'\right), \quad D\left(p | (Y_1, \dots, Y_m)_{p'}\right) \\ &\text{та } D\left((X_1, \dots, X_l)_p | (Y_1, \dots, Y_m)_{p'}\right), \end{aligned} \tag{5.16}$$

для будь-яких векторних полів, визначених на M .

Зауваження 5.1 В щойно наведених формулах векторні поля – це диференціальні оператори вигляду:

$$X = X^1 \frac{\partial}{\partial \xi^1} + X^2 \frac{\partial}{\partial \xi^2} + \dots + X^m \frac{\partial}{\partial \xi^m}.$$

Взагалі, якщо на деякому многовиді задані $X = X^1 e_1 + X^2 e_2 + \dots + X^m e_m$ – векторне поле, $f(x_1, x_2, \dots, x_m)$ – функція, в диференціальній геометрії вважається, що запис Xf означає:

$$Xf = X^1 \frac{\partial f}{\partial x^1} + X^2 \frac{\partial f}{\partial x^2} + \dots + X^m \frac{\partial f}{\partial x^m}.$$

Саме таким чином треба розуміти формули (5.16).

Тепер розглянемо обмеження цих формул на діагональ $\{(p, p) | p \in M\} \subset M \times M$ і позначимо відображення, індуковані на M таким чином:

$$D(X_1, \dots, X_l | \cdot) : p \rightarrow D((X_1, \dots, X_l)_p | p),$$

$$D(\cdot | Y_1, \dots, Y_m) : p \rightarrow D(p | (Y_1, \dots, Y_m)_p),$$

$$D(X_1, \dots, X_l | Y_1, \dots, Y_m) : p \rightarrow D((X_1, \dots, X_l)_p | (Y_1, \dots, Y_m)_p).$$

З рівнянь (5.14) випливає:

$$D(\partial_i | \cdot) = D(\cdot | \partial_i) = 0, \quad (5.17)$$

$$D(\partial_i \partial_j | \cdot) = D(\cdot | \partial_i \partial_j) = -D(\partial_i | \partial_j) \left(\stackrel{\text{def}}{=} g_{ij}^{(D)} \right). \quad (5.18)$$

Крім того, матриця $(g_{ij}^{(D)})$ породжує позитивно напіввизначену квадратичну форму. Якщо породжена форма строго позитивно визначена в усіх точках M , то ми кажемо, що D є *дивергенцією*, або *контрастною функцією* на M . Дивергенція однозначно задає ріманову метрику як:

$$\partial_i, \partial_j^{(D)} = g_{ij}^{(D)},$$

або у інваріантній формі:

$$X, Y^{(D)} = -D(X|Y).$$

Ця метрика з'являється, якщо ми випишемо для дивергенції D апроксимацію до другого порядку включно:

$$D(p|q) = \frac{1}{2} g_{ij}^{(D)}(q) \Delta \xi^i \Delta \xi^j + o(\Delta \xi^2), \quad (5.19)$$

де $\Delta \xi^i = \xi^i(p) - \xi^i(q)$ і $o(\Delta \xi^2)$ означає доданки, що прямують до нуля швидше, ніж $\Delta \xi^2$ по мірі наближення p до q . Додамо, що доданків першого порядку у (5.19) немає внаслідок (5.17).

Маючи дивергенцію D можна також задати афінну зв'язність $\nabla^{(D)}$ з коефіцієнтами:

$$\Gamma_{ij,k}^{(D)} = -D(\partial_i \partial_j | \partial_k), \quad (5.20)$$

або

$$\nabla_X^{(D)} Y, Z^{(D)} = -D(XY|Z).$$

Зауважимо, що отримана зв'язність – симетрична: $\Gamma_{ij,k}^{(D)} = \Gamma_{ji,k}^{(D)}$. Використовуючи метрику і коефіцієнти зв'язності можна отримати для дивергенції D апроксимацію третього порядку:

$$D(p|q) = \frac{1}{2} g_{ij}^{(D)}(q) \Delta \xi^i \Delta \xi^j + \frac{1}{6} h_{ijk}^{(D)}(q) \Delta \xi^i \Delta \xi^j \Delta \xi^k + o(\Delta \xi^3), \quad (5.21)$$

де

$$h_{ijk}^{(D)} = D(\partial_i \partial_j | \partial_k | \cdot).$$

Таким чином коефіцієнти $h_{ijk}^{(D)}$ пов'язані з $g_{ij}^{(D)}$ та $\Gamma_{ij,k}^{(D)}$ таким чином

$$h_{ijk}^{(D)} = \partial_i g_{jk}^{(D)} + \Gamma_{jk,i}^{(D)}. \quad (5.22)$$

І навпаки, як метрика так і коефіцієнти зв'язності визначаються через розкладання (5.21) та співвідношення (5.22).

Тепер розглянемо дивергенцію $D^*(p|q) = D(q|p)$. Її називають дуальною до $D(p|q)$. Розрахувавши метрику, використовуючи дуальну дивергенцію отримуємо:

$$g_{ij}^{(D^*)} = g_{ij}^{(D)}$$

і

$$\Gamma_{ij,k}^{(D^*)} = -D(\partial_k | \partial_i \partial_j). \quad (5.23)$$

Є очевидною теорема.

Теорема 5.2 Зв'язності $\nabla^{(D)}$ та $\nabla^{(D^*)}$ є взаємно дуальними по відношенню до метрики $g_{ij}^{(D)}$.

Використовуючи дуальні величини, отримуємо рівняння, що є аналогічними (5.21) і (5.22).

$$D(p|q) = D^*(q|p) = \frac{1}{2} g_{ij}^{(D)}(p) \Delta \xi^i \Delta \xi^j - \frac{1}{6} h_{ijk}^{(D^*)}(p) \Delta \xi^i \Delta \xi^j \Delta \xi^k + o(\Delta \xi^3), \quad (5.24)$$

$$h_{ijk}^{(D^*)} = D(\cdot | \partial_i \partial_j \partial_k) = \partial_i g_{jk}^{(D)} + \Gamma_{jk,i}^{(D^*)}. \quad (5.25)$$

Таким чином, ми бачимо, що будь-яка дивергенція породжує дуальну структуру з нульовим крученням. Навпаки, будь-який набір (g, ∇, ∇^*) , тобто метрика і пара узгоджених з нею взаємно дуальних зв'язностей може породжувати дивергенцію. Маючи таку трійку в координатній системі $\xi^1, \xi^2, \dots, \xi^m$ можна локально задати дивергенцію:

$$D(p|q) \stackrel{\text{def}}{=} \frac{1}{2} g_{ij}(q) \Delta \xi^i \Delta \xi^j + \frac{1}{6} h_{ijk}(q) \Delta \xi^i \Delta \xi^j \Delta \xi^k, \quad (5.26)$$

де $\Delta \xi^i = \xi^i(p) - \xi^i(q)$ і

$$h_{ijk}(q) \stackrel{\text{def}}{=} \partial_i g_{jk} + \Gamma_{jk,i} = \Gamma_{ij,k} + \Gamma_{ik,j}^* + \Gamma_{jk,i}. \quad (5.27)$$

Оскільки $\Gamma_{ik,j}^*$ відрізняється від $\Gamma_{ik,j}$ на величину, пропорційну симетричному по всіх трьох індексах коваріантному тензору (див. (2.10), (2.11)), то ми бачимо, що h_{ijk} має бути також абсолютно симетричною величиною. Тому з (5.26) та (5.27) ми отримуємо $D(\partial_i \partial_j | \cdot) = g_{ij}$ та $D(\partial_i \partial_j \partial_k | \cdot) = h_{ijk}$, з чого випливає, що трійка (g, ∇, ∇^*) породжує дивергенцію, яка в свою чергу однозначно породжує таку ж саме трійку. Додамо, що виконання умови невід'ємності у (5.15) забезпечується саме позитивною визначеністю метричного тензора g_{ij} . Тим не менш, дивергенцію, що відповідає цій трійці (g, ∇, ∇^*) – можна побудувати і іншим шляхом. Зокрема так:

$$D(p|q) \stackrel{\text{def}}{=} \frac{1}{2} g_{ij}(p) \Delta \xi^i \Delta \xi^j - \frac{1}{6} h_{ijk}^*(p) \Delta \xi^i \Delta \xi^j \Delta \xi^k,$$

$$\text{де } \Delta \xi^i = \xi^i(p) - \xi^i(q) \text{ і}$$

$$h_{ijk}^* \stackrel{\text{def}}{=} \partial_i g_{jk} + \Gamma_{jk,i}^* = \Gamma_{ij,k}^* + \Gamma_{ik,j}^* + \Gamma_{jk,i}^*.$$

Як ми вже казали, що запропоновані вирази для обчислення дивергенції отримані нами з локальних міркувань, але у [18] доведено, що для будь-якої дуальної структури існує глобально визначена дивергенція, що її породжує.

Тепер розглянемо дуже важливий клас дивергенцій, що є застосовними для дослідження статистичних моделей [19]. Нехай $f(u)$ є опуклою функцією. Для будь-якої пари розподілів $p(x, \xi)$ та $p(x, \xi)$ задаємо дивергенцію наступним чином:

$$D_f(p(x, \xi) | p(x, \xi)) \stackrel{\text{def}}{=} \int p(x, \xi) f\left(\frac{p(x, \xi)}{p(x, \xi)}\right) dx.$$

Визначена таким контрастна функція має назву f -дивергенції. Познайомимося з її властивостями. По-перше, згадаємо так звану нерівність Йєнсена. Вона полягає в тому, що для будь-якої опуклої функції $f(x)$ і будь-якого розподілу $p(x, \xi)$ завжди виконується

$$f\left(\int p(x, \xi) g(x) dx\right) \leq \int p(x, \xi) f(g(x)) dx.$$

Застосовуючи його для f -дивергенції, маємо:

$$\begin{aligned} D_f(p(x, \xi) | p(x, \xi)) &= \int p(x, \xi) f\left(\frac{p(x, \xi)}{p(x, \xi)}\right) dx \geq \\ &\geq f\left(\int p(x, \xi) \frac{p(x, \xi)}{p(x, \xi)} dx\right) = f\left(\int p(x, \xi) dx\right) = f(1). \end{aligned} \quad (5.28)$$

З іншого боку, якщо $\xi = \xi$, то згідно з (5.14) $D_f(p(x, \xi) | p(x, \xi)) = 0$, отже враховуючи (5.28) маємо, що $f(1) = 0$.

Важливим прикладом гладкої f -дивергенції є α -дивергенція, яка позначається $D^{(\alpha)} = D_{f(\alpha)}$ для дійсного числа α . Вона породжується функцією:

$$f^{(\alpha)}(u) \stackrel{\text{def}}{=} \begin{cases} \frac{4}{1-\alpha^2} \left(1 - u^{\frac{1+\alpha}{2}} \right) (\alpha \neq \pm 1) \\ u \ln u \quad (\alpha = 1) \\ -\ln u \quad (\alpha = -1) \end{cases}. \quad (5.29)$$

Для $\alpha \neq \pm 1$ ми маємо:

$$D^{(\alpha)}(p(x, \xi) | p(x, \xi)) = \frac{4}{1-\alpha^2} \left(1 - \int (p(x, \xi))^{\frac{1-\alpha}{2}} (p(x, \xi))^{\frac{1+\alpha}{2}} dx \right). \quad (5.30)$$

У випадку $\alpha = \pm 1$:

$$D^{(-1)}(p(x, \xi) | p(x, \xi)) = D^{(1)}(p(x, \xi) | p(x, \xi)) = \int p(x, \xi) \ln \left(\frac{p(x, \xi)}{p(x, \xi)} \right) dx. \quad (5.31)$$

Ми бачимо, що α -дивергенція $D^{(\alpha)}$ індукує структуру $(g, \nabla^{(\alpha)})$. Варто зазначити, що рівність

$$D^{(\alpha)}(p(x, \xi) | p(x, \xi)) = D^{(-\alpha)}(p(x, \xi) | p(x, \xi)).$$

Виконується для будь-якого α . Зокрема, $D^{(0)}(p(x, \xi) | p(x, \xi))$ є симетричною, більше того, $\sqrt{D^{(0)}(p(x, \xi) | p(x, \xi))}$ задовільняє усім аксіомам відстані. Дійсно, з (5.31) при $\alpha = 0$ маємо

$$\begin{aligned} D^{(0)}(p(x, \xi) | p(x, \xi)) &= 4 \left(1 - \int (p(x, \xi))^{\frac{1}{2}} (p(x, \xi))^{\frac{1}{2}} dx \right) = \\ &= 2 \left(2 - 2 \int (p(x, \xi))^{\frac{1}{2}} (p(x, \xi))^{\frac{1}{2}} dx \right) = \\ &= 2 \left(\int p(x, \xi) dx + \int p(x, \xi) dx - 2 \int (p(x, \xi))^{\frac{1}{2}} (p(x, \xi))^{\frac{1}{2}} dx \right) = \\ &= 2 \int \left(p(x, \xi) + p(x, \xi) - 2(p(x, \xi))^{\frac{1}{2}} (p(x, \xi))^{\frac{1}{2}} \right) dx = \end{aligned}$$

$$= 2 \int \left((p(x, \xi))^{\frac{1}{2}} - (p(x, \xi))^{-\frac{1}{2}} \right)^2 dx.$$

Ми тут суттєво використовували той факт, що

$$\int p(x, \xi) dx = \int p(x, \xi) dx = 1.$$

$\sqrt{D^{(0)}(p(x, \xi)|p(x, \xi))}$ має назву *відстані Гелінгера*.

Випадок $\alpha = \pm 1$, так звана ± 1 – дивергенція має декілька назв, зокрема *дивергенція Каллбека*, *інформація Кулбека-Ляйблера*, *відносна ентропія*, або просто дивергенція.

Знайдемо явні вирази для метрики та коефіцієнтів для $\nabla^{(\alpha)}$ зв'язності.

Щоб знайти метрику

$$g_{ij}^{(D)} = D(\partial_i \partial_j | \cdot),$$

знайдемо спочатку

$$\begin{aligned} & \frac{\partial}{\partial \xi^i} \frac{\partial}{\partial \xi^j} \left(\frac{4}{1-\alpha^2} \left(1 - \int (p(x, \xi))^{\frac{1-\alpha}{2}} (p(x, \xi'))^{\frac{1+\alpha}{2}} dx \right) \right) = \\ &= \frac{4}{1-\alpha^2} \frac{1-\alpha}{2} \frac{\partial}{\partial \xi^j} \left(- \int \frac{\partial p(x, \xi)}{\partial \xi^i} (p(x, \xi))^{-\frac{1+\alpha}{2}} (p(x, \xi'))^{\frac{1+\alpha}{2}} dx \right) = \\ &= - \frac{1}{1+\alpha} \left(\int \left(\frac{\frac{\partial p(x, \xi)}{\partial \xi^i} \frac{\partial p(x, \xi)}{\partial \xi^j} (p(x, \xi))^{-\frac{1+\alpha}{2}} - \right. \right. \\ & \quad \left. \left. - \frac{+\alpha}{\partial \xi^i} \frac{\partial p(x, \xi)}{\partial \xi^j} (p(x, \xi))^{-\frac{1+\alpha}{2}} \right) (p(x, \xi'))^{\frac{1+\alpha}{2}} dx \right) = \\ &= - \frac{2}{1+\alpha} \int \frac{\partial^2 p(x, \xi)}{\partial \xi^i \partial \xi^j} (p(x, \xi))^{-\frac{1+\alpha}{2}} (p(x, \xi'))^{\frac{1+\alpha}{2}} dx + \\ & \quad + \int \frac{\partial p(x, \xi)}{\partial \xi^i} \frac{\partial p(x, \xi)}{\partial \xi^j} (p(x, \xi))^{-\frac{3+\alpha}{2}} (p(x, \xi'))^{\frac{1+\alpha}{2}} dx. \end{aligned}$$

Враховуючи, що

$$D(\partial_i \partial_j | \cdot) = \frac{\partial}{\partial \xi^i} \frac{\partial}{\partial \xi^j} D(p(x, \xi) | p(x, \xi')) \Big|_{\xi=\xi'},$$

Отримуємо:

$$g_{ij}^{(D)} = D(\partial_i \partial_j | \cdot) = -\frac{2}{1+\alpha} \int \frac{\partial^2 p(x, \xi)}{\partial \xi^i \partial \xi^j} dx + \int \frac{\partial p(x, \xi)}{\partial \xi^i} \frac{\partial p(x, \xi)}{\partial \xi^j} \frac{1}{p(x, \xi)} dx.$$

Як ми бачимо, перший доданок внаслідок (2.4) обертається в нуль, а другий – власне і є інформаційною метрикою Фішера згідно з (2.5):

$$g_{ij}^{(D)} = D(\partial_i \partial_j | \cdot) = \int \frac{\partial p(x, \xi)}{\partial \xi^i} \frac{\partial p(x, \xi)}{\partial \xi^j} \frac{1}{p(x, \xi)} dx = E \left[\partial_i l_\xi \partial_j l_\xi \right].$$

Тепер знайдемо коефіцієнти зв'язності:

$$\Gamma_{ij,k}^{(D)} = -D(\partial_i \partial_j | \partial_k).$$

Спочатку знайдемо

$$\begin{aligned} & \frac{\partial}{\partial \xi'^k} \frac{\partial}{\partial \xi^i} \frac{\partial}{\partial \xi^j} D(p(x, \xi) | p(x, \xi')) = \\ & = \frac{\partial}{\partial \xi'^k} \left(-\frac{2}{1+\alpha} \int \frac{\partial^2 p(x, \xi)}{\partial \xi^i \partial \xi^j} (p(x, \xi))^{-\frac{1+\alpha}{2}} (p(x, \xi'))^{\frac{1+\alpha}{2}} dx + \right. \\ & \quad \left. + \int \frac{\partial p(x, \xi)}{\partial \xi^i} \frac{\partial p(x, \xi)}{\partial \xi^j} (p(x, \xi))^{-\frac{3+\alpha}{2}} (p(x, \xi'))^{\frac{1+\alpha}{2}} dx \right) = \\ & = -\int \frac{\partial^2 p(x, \xi)}{\partial \xi^i \partial \xi^j} (p(x, \xi))^{-\frac{1+\alpha}{2}} (p(x, \xi'))^{\frac{\alpha-1}{2}} \frac{\partial p(x, \xi)}{\partial \xi^k} dx + \\ & + \frac{1+\alpha}{2} \int \frac{\partial p(x, \xi)}{\partial \xi^i} \frac{\partial p(x, \xi)}{\partial \xi^j} (p(x, \xi))^{-\frac{3+\alpha}{2}} (p(x, \xi'))^{\frac{\alpha-1}{2}} \frac{\partial p(x, \xi')}{\partial \xi'^k} dx. \end{aligned}$$

Остаточно маємо:

$$\begin{aligned}\Gamma_{ij,k}^{(\alpha)} &= -D(\partial_i \partial_j | \partial_k) = -\frac{\partial}{\partial \xi^k} \frac{\partial}{\partial \xi^i} \frac{\partial}{\partial \xi^j} D(p(x, \xi) | p(x, \xi')) \Big|_{\xi=\xi} = \\ &= \int \frac{\partial^2 p(x, \xi)}{\partial \xi^i \partial \xi^j} \frac{\partial p(x, \xi)}{\partial \xi^k} \frac{1}{p(x, \xi)} dx - \\ &\quad - \frac{1+\alpha}{2} \int \frac{\partial p(x, \xi)}{\partial \xi^i} \frac{\partial p(x, \xi)}{\partial \xi^j} \frac{\partial p(x, \xi)}{\partial \xi^k} \frac{1}{(p(x, \xi))^2} dx.\end{aligned}\quad (5.32)$$

Це цілком збігається із формулою (2.12). Більше того, якщо ми повторимо наші обчислення метрики та коефіцієнтів зв'язності для дивергенції Кулбека – Ляйблера (5.31), то отримаємо:

$$\begin{aligned}g_{ij}^{(KL)} &= D(\partial_i \partial_j | \cdot) = \int \frac{\partial p(x, \xi)}{\partial \xi^i} \frac{\partial p(x, \xi)}{\partial \xi^j} \frac{1}{p(x, \xi)} dx, \\ {}^{(KL)}_{ij,k} &= \Gamma_{ij,k}^{(-1)} = \int \frac{\partial^2 p(x, \xi)}{\partial \xi^i \partial \xi^j} \frac{\partial p(x, \xi)}{\partial \xi^k} \frac{1}{p(x, \xi)} dx. \\ {}^{(KL*)}_{ij,k} &= \Gamma_{ij,k}^{(1)} = \int \frac{\partial^2 p(x, \xi)}{\partial \xi^i \partial \xi^j} \frac{\partial p(x, \xi)}{\partial \xi^k} \frac{1}{p(x, \xi)} dx - \\ &\quad - \int \frac{\partial p(x, \xi)}{\partial \xi^i} \frac{\partial p(x, \xi)}{\partial \xi^j} \frac{\partial p(x, \xi)}{\partial \xi^k} \frac{1}{(p(x, \xi))^2} dx.\end{aligned}$$

Це цілком збігатиметься з тим, якщо ми підставимо $\alpha = -1$ та $\alpha = 1$ у (5.22) відповідно.

5.4. Дуально плоскі простори

Нехай (g, ∇, ∇^*) – деяка дуальна структура на статистичному многовиді M . З теореми 5.2 випливає, що з того, що ∇ є плоскою – випливає, що плоскою також є і ∇^* . Зокрема, ми бачили, що експоненціальні структури є 1-плоскими, а мішані структури в свою чергу є -1 -плоскими (секція 3.2). Як виявилось, вони обидві є ± 1 -плоскими. З іншого боку, це вже було видно по виразу тензора кривини експоненціального простору (3.11):

$${}^{\alpha}R_{ijk}^h = \frac{1-\alpha^2}{2} (g^{ht}T_{tsk}g^{sp}T_{ijp} - g^{ht}T_{tsj}g^{sp}T_{ikp}).$$

Ми бачили, що він обертається в нуль не тільки коли $\alpha = \pm 1$.

В загалі, многовид M із заданою на ньому дуальною структурою (g, ∇, ∇^*) – позначатимемо (M, g, ∇, ∇^*) – будемо називати *дуально плоским простором*, якщо обидві здуальні зв'язності ∇ та ∇^* є плоскими.

При певних обставинах така властивість може розповсюджуватися і на підмноговиди. Нагадаємо (секція 1.8), що підмноговид $S \subset M$ називають *автопаралельним*, якщо для кожного вектора, дотичного $T_{x_0}(S)$ і кожної кривої γ , що виходить з x_0 , паралельний перенос цього вектору вздовж γ (відносно афінної зв'язності охоплюючого простору M) призводить до того, що перенесений вектор лишається дотичним до S . Звісно, треба завжди пояснювати в контексті якої зв'язності многовид є автопаралельним. Зараз, звичайно мова йтиме про α -зв'язності. Зрозуміло, що при одному значенні α вкладений підмноговид може бути автопаралельним, при іншому – ні. Тому, в таких випадках коректно говорити про α -автопаралельні многовиди.

Існує така теорема [3].

Теорема 5.3 Нехай (M, g, ∇, ∇^*) є дуально плоским простором. Якщо підмноговид $S \subset M$ є автопаралельним по відношенню до ∇ або ∇^* , тоді S є дуально плоским многовидом по відношенню до дуальної структури $(g_M, \nabla_M, \nabla_M^*)$, індукованою на S дуальною структурою (g, ∇, ∇^*) .

Розглянемо загальну структуру дуально плоского простору (M, g, ∇, ∇^*) . Безпосередньо з означення випливає існування на многовиді M ∇ -афінної координатної системи, яку ми позначимо $[\theta^i]$, і ∇^* -афінної координатної системи, яку ми позначимо відповідно $[\eta_j]$. Відповідно, задаємо диференціальні оператори $\partial_i = \frac{\partial}{\partial \theta^i}$, які визначають дотичні вектори до координатних кривих $[\theta^i]$. Аналогічно задаємо диференціальні оператори $\partial^j = \frac{\partial}{\partial \eta_j}$.

Зауваження 5.2 Ми вже писали, зокрема в Зауваженні 5.1, що дуже часто диференціальні оператори інтепретують як вектори.

Зокрема, диференціальний оператор $\partial_i = \frac{\partial}{\partial x^i}$ прямо називають базисним вектором, що є дотичним до координатної кривої x^i . Це пов'язано в першу чергу з тим, що при заміні координат диференціальні оператори перетворюються так само, як і базисні вектори. Порівняйте:

$$\text{Перетворення базисних векторів } e'_i = \frac{\partial x^i}{\partial x^{i'}} e_i,$$

$$\text{і перетворення диференціальних операторів } \frac{\partial}{\partial x^{i'}} = \frac{\partial x^i}{\partial x^{i'}} \frac{\partial}{\partial x^i}.$$

Це має прояв, наприклад, коли ми перераховуємо градієнт у новій системі координат, користуючись ланцюговим правилом:

$$\frac{\partial f}{\partial x^{i'}} = \frac{\partial x^i}{\partial x^{i'}} \frac{\partial f}{\partial x^i}.$$

згідно з ланцюговим правилом. Формула (1.2), згідно з якою відбувається перерахунок компонент тензора загального вигляду – якраз є узагальненням щойно згаданого правила перерахунку градієнта.

Отже, надалі ми будемо сприймати $\partial_i = \frac{\partial}{\partial x^i}$ як базисний вектор e_i , що є дотичним до координатної кривої x^i або як векторне поле у якого усі контраваріантні координати дорівнюють нулю, крім i -ї компоненти, яка дорівнює одиниці.

Зауваження 5.3 Дуже важливо відрізнати загальну ситуацію, коли ми позначали систему координат статистичного многовиду як $[\xi^i]$ від нинішньої, коли у нас є дві спеціальні системи координат, а саме, $[\theta^i]$ та $[\eta_j]$. В першому, загальному випадку системи координат $[\xi^i]$ верхній і нижній індекси є відповідно контра- і коваріантними координатами тензорів, що враховується при перерахунку за формулою (1.2).

Коли ми працюємо в системах координат $[\theta^i]$ та $[\eta_j]$ – то ми стикаємося з такими властивостями, які проявляються лише конкретно в даних спеціальних системах координат. Тому

врахування контра- і коваріантності при заміні систем координат – просто втрачає сенс. У зв'язку з тим – індекси розташовують, головним чином, згідно з тим, наскільки це буде зручно при застосування правила Ейнштейна для сумациї за одноіменними індексами. Наприклад, вектор для базисного вектору $\partial^j = \frac{\partial}{\partial \eta_j}$ в системі координат $[\eta_j]$ ми будемо використовувати в тому числі позначення e^j . Це ні в якому разі не означає що такі вектори змінюватимуться в загальному випадку до чогось «контраваріантно».

Але, узагальнюючи, можна рекомендувати таке правило. У системі координат $[\eta_j]$, індекси що зазвичай контраваріантні – записуються як коваріантні і навпаки.

Очевидно, що векторні $\partial_i = \frac{\partial}{\partial \theta^i}$, будуть плоскими у зв'язності ∇ в системі координат $[\theta^i]$, а $\partial^j = \frac{\partial}{\partial \eta_j}$ – плоскими у зв'язності ∇^* в системі координат $[\eta_j]$:

$$\nabla_{\frac{\partial}{\partial \theta^i}} \frac{\partial}{\partial \theta^i} = \sum_{t=1}^m \Gamma_{ij}^t(\theta) \frac{\partial}{\partial \theta^t} = 0, \nabla_{\frac{\partial}{\partial \eta_j}} \frac{\partial}{\partial \eta_i} = \sum_{t=1}^m \Gamma_{ti}^{j*}(\eta) \frac{\partial}{\partial \eta_t} = 0. \quad (5.33)$$

Очевидно, що (5.33) є інваріантними відносно лінійного перетворення систем координат. Тому, для дуально плоского простору (M, g, ∇, ∇^*) , маючи ∇ -афінну координатну систему $[\theta^i]$, підібрати ∇^* -афінну координатну систему $[\eta_j]$ так, щоб для скалярного добутку двох векторів $\partial_i = \frac{\partial}{\partial \theta^i}$ та $\partial^j = \frac{\partial}{\partial \eta_j}$ виконувалось [3]:

$$\partial_i, \partial^j = \delta_i^j, \quad (5.34)$$

де δ_i^h є вже знайомим символом Кронекера.

Якщо координатні системи $[\theta^i]$ та $[\eta_j]$, задані на рімановому многовиді задовільняють (5.33), то їх називають взаємно дуальними. Для дуально плоскої структури (M, g, ∇, ∇^*) така пара координатних систем завжди існує. З іншого боку, можна стверджувати, існування для деякого ріманового многовиду (M, g) таких систем $[\theta^i]$ та $[\eta_j]$,

що відносно них зв'язності ∇, ∇^* є відповідно афінними, і $\frac{\nabla + \nabla^*}{2}$

буде зв'язністю Леві – Чівіта – забезпечує виконання того факту, що (M, g, ∇, ∇^*) буде дуально плоскою структурою.

Компоненти ріманової метрики для такого простору визначаються за формулами:

$$\partial_i, \partial_j = g_{ij}, \partial^i, \partial^j = g^{ij}. \quad (5.35)$$

Базисні вектори в обох системах координат пов'язані наступним чином:

$$\partial^j = (\partial^j \theta^i) \partial_i = g^{ji} \partial_i, \partial_i = (\partial_i \eta_j) \partial^j = g_{ij} \partial^j.$$

Отже, очевидним є

$$\frac{\partial \eta_j}{\partial \theta^i} = g_{ij}, \frac{\partial \theta^i}{\partial \eta_j} = g^{ij}. \quad (5.36)$$

Звідси ми бачимо, що $g_{ik} g^{kj} = \delta_i^j$.

Оскільки матриця g_{ij} є симетричною, це означає, що

$$\frac{\partial \eta_j}{\partial \theta^i} - \frac{\partial \eta_i}{\partial \theta^j} = 0,$$

Тобто диференціальна форма $\eta_j d\theta^j$ є замкненою. Оскільки за лемою Пуанкаре принаймні локально існує така функція (скаляр) $\psi(\theta): M \rightarrow \mathbb{R}$, що

$$\eta_i = \frac{\partial \psi}{\partial \theta^i}. \quad (5.37)$$

З (5.36) та (5.37) випливає, що метрика в системі координат $[\theta^i]$ може бути представленою у вигляді:

$$g_{ij}(\theta) = \frac{\partial^2 \psi(\theta)}{\partial \theta^i \partial \theta^j}, \quad (5.38)$$

як це було зроблено у секції 3.2, рівняння (3.8). Нагадаємо, що многовиди з такою метрикою, утворюють *структури Гессе* (див. Теорему 3.3). Оскільки g_{ij} , то відповідна квадратична форма є позитивно визначеною, звідси $\psi(\theta)$ є строго опуклою функцією змінних $\theta^1, \theta^2, \dots, \theta^m$.

Подібним чином ми можемо проаналізувати рівняння,

$$\theta^i = \frac{\partial \varphi}{\partial \eta_i}. \quad (5.39)$$

яке виникає з огляду того, що

$$\frac{\partial \theta^i}{\partial \eta_j} - \frac{\partial \theta^j}{\partial \eta_i} = 0.$$

Отже,

$$g^{ij}(\eta) = \frac{\partial^2 \varphi(\eta)}{\partial \eta_i \partial \eta_j}, \quad (5.40)$$

де $\varphi(\eta)$ є строго опуклою функцією змінних $\eta_1, \eta_2, \dots, \eta_m$.

Якщо виписати зв'язок між $\varphi(\eta)$ та $\psi(\theta)$, отримаємо

$$\varphi(\eta) = \theta^i \eta_i - \psi(\theta). \quad (5.41)$$

Однак, в (5.41) не відзеркалено зв'язок між змінними $\theta^1, \theta^2, \dots, \theta^m$ та $\eta_1, \eta_2, \dots, \eta_m$ що описується у (5.37) та (5.39). Якщо ми хочемо в (5.41) врахувати зв'язок (5.37) та строгу опуклість $\psi(\theta)$, то це потрібно записати так

$$\varphi(\eta) = \max_{\theta} (\theta^i \eta_i - \psi(\theta)). \quad (5.42)$$

Аналогічно, для врахування опуклості $\varphi(\eta)$ та (5.39) – маємо

$$\psi(\theta) = \max_{\eta} (\theta^i \eta_i - \varphi(\eta)). \quad (5.43)$$

Перетворення систем координат $[\theta^i] \leftrightarrow [\eta_j]$, що представлені (5.42) та (5.43) мають назву перетворень Лежандра. Таким чином, функції $\psi(\theta)$ та $\varphi(\eta)$ є потенціалами, які породжують взаємно дуальні структури Гессе.

Також

$$\Gamma_{ij,k}^*(\theta) = \nabla_{\partial_i}^* \partial_j, \partial_k = \partial_i \partial_j \partial_k \psi(\theta), \quad (5.44)$$

$$\Gamma^{ij,k}(\eta) = \nabla_{\partial^i} \partial^j, \partial^k = \partial^i \partial^j \partial^k \varphi(\eta). \quad (5.45)$$

Зважимо, що

$$\Gamma^{*ij,k}(\eta) = \Gamma_{ij,k}(\theta) = 0, \quad (5.46)$$

оскільки згідно з (5.12) та (5.38)

$$\partial_i \partial_j \partial_k \psi(\theta) = \frac{\partial g_{ij}}{\partial x^k} = \Gamma'_{ik} g_{ij} + \Gamma^*_{jk} g_{it} = \Gamma_{ik,j} + \Gamma^*_{jk,i}.$$

$$\partial^i \partial^j \partial^k \varphi(\eta) = \Gamma^{ik}_{,t}(\eta) g^{ij}(\eta) + \Gamma^{*jk}_{,t}(\eta) g^{ti}(\eta) = \Gamma^{ik,j}(\eta) + \Gamma^{*jk,i}(\eta).$$

Усі наведені міркування ми можемо підсумувати у вигляді теореми.

Теорема 5.4 Якщо для дуально плоского многовиду (M, g, ∇, ∇^*) існує ∇ -афінна координатна система $[\theta^i]$, то також існує дуальна до $[\theta^i]$ ∇^* -афінна координатна система $[\eta_j]$. Ці координатні системи пов'язані одна з одною через перетворення Лежандра (5.42), (5.43) потенціалів $\psi(\theta)$ та $\varphi(\eta)$. Для безпосереднього перерахунку координат використовують формули (5.37) та (5.39). Компоненти метричного тензора в цих системах координат обчислюються відповідно через формули (5.38) та (5.40).

Коефіцієнти об'єктів ∇ - та ∇^* -зв'язностей обчислюються за допомогою (5.44)–(5.46).

Зауваження 5.3 Варто нагадати, що коли ми для короткості називали об'єкти $\Gamma_{ik,j}$, $\Gamma^*_{jk,i}$, $\Gamma^{ik,j}$ та $\Gamma^{*jk,i}$ – коефіцієнтами зв'язності (зокрема, щойно посилаючись на (5.44)–(5.46)) – то це було не зовсім вірно. Перш ніж підставити отримані «гамми» у рівняння для коваріантного диференціювання – потрібно метричним тензором змінити локацію останнього індексу. А саме, у координатах $[\theta^i]$ коефіцієнти зв'язності мають вигляд $\Gamma^{ik}_{ij}(\theta) = \Gamma_{ij,t}(\theta) g^{tk}(\theta)$ та $\Gamma^{*k}_{ij}(\theta) = \Gamma^*_{ij,t}(\theta) g^{tk}(\theta)$.

Відповідно, у координатах $[\eta_j]$:

$$\Gamma^{ij}_{,k}(\eta) = \Gamma^{ij,t}(\eta) g_{tk}(\eta) \text{ та } \Gamma^{*ij}_{,k}(\eta) = \Gamma^{*ij,t}(\eta) g_{tk}(\eta).$$

Наприклад, у координатах $[\theta^i]$ рівняння e -геодезичної виглядає так:

$$\frac{d^2 \theta^i}{ds^2} + \Gamma_{jk,t}(\theta) g^{ti}(\theta) \frac{d\theta^j}{ds} \frac{d\theta^k}{ds} = 0. \quad (5.47)$$

У тих саме координатах рівняння m -геодезичної запишемо таким чином:

$$\frac{d^2 \theta^i}{ds^2} + \Gamma^*_{ij,t}(\theta) g^{tk}(\theta) \frac{d\theta^j}{ds} \frac{d\theta^k}{ds} = 0. \quad (5.48)$$

У координатах $[\eta_j]$ рівняння m -геодезичної:

$$\frac{d^2\eta_i}{ds^2} + \Gamma^{*jk,i}(\eta) g_{ii}(\eta) \frac{d\eta_j}{ds} \frac{d\eta_k}{ds} = 0. \quad (5.49)$$

І нарешті, e -геодезична у координатах $[\eta_j]$:

$$\frac{d^2\eta_i}{ds^2} + \Gamma^{jk,i}(\eta) g_{ii}(\eta) \frac{d\eta_j}{ds} \frac{d\eta_k}{ds} = 0. \quad (5.50)$$

Враховуючи (5.44)–(5.46), маємо:

e -геодезична в системі $[\theta^i]$

$$\frac{d^2\theta^i}{ds^2} = 0. \quad (5.47')$$

m -геодезична в системі $[\theta^i]$

$$\frac{d^2\theta^i}{ds^2} + \partial_i \partial_j \partial_i \psi(\theta) g^{ik}(\theta) \frac{d\theta^j}{ds} \frac{d\theta^k}{ds} = 0. \quad (5.48')$$

m -геодезична в системі $[\eta_j]$

$$\frac{d^2i}{ds^2} = 0. \quad (5.49')$$

e -геодезична в системі $[\eta_j]$

$$\frac{d^2\eta_i}{ds^2} + \partial^j \partial^k \partial^l \varphi(\eta) g_{ii}(\eta) \frac{d\eta_j}{ds} \frac{d\eta_k}{ds} = 0. \quad (5.50')$$

Саме простота рівнянь (5.48') та (5.50') демонструє, яким потужним знаряддям у інформаційній геометрії є апарат взаємно дуальних зв'язностей.

5.5. Канонічна дивергенція

У секції 5.2 ми розглядали дивергенції, і побачили що деяка коректно задана дивергенція однозначно породжує метрику і дві дуальних зв'язності, пов'язаних із цією метрикою. З іншого боку, одна і та сама дуальна структура може породжуватись різними дивергенціями. Але існує так звана *канонічна дивергенція*, яка на дуально плоскому просторі задається однозначно.

Нехай заданий дуально плоский многовид (M, g, ∇, ∇^*) на якому існують взаємно дуальні координатні системи $[\theta^i]$ та $[\eta_j]$ і відповідні потенціали $\psi(\theta)$ та $\phi(\eta)$. На цьому многовиді виберемо довільним чином два різних розподіла $p = p(x, \xi)$ та $p' = p(x, \xi')$, таких що $p, p' \in M$. Літерою ξ ми позначаємо параметри розподілів що задані в якійсь довільній координатній системі $\xi = (\xi^1, \xi^2, \dots, \xi^m)$, що була задана на M з самого початку. Звісно, оперуючи з нашими спеціальними координатними системами нам доведеться із довільної переходити до $\theta = (\theta^1, \theta^2, \dots, \theta^m)$ або до $\eta = (\eta_1, \eta_2, \dots, \eta_m)$. Тобто, ми матимемо функції $\theta = \theta(\xi)$, $\eta = \eta(\xi)$. Оскільки розподіли $p = p(x, \xi)$ та $p' = p(x, \xi')$ є різними, то різними будуть і їх представлення у нових координатних системах: для $p = p(x, \xi) \theta = \theta(\xi)$ та $\eta = \eta(\xi)$, а для $p' = p(x, \xi')$ відповідно $\theta' = \theta(\xi')$ та $\eta' = \eta(\xi')$.

Враховуючи викладене, ми можемо ввести на многовиді (M, g, ∇, ∇^*) дивергенцію наступним чином:

$$D(p(x, \xi) | p(x, \xi')) \stackrel{\text{def}}{=} \psi(\theta(\xi)) + \phi(\eta(\xi)) - \theta^i(\xi) \eta_i(\xi). \quad (5.51)$$

З іншого боку, якщо для нас не буде суттєвою якась третя система координат, і будуть задані два різних розподіла – скажімо p та q з їхніми координатами у відповідних системах, а саме $\theta(p)$ та $\theta(q)$ у системі $[\theta^i]$ і $\eta(p)$ та $\eta(q)$ у системі $[\eta_j]$ – тоді дивергенція двох розподілів виглядатиме так:

$$D(p | q) \stackrel{\text{def}}{=} \psi(\theta(p)) + \phi(\eta(q)) - \theta^i(p) \eta_i(q). \quad (5.52)$$

Величину, що обчислюється за формулою (5.51) або (5.52) – називатимемо *канонічною дивергенцією для структури (M, g, ∇, ∇^*)* , або задля короткості *(g, ∇) -дивергенцією*.

Ми бачимо, що згідно (5.42) та (5.43) виконуються властивості $D(p|q) \geq 0$ та $D(p|q) = 0$ якщо і тільки якщо $p = q$. Є також справедливим

$$D\left(\left(\partial_i \partial_j\right)_p | q\right) = g_{ij}(p) \text{ та } D\left(p | \left(\partial_i \partial_j\right)_q\right) = g^{ij}(q). \quad (5.53)$$

З виразу (5.53) ми бачимо, що з міркувань симетрії, можна ввести спряжену (g, ∇^*) -дивергенцією

$$D^*(p|q) = D(q|p). \quad (5.54)$$

Якщо існує підмноговид $S \subset M$, причому він є автопаралельним по відношенню як для зв'язності ∇ так і для ∇^* то на ньому очевидно буде індуковано дуальну плоску структуру $(S, g_S, \nabla_S, \nabla_S^*)$. Можна довести, що на ньому індукована (g_S, ∇_S) -дивергенція, така, що для будь-яких $p, q \in S$

$$D_S(p|q) = D(q|p). \quad (5.55)$$

Аналогічне твердження є справедливим для спряженої дивергенції:

$$D_S(q|p) = D(q|p) = D^*(p|q) = D_S^*(p|q).$$

Якщо ∇ є зв'язністю Леві – Чівіта, узгодженою із метрикою Фішера, то маємо $\nabla = \nabla^*$. Для дуально-плоского простору це просто означає, що в ньому можна задати систему координат $[\theta^i]$, що є евклідовою. Така координатна система буде «автодуальною», тобто $\theta^i = \eta_i$. Очевидно, що потенціал бути мати вигляд:

$$\Psi = \varphi = \frac{1}{2} \sum_{i=1}^m (\theta^i)^2.$$

Підставивши це у (5.54), отримуємо:

$$\begin{aligned} D(p|q) &= \frac{1}{2} \sum_{i=1}^m \left((\theta^i(p))^2 + (\theta^i(q))^2 - 2\theta^i(p)\theta^i(q) \right) = \\ &= \frac{1}{2} \sum_{i=1}^m (\theta^i(p) - \theta^i(q))^2 = \frac{1}{2} (d(p, q))^2. \end{aligned} \quad (5.56)$$

Тут через $d(p, q)$ позначено евклідову відстань

$$d(p, q) = \sqrt{\sum_{i=1}^m (\theta^i(p) - \theta^i(q))^2}.$$

Загалом, канонічна дивергенція $D(p|q)$ на дуально плоскому просторі лише приблизно дорівнює $\frac{1}{2}(d(p, q))^2$ в сенсі формули (5.19), тобто з точністю до доданків $o(\Delta \xi^2)$.

5.6. Найважливіші властивості канонічної дивергенції

Сформулюємо декілька теорем, що відзеркалюють властивості канонічної дивергенції [3].

Теорема 5.5 Нехай на дуально плоскому многовиді (M, g, ∇, ∇^*) задані взаємно дуальні координатні системи $[\theta^i]$ та $[\eta_j]$, і D є дивергенцією на M . Для того, щоб D була канонічною дивергенцією, необхідно і достатньо, щоб для будь-яких розподілів $p, q, r \in M$ виконувалось наступне триангуляційне співвідношення:

$$D(p|q) + D(q|r) - D(p|r) = \sum_{i=1}^m (\theta^i(p) - \theta^i(q))(\eta_i(r) - \eta_i(q)). \quad (5.57)$$

Також, ми можемо узагальнити теорему Піфагора, яку було сформульовано для евклідових відстаней на випадок канонічної дивергенції що є заданою на довільному дуально плоскому просторі (M, g, ∇, ∇^*) .

Теорема 5.6 Нехай на дуально плоскому многовиді (M, g, ∇, ∇^*) задані три точки (розподіли) $p, q, r \in M$. При цьому, ∇ -геодезична γ_1 з'єднує точки p та q , а ∇^* -геодезична γ з'єднує точки q та r так, що в точці q обидві геодезичні є ортогональними згідно із заданою метрикою g . Тоді має місце рівність Піфагора (рис. 5).

$$D(p|r) = D(p|q) + D(q|r). \quad (5.58)$$

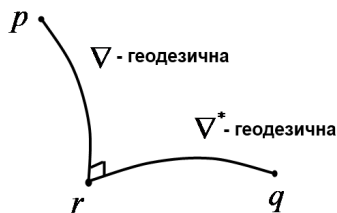


Рисунок 5 – Рівність Піфагора для (g, ∇) – дивергенції

Наступне твердження є наслідком Теорема 5.6.

Наслідок 5.7 Нехай на дуально плоскому многовиді (M, g, ∇, ∇^*) задано точку $p \in M$ і нехай $S \in \nabla^*$ -автопаралельним підмноговидом M , причому точка $q \in S$. Тоді необхідною та достатньою умовою $D(p|q) = \min_{r \in S} D(p|r)$ є ортогональність до S цієї ∇ -геодезичної, що з'єднує точки p та q (рис. 6).

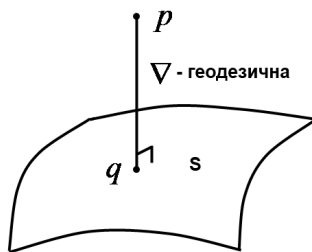


Рисунок 6 – Теорема про проєкцію для (g, ∇) – дивергенції

Точка q про яку йде мова у Наслідку 5.7 має назву ∇ -проєкції точки p на підмноговид S .

Наступне твердження є справедливим для будь-якого підмноговида S .

Теорема 5.8 Нехай на дуально плоскому многовиді (M, g, ∇, ∇^*) задано точку $p \in M$ і нехай $S \in$ підмноговидом M , причому точка $q \in S$. Тоді необхідною та достатньою умовою, щоб q була стаціонарною точкою функції $D(p|\cdot): r \rightarrow D(p|r)$ обмеженої на підмноговид S (в тому сенсі, що усі частинні похідні відносно координат на S мають бути рівні нулю) є ортогональність до S ∇ -геодезичної, що з'єднує точки p та q

Наслідок 5.9 Нехай на дуально плоскому многовиді (M, g, ∇, ∇^*) задано точку $p \in M$ і нехай $S \in$ гіперповерхнею у M , так, що $S = \{q \in M | D(p|q) = c\}$, тобто $S \in$ « D -сферою» радіуса $c = \text{const}$. Тоді кожна ∇ -геодезична, що походить через центр, яким є точка p отогонально перетинає S .

5.7. Взаємно дуальні шарування

Нехай заданий дуально плоский m -вимірний многовид $(M_m, g, \nabla, \nabla^*)$ на якому існують взаємно дуальні координатні системи $[\theta^i]$ та $[\eta_i]$. Розіб'ємо множину значень індексу $i = 1, 2, \dots, m$ на дві частини: $i = 1, 2, \dots, k$ та $i = k + 1, k + 2, \dots, m$. Назвемо ці частини секцією I та секцією II відповідно. Нехай $S(c_I)$ буде множиною точок, чії координати $[\eta_i]$ в секції I є фіксованими наборами констант $c_I = (c_{I,i})$ для $i = 1, 2, \dots, k$

$$S(c_I) = \{p \in M_m | \eta_1(p) = c_{I,1}, \eta_2(p) = c_{I,2}, \dots, \eta_k(p) = c_{I,k}\}.$$

Таким чином утворений $(m - k)$ -вимірний ∇^* – автопаралельний підмноговид. Множина $S(c_I)$ визначається конкретними значеннями набору c_I , і якщо $c_I \neq c_I$, то

$$S(c_I) \cap S(c_I) = \emptyset,$$

і є очевидним, що мом можемо представити весь многовид M_m , як об'єднання усіх можливих

$$\bigcup_{c_I} S(c_I) = M_m.$$

Тобто, M_m є розбитим на підмноговиди. Таке розбиття називають шаруванням M_m .

Подібним чином розглянемо множину $A(d_{II})$ точок, чії координати $\theta^{k+1}, \theta^{k+2}, \dots, \theta^m$ у θ -координатній системі є фіксованими набором $d_{II}^{k+1}, \dots, d_{II}^m$:

$$A(d_{II}) = \{p \in M_m | \theta^{k+1}(p) = d_{II}^{k+1}, \theta^{k+2}(p) = d_{II}^{k+2}, \dots, \theta^m(p) = d_{II}^m\}.$$

Ми отримали ∇ -автопаралельний підмноговид, і $\{A(d_{II})\}$ формує ще одне шарування многовиду M_m .

Будь-яка точка p однозначно визначає певні $S(c_I)$ та $A(d_{II})$, що містить цю точку. Таким чином, точка p є перетином цих $S(c_I)$ та $A(d_{II})$. Дотичним простором $T_p(S)$ до $S(c_I)$ в цій точці є $(m-k)$ -вимірний простір, натягнутий на базисні вектори $\{\partial^{k+1}, \partial^{k+2}, \dots, \partial^m\}$ а дотичним простором $T_p(A)$ до $A(d_{II})$ є відповідно k -вимірний простір, натягнутий на базисні вектори $\{\partial_1, \partial_2, \dots, \partial_k\}$. Дуже важливим фактом є те, що

$$\partial_i, \partial^j = 0, \text{ якщо } i \neq j.$$

Тому $T_p(S)$ та $T_p(A)$ є взаємно ортогональними. Такі шарування, що є ортогональними одне до одного називають *взаємно дуальними*.

Можна розділити координати точки p на перші $(m-k)$ виміри і решту k вимірів: $\eta = (\eta_I, \eta_{II})$. Або подібним чином можна розбити θ -координати: $\theta = (\theta^I, \theta^{II})$. Але, оскільки точка p є перетином $S(\eta_I)$ та $A(\theta^{II})$, то дуже корисним є вибрати координати наступним чином: $\xi = (\eta_I, \theta^{II})$.

Такий вибір має назву *мішаної системи координат* (рис. 7).

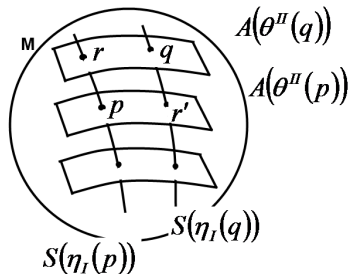


Рисунок 7 – Взаємно дуальні шарування многовиду M

Нехай p та q є точками, чийми мішаними координатами є $(\eta_I(p), \theta^{II}(p))$ та $(\eta_I(q), \theta^{II}(q))$ відповідно. Далі, розглянемо точки з такими мішаними координатами r : $(\eta_I(p), \theta^{II}(q))$ та r' : $(\eta_I(q), \theta^{II}(p))$. Тоді r є ∇ -проекцією q на $S(\eta_I(p))$ та r' є ∇^* -проекцією q на $A(\theta^{II}(p))$.

Звідси ми бачимо, що

$$D(p|q) = D(p|r) + D(r|q) \text{ та}$$

$$D(q|p) = D(q|r) + D(r|p),$$

де символом D позначено (g, ∇^*) -дивергенцією.

6. Висновки

Виконане дослідження доводить, що інформаційна геометрія є актуальною та потужною математичною основою для сучасної кібербезпеки, дозволяючи моделювати ймовірнісні розподіли, аналізувати інформаційні відстані та виявляти аномалії у великих потоках даних. Геометричні методи, які ґрунтуються на статистичних многовидах, метриці Фішера та дивергенції, показують переваги у виявленні складних атак і оптимізації захисту даних, перевищуючи класичні підходи, особливо для аналізу гібридних загроз та в умовах невизначеності. Практична цінність полягає у поєднанні інформаційної геометрії з алгоритмами штучного інтелекту, що дозволяє підвищувати ефективність систем раннього попередження і знижувати обчислювальні витрати. Водночас існують обмеження: висока обчислювальна складність, нестача стандартів впровадження та потреба адаптації до українського законодавства. Перспективи подальших досліджень охоплюють інтеграцію з квантовими технологіями та блокчейном, оптимізацію алгоритмів, емпіричне тестування на реальних даних, застосування до нових сфер (наприклад, соціальних мереж чи біометрії), а також розробку стандартів та освітніх програм у співпраці з українськими установами. Інформаційна геометрія має всі шанси стати основою для більш стійких і адаптивних систем кіберзахисту. Ці перспективи дозволять не лише теоретично вдосконалити методи, але й забезпечити їх практичне впровадження для підвищення рівня кібербезпеки в Україні та світі.

Список використаних джерел

1. National Institute of Standards and Technology. (2023). Cybersecurity Framework. Gaithersburg: NIST.
2. Про основні засади забезпечення кібербезпеки України : Закон України від 05.10.2017 № 2163-VIII : станом на 20 квітня 2025 року. URL: <https://zakon.rada.gov.ua/laws/show/2163-19#Text>
3. Amari S. Methods of information geometry. Providence, RI : American Mathematical Society, 2000. 206 p.
4. Dodson, C. T. J. (2017). On Some Information Geometric Approaches to Cyber Security. DOI: 10.1007/978-3-319-51500-7_9
5. Cherevko Y., Chepurna O., Kuleshova Y. On information geometry methods for data analysis. *Geometry, Integrability and Quantization*. 2024. Issue 29. P. 11–22. DOI: 10.7546/giq-29-2024-11-22
6. Chuang I. L., Nielsen M. A. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*. 1997. Vol. 44, No. 11–12. P. 2455–2467. DOI: <https://doi.org/10.1080/09500349708231894>
7. Wang X., Li Y., Zhang Z. Data Privacy and Security Risks in AI-Driven Organizations. *Journal of Information Privacy and Security*. 2019. Vol. 37, No. 187–200. P. 112–130.
8. Watanabe S. Mathematical Theory of Bayesian Statistics. Boca Raton: CRC Press. 2018.
9. Chen Y., Liu H., Zhang Q. Information geometry for big data security analytics. *IEEE Transactions on Information Forensics and Security*. 2020. Issue 15. P. 2100–2115.
10. Семенов П. В. Нейронні мережі та їх застосування. Київ : Техніка, 2023. 230 с.
11. Semenova O. O., Semenov A. O., Voitsekhovska O. O. Using a neural network for the vertical handover procedure. *Information technology and computer engineering*. 2020. Vol. 49, No. 3. P. 14–21. DOI: <https://doi.org/10.31649/1999-9941-2020-49-3-14-21> (date of access: 28.07.2025).
12. Mikeš J. et al. Differential Geometry of Special Mappings. Palacky Univ. Press. Olomouc. 2019.
13. Eisenhart L. Riemannian Geometry. Princeton Univ. Press. 1926.

14. Amari S. Differential Geometry of Curved Exponential Families. Curvature and Information Loss. *Ann. Stat.* 1982. Issue 10. P. 357–385.

15. Cartan, Elie Leçons sur la géométrie des espaces de Riemann (2ème. ed. rev. et aug. ed.). Paris: Gauthier-Villars. p. VIII, 378. ISBN 287647008X. Zbl 0060.38101.

16. Shima H. The Geometry of Hessian Structures. World Sci. Publ. 2007.

17. Matumoto T. Any statistical manifold has a cotrast function -on the C3-functions taking the minimum at the diagonal of the product manifold. *Hiroshima Mathematical Journal.* 1993. Issue 23. P. 327–332.

18. Csiszar I. Information type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica.* 1967. Issue 2. P. 299–318.

19. Kennedy, T. G. (Spring 2006). Some notes on sufficient statistics. Math 466/566 lecture notes, University of Arizona.

20. Lauritzen, S. L. Sufficiency and Unbiased Estimation. (2004, 19 October). Lecture 3 of BS2 Statistical Inference. University of Oxford. URL: <https://www.stats.ox.ac.uk/~steffen/teaching/bs2siMT04/si3c.pdf>

21. Lauritzen, S. L. Extremal Families and Systems of Sufficient Statistics Lecture Notes in Statistics. Vol. 49. (1988). Springer.