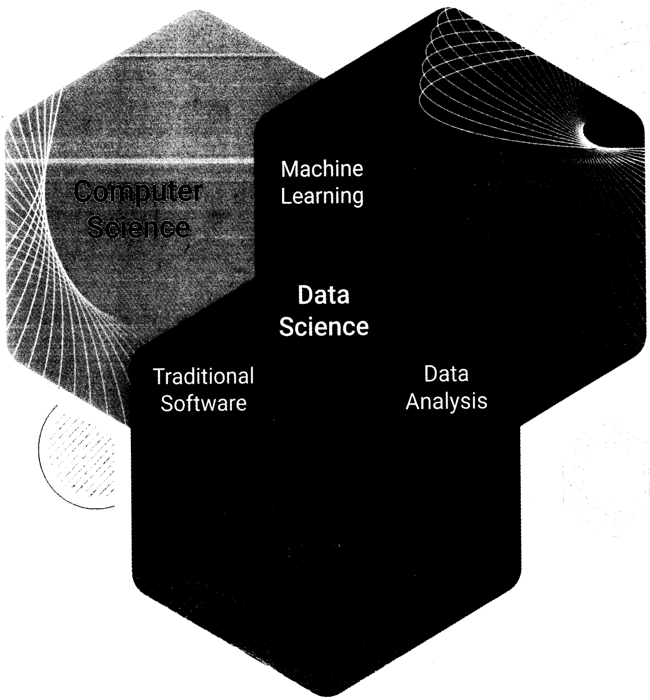
ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN



CƠ SỞ TOÁN CHO KHOA HỌC DỮ LIỆU

NGUYỄN THANH BÌNH (Chủ biên)
ĐINH NGỌC THANH - NGUYỄN ĐÌNH THÚC - NGUYỄN ĐĂNG MINH

NHÀ XUẤT BẢN
ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

CƠ SỞ TOÁN CHO KHOA HỌC DỮ LIỆU

NGUYỄN THANH BÌNH (Chủ biên)
ĐINH NGỌC THANH - NGUYỄN ĐÌNH THÚC - NGUYỄN ĐĂNG MINH

NHÀ XUẤT BẢN
ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH

LỜI GIỚI THIỆU

Trong mười năm gần đây, Việt Nam bước vào giai đoạn *chuyển đổi số* với các lĩnh vực Dữ liệu lớn (Big data), Trí tuệ nhân tạo (AI), Internet vạn vật (IoT), Điện toán đám mây (Cloud computing). Các lĩnh vực này đều thu nhận rất nhiều dữ liệu cần phải xử lý và là đối tượng của một ngành khoa học đặc biệt có tên là *Khoa học Dữ liệu* (KHDL).

Do nhận thức được nhu cầu của giai đoạn mới, nhiều cơ sở đào tạo đã mở ngành đào tạo về KHDL. Trong Đại học Quốc gia TP. HCM có thể kể đến các chương trình KHDL của Đại học Khoa học Tự nhiên, Đại học Công nghệ Thông tin, Đại học Quốc tế. Ngoài ĐHQG TP. HCM có thể kể đến các chương trình của ĐHKHTN Hà Nội, ĐHBK Hà Nội, ĐH Quy Nhơn,... Các chương trình này đều tuyển được nhiều sinh viên, học viên theo học. Để học được ngành KHDL, người học cần phải có một số lượng kiến thức toán tối thiểu (KTTTT). Các kiến thức này khá đa dạng, bao gồm Đại số tuyến tính (ma trận, giá trị riêng, vectơ riêng, phân tích giá trị kỳ dị,...), Phép tính vi tích phân (phương pháp bình phương nhỏ nhất, phương trình vi phân,...) và Xác suất Thống kê (thống kê nhiều chiều, thống kê phi tham số,...).

Trong thực tế, người quan tâm đến KHDL được đào tạo cơ bản về Toán rất đa dạng. Do đó không có sự đồng đều trong KTTTT. Người được đào tạo trong lĩnh vực Khoa học tự nhiên và kỹ thuật được trang bị các kiến thức về toán khá nhiều. Tuy nhiên một số kiến thức như phân tích giá trị kỳ dị, thống kê nhiều chiều,... không có trong chương trình Toán tự nhiên và kỹ thuật. Người được đào tạo trong lĩnh vực Kinh tế xã hội (Kinh tế, Tài chính, Khoa học xã hội...) thì lượng KTTTT cần thiết để học chương trình KHDL bị thiếu hụt khá nhiều.

Để giúp người học có thể chuẩn hóa KTTTT cần thiết cho ngành Khoa học Dữ liệu, nhóm tác giả của Khoa Toán-Tin học và Khoa Công nghệ Thông tin Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia TP. HCM đã soạn ra một số chuyên đề trong ba lĩnh vực Đại số Tuyến tính, Vi tích phân, Xác suất Thống kê. Tài liệu *Cơ sở Toán cho Khoa học Dữ liệu* giúp người học chuẩn hóa kiến thức về Toán để có thể học sâu hơn về KHDL.

LỜI TỰA

Trong hai mươi năm trở lại đây, việc thu nhận một số lượng khổng lồ các dữ liệu làm xã hội nhận ra được nhu cầu phân tích dữ liệu trong tất cả các lĩnh vực khoa học kỹ thuật, kinh tế xã hội của đời sống. Xử lý các dữ liệu là đối tượng của ngành khoa học đã có từ lâu đời: Khoa học Dữ liệu (KHDL). Đây là ngành khoa học nghiên cứu về việc tổ chức, phân tích dữ liệu và hỗ trợ cho việc rút ra các quy luật tự nhiên và xã hội từ dữ liệu để đưa ra các quyết định dẫn dắt hành động cho con người hay máy móc.

Trong giai đoạn trước Thế chiến thứ hai, Khoa học dữ liệu bao gồm hai trụ cột chính là (1) Toán học và (2) các lĩnh vực khoa học chuyên ngành (KHCN - Lý, Hóa, Sinh, Thiên văn, Kinh tế xã hội,...). Các KHCN cung cấp các dữ liệu, các mô hình toán học và giải thích các kết quả tính toán của Toán học. Toán học dùng các thuật toán để tính toán ra các kết quả mà KHCN cần tìm. Một ví dụ minh họa cho ứng dụng của KHDL là việc tính toán quỹ đạo của tiểu hành tinh Ceres từ dữ liệu hạn chế do nhà toán học vĩ đại người Đức Carl Friedrich Gauss thực hiện. Vào ngày 1 tháng 1 năm 1801, nhà thiên văn học người Sicily và linh mục Công giáo Giuseppe Piazzi (1746–1826) nhận thấy một thiên thể mờ mà ông tin rằng có thể là một "hành tinh mất tích". Ông đặt tên cho vật thể là Ceres và thực hiện được một số lần quan sát định vị Ceres (dữ liệu của KHCN) trước khi bị mất dấu khi Ceres đến gần Mặt trời. Gauss, khi đó mới 24 tuổi, đã thực hiện các tính toán quỹ đạo của

Ceres từ dữ liệu đo đạc của Piazzi bằng cách đưa ra một kỹ thuật gọi là "phương pháp bình phương nhỏ nhất" (thuật toán của Toán học). Kỹ thuật này đưa đến việc giải hệ phương trình tuyến tính bằng phương pháp mà ngày nay chúng ta gọi là "phép khử Gauss". Kết quả công bố của Gauss đã gây chấn động trong giới Thiên văn học (KHCN) khi Ceres xuất hiện trở lại một năm sau đó trong chòm sao Xử Nữ ở vị trí gần như chính xác với nơi ông đã dự đoán! (sự giải thích kết quả bằng KHCN). Ý tưởng cơ bản của phương pháp này đã được kỹ sư người Đức Wilhelm Jordan phổ biến rộng rãi hơn trong cuốn sách của ông về trắc địa (khoa học đo hình dạng Trái đất) mang tên *Handbuch der Vermessungskunde* và được xuất bản vào năm 1888.

Để thực hiện các tính toán, Gauss đã tính bằng tay và sử dụng khoảng 100 giờ tính toán. Đây là một trở ngại lớn cho việc áp dụng các thuật toán Toán học vào các khoa học chuyên ngành. Trong Thế chiến thứ hai, nhóm các nhà mật mã học tại Bletchley Park do nhà toán học Alan Turing lãnh đạo đối diện với bài toán phá mã máy Enigma của Đức Quốc Xã, thiết bị mã hóa cải tiến so với phiên bản dân sự do kỹ sư Đức Arthur Scherbius phát minh vào giai đoạn cuối Thế chiến thứ nhất. Máy này vận hành dựa trên các "khóa mã" lựa chọn trong 158,9 triệu, triệu, triệu khả năng và được thay đổi mỗi ngày. Điều này có nghĩa là Turing không thể có 100 giờ như Gauss để tìm khóa mã. Thay vì tính toán bằng tay, Turing và các đồng nghiệp của ông đã sáng chế ra một cỗ máy, mà nay mô hình toán học tổng quát cho nó được gọi là "máy Turing", tiền thân của các máy tính hiện đại. Sau Thế chiến thứ hai, lĩnh vực khoa học máy tính dần dần phát triển và trở thành trợ thủ không thể thiếu cho việc xử lý dữ liệu.

Ngày nay, Khoa học dữ liệu là một khoa học liên ngành dựa vào ba trụ cột chính: (1) Tri thức của Khoa học Chuyên ngành, (2) Toán học (TH), và (3) Công nghệ thông tin (CNTT), trong đó Thống kê Toán học (TKTH) đóng một vai trò quan trọng, và kết hợp với CNTT, ngoài việc giải quyết các bài toán với cỡ mẫu nhỏ, TKTH và CNTT phát triển thêm các công cụ, giải thuật cho dữ liệu lớn. Phần giao thoa này được biết dưới tên là Máy Học (Machine Learning).

Phần TH cho KHDL là phần rất quan trọng và đòi hỏi các hiểu biết vượt quá chương trình Toán học cơ bản dành cho chương trình đại học của các khoa học chuyên ngành đang giảng dạy ở Việt Nam hiện nay. Một số kiến thức như Phân tích giá trị kỳ dị (singular value decomposition - SVD), thống kê nhiều chiều,... không có trong chương trình Toán tự nhiên và kỹ thuật. Người được đào tạo trong lĩnh vực Kinh tế xã hội (Kinh tế, Tài chính, Khoa học Xã hội...) thì lượng kiến thức về Toán còn ít hơn.

Nhằm cung cấp kiến thức cần thiết cho người học chuyên ngành KHDL, chúng tôi đưa ra mô hình gồm hai tài liệu song hành *Cơ sở Toán cho Khoa học Dữ liệu*, và *Cơ sở Tin cho Khoa học Dữ liệu*, trong đó, các kiến thức Toán cơ bản cho KHDL được giới thiệu nguồn gốc và cơ sở lý thuyết, thành lập các giải thuật tính toán, kiểm nghiệm cụ thể trên các mô hình nhỏ trong tài liệu thứ nhất, và được lập trình tính toán cho mô hình tổng quát trong tài liệu thứ hai.

Nội dung giáo trình này gồm 5 chương, trong đó chương đầu trình bày một số kiến thức cần thiết của Đại số Tuyến tính nhằm cung cấp các công cụ giải số cho các giải thuật trình bày trong các chương sau. Chương 2 giới thiệu một lĩnh vực quan trọng trong Lý thuyết Tối ưu, Tối ưu Lồi, cho bài toán xác định các điểm cực trị toàn cục. Ba chương còn lại giới thiệu ba mô hình khác nhau trong việc tìm hiểu mối liên quan giữa các biến trong bộ dữ liệu. Chương 3 khảo sát sự liên hệ giữa một biến phụ thuộc theo các biến độc lập, với công cụ chính là phương pháp bình phương nhỏ nhất của Gauss. Chương 4 giới thiệu một số phương pháp tính số cho mô hình động (dynamic models), mô hình quan tâm đến tốc độ thay đổi của các biến, biểu diễn bằng phương trình vi phân thường (ODE). Chương 5 giới thiệu một phương pháp của Xác suất Thống kê, khảo sát dữ liệu của một biến ngẫu nhiên bằng phương pháp hợp lý cực đại. Cuối cùng, một số kiến thức cơ bản mà chúng tôi giả định rằng người học đã biết nhưng có thể không nhớ, sẽ được trình bày lại một cách vắn tắt trong phần Phụ lục.

Đối chiếu với tiêu chí nằm lòng của giới thống kê,

"All models are wrong, but some are useful."

của Giáo sư George Box, người được mệnh danh là "one of the great statistical minds of the 20^{th} century", chúng tôi hy vọng rằng "mô hình" chúng tôi đưa ra nhằm cung cấp kiến thức cần thiết cho KHDL có ít nhiều "hữu ích".

Tài liệu được xây dựng không từ một khuôn mẫu có sẵn nên chắc chắn còn rất nhiều thiếu sót. Chúng tôi rất mong nhận được phản hồi từ người đọc để "mô hình sau" sẽ hữu ích hơn "mô hình trước" theo đúng cách mà khoa học phát triển. Mọi chi tiết xin liên hệ tới chủ biên (ngtbinh@hcmus.edu.vn) hoặc nhóm tác giả qua email datascience@hcmus.edu.vn.

Nhóm tác giả

TỪ KHÓA

Chương 1

Ma trận chính tắc (standard matrix). **Nhân** (kernel). **Ảnh** (image). **Không gian cột** (column space). **Không gian dòng** (row space). **Hạng** (rank). **Toán tử tuyến tính** (linear operator). **Đồng dạng** (similar). **Chéo hóa được** (diagonalizable). **Chéo hóa** (diagonalize). **Vectơ riêng** (eigenvector). **Trị riêng** (eigenvalue). **Không gian riêng** (eigenspace). **Nghiệm không tầm thường** (nontrivial solution). **Đa thức đặc trưng** (characteristic polynomial). **Phương trình đặc trưng** (characteristic equation). **Tích vô hướng** (dot product/scalar product). **Trực giao** (orthogonal). **Vectơ đơn vị** (unit vector). **Trực chuẩn** (orthonormal). **Đồng dạng trực giao** (orthogonally similar). **Chéo hóa trực giao được** (orthogonally diagonalizable). **Chéo hóa trực giao** (orthogonally diagonalizes). **Phân tích giá trị kỳ dị** (Singular Value Decomposition – SVD). **Giá trị kỳ dị** (singular values). **Vectơ kỳ dị trái** (left singular vectors). **Vectơ kỳ dị phải** (right singular vectors).

Chương 2

Trơn (smooth). **Cực đại toàn cục** (global maximum). **Cực tiểu toàn cục** (global minimum). **Lồi** (convex). **Lõm** (concave). **Dây cung** (chord). **Nửa xác định dương** (positive semidefinite). **Nửa xác định âm** (negative semidefinite). **Dạng tuyến tính** (linear form). **Dạng toàn phương** (quadratic form).

Chương 3

Phân tích hồi quy (regression analysis). **Hồi quy đơn** (simple regression). **Hồi quy bội** (multiple regression). **Biến**

độc lập (independent variables). **Biến phụ thuộc** (response variable). **Dữ liệu** (data). **Mô hình** (model). **Dự báo** (prediction). **Hàm dự báo** (prediction function). **Tuyến tính theo tham số** (linear in the parameters model). **Hàm số cơ sở** (basis functions). **Tham số mô hình** (model parameters). **Phù hợp** (consistent). **Phương pháp bình phương nhỏ nhất** (Least squares). **Sai số dự báo** (prediction error). **Phần dư** (residual). **Tổng bình phương các phần dư** (Residual Sum of Squares). **Thừa số tương tác** (interaction term).

Chương 4

Tỷ lệ thay đổi (rate of changes). **Mô hình động** (dynamic model). **Phương trình vi phân** (ordinary differential equation – ODE). **Mô hình dân số** (population models). **Bài toán giá trị đầu** (initial value problem). **Luật Malthus** (Malthusian law). **Luật mũ** (exponential law). **Đại lượng tương tác** (interaction term). **Mô hình logistic** (logistic model). **Hàm logistic** (logistic function). **Bước** (step). **Kích thước bước** (step size).

Chương 5

Hàm phân phối tích lũy (cumulative distribution function – cdf). **Rời rạc** (discrete). **Biến số ngẫu nhiên rời rạc** (discrete random variable). **Hàm mật độ xác suất** (probability density function – pdf, hay probability mass function – pmf). **Độc lập và có cùng phân phối** (independent, identical distribution – iid). **Mẫu ngẫu nhiên** (random sample). **Hàm mật độ xác suất đồng thời** (joint-pdf). **Hàm hợp lý** (likelihood function). **Ước lượng hợp lý cực đại** (maximum likelihood estimator). **Hàm log-hợp lý** (log-likelihood function). **Biến số ngẫu nhiên liên tục** (continuous random variable). **Biến số ngẫu nhiên hỗn hợp** (mixed random variable).

MỤC LỤC

Chương 1

ĐẠI SỐ TUYẾN TÍNH

1.1. MA TRẬN VÀ ÁNH XẠ TUYẾN TÍNH

1.1.1. Định nghĩa. Xét ánh xạ $T: \mathbb{R}^m \to \mathbb{R}^n$. Ta nói T là một **ánh xạ tuyến tính** khi

(i) $T(x+y) = T(x) + T(y)$, và

(ii) $T(kx) = kT(x)$,

với mọi $x, y \in \mathbb{R}^m$, và $k \in \mathbb{R}$. Tập hợp tất cả các ánh xạ tuyến tính từ $\mathbb{R}^m$ vào $\mathbb{R}^n$ được ký hiệu là $L(\mathbb{R}^m, \mathbb{R}^n)$.

Với $x = (x_1, x_2, \cdots, x_m) \in \mathbb{R}^m$ biểu diễn dưới cơ sở chính tắc $S_m = \{e_1, e_2, \cdots, e_m\}$ trong $\mathbb{R}^m$ bằng ma trận cột

$$x = [x]_{S_m} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \in \mathbb{R}^{m \times 1},$$

nghĩa là $x = x_1e_1 + x_2e_2 + \cdots + x_me_m$, ta có

$$T(x) = T(x_1e_1 + x_2e_2 + x_me_m) = x_1T(e_1) + x_2T(e_2) + \cdots + x_mT(e_m). \quad (1)$$

Điều này cho thấy ánh xạ tuyến tính T được hoàn toàn xác định bằng các vectơ $T(e_1), T(e_2), \cdots, T(e_m) \in \mathbb{R}^n$. Hơn nữa, với $T(e_j) = (a_{1j}, a_{2j}, \cdots, a_{nj})$, $j = 1,2,\cdots,m$, biểu diễn dưới cơ sở chính tắc S_n trong $\mathbb{R}^n$ bằng ma trận cột

$$T(e_j) \equiv [T(e_j)]_{S_n} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix} \in \mathbb{R}^{n\times 1},$$

xét ma trận $A = [a_{ij}] \in \mathbb{R}^{n\times m}$, nhận $T(e_j)$, $j = 1,2,\cdots,m$, làm các vectơ cột,

$$A = [T(e_1) \quad T(e_2) \quad \cdots \quad T(e_m)]. \tag{2}$$

Đẳng thức (1) được viết lại thành

$$T(x) = Ax.$$

Chú ý rằng do tính chất của phép nhân ma trận, ứng với mỗi ma trận $A \in \mathbb{R}^{n\times m}$, ánh xạ $T: \mathbb{R}^m \to \mathbb{R}^n$, xác định bởi $T(x) = Ax$, $x \in \mathbb{R}^m$, là một ánh xạ tuyến tính.

1.1.2. Định nghĩa. Ứng với mỗi $T \in L(\mathbb{R}^m, \mathbb{R}^n)$, ma trận $A \in \mathbb{R}^{n\times m}$, xác định bởi (2), được gọi là **ma trận chính tắc** (standard matrix) của T. Ký hiệu $T \equiv T_A \in L(\mathbb{R}^m, \mathbb{R}^n)$.

Nhận xét 1. Ánh xạ $A \in \mathbb{R}^{n\times m} \mapsto T \equiv T_A \in L(\mathbb{R}^m, \mathbb{R}^n)$, xác định bởi

$$T(x) = Ax, \text{ với mọi } x \in \mathbb{R}^m,$$

là một song ánh tuyến tính*. Ngoài ra, với $S \equiv S_A \in L(\mathbb{R}^m, \mathbb{R}^n)$ và $T \equiv T_B \in L(\mathbb{R}^n, \mathbb{R}^p)$, ta có ánh xạ hợp $T \circ S \in L(\mathbb{R}^m, \mathbb{R}^p)$, và vì

$$(T \circ S)(x) = T[S(x)] = T(Ax) = B(Ax) = (BA)x,$$

với mọi $x \in \mathbb{R}^m$, nên ta đồng nhất ánh xạ hợp $T_A \circ S_B$ với ma trận tích $BA \in \mathbb{R}^{p\times m}$.

1.1.3. Định lý. Xét $T \equiv T_A \in L(\mathbb{R}^m, \mathbb{R}^n)$, với

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \in \mathbb{R}^{n\times m}.$$

(i) **Nhân** (kernel) của T, $Ker(T) \equiv \{x \in \mathbb{R}^m | T(x) = 0\}$, chính là **không gian nghiệm** của hệ phương trình tuyến tính thuần nhất $Ax = 0$, ký hiệu $null(A)$.

(ii) **Ảnh** (image) của T, $Im(T) \equiv \{y \in \mathbb{R}^n | \exists x \in \mathbb{R}^m, T(x) = y\}$, chính là không gian con của $\mathbb{R}^n$ sinh bởi các vectơ cột của A,

$$c_j = \left(a_{1j}, a_{2j}, \cdots, a_{nj}\right) \in \mathbb{R}^n, j = 1,2, \cdots, m,$$

mà ta gọi là **không gian cột** (column space) của A, ký hiệu $col(A)$.

---

* Song ánh tuyến tính giữa hai không gian vectơ V, W còn được gọi là một **đẳng cấu** (isomorphism). Gốc Hy Lạp của từ isomorphism là "iso" (tiếng Anh là identical) và "morphe" (tiếng Anh là form). Khi tồn tại một đẳng cấu giữa hai không gian vectơ V, W, ta có thể đồng nhất chúng với nhau.

Xét không gian vectơ con của $\mathbb{R}^m$ sinh bởi các vectơ dòng của A,

$$r_i = (a_{i1}, a_{i2}, \cdots, a_{im}) \in \mathbb{R}^m, i = 1,2,\cdots,n,$$

mà ta còn gọi là **không gian dòng** (row space) của A, ký hiệu $row(A)$. Ta có

1.1.4. Định lý. Với mọi $A \in \mathbb{R}^{n\times m}$,

$$dim(row(A)) = dim(col(A)).$$

1.1.5. Định nghĩa. Cho $A \in \mathbb{R}^{n\times m}$. Số chiều chung của không gian dòng và không gian cột của A được gọi là **hạng** (rank) của ma trận A, ký hiệu $rank(A)$,

$$rank(A) = dim(row(A)) = dim(col(A)).$$

Ký hiệu $nullity(A)$ cho số chiều của không gian nghiệm của A, $null(A)$. Ta có

1.1.6. Định lý. Với mọi $A \in \mathbb{R}^{n\times m}$, ta có

$$rank(A) + nullity(A) = m,$$

số cột của ma trận A.

Ví dụ 1. Xét ánh xạ tuyến tính $T = T_A \in L(\mathbb{R}^6, \mathbb{R}^4)$ xác định bởi

$$A = \begin{bmatrix} -1 & 2 & 0 & 4 & 5 & -3 \\ 3 & -7 & 2 & 0 & 1 & 4 \\ 2 & -5 & 2 & 4 & 6 & 1 \\ 4 & -9 & 2 & -4 & -4 & 7 \end{bmatrix} \in \mathbb{R}^{4\times 6}.$$

Bằng phép khử Gauss-Jordan, ta được

$$A \to A' = \begin{bmatrix} 1 & 0 & -4 & -28 & -37 & 13 \\ 0 & 1 & -2 & -12 & -16 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3}$$

Giải hệ phương trình tuyến tính thuần nhất $Ax = 0$ với $x = (x_1, x_2, x_3, x_4, x_5, x_6) \in \mathbb{R}^6$, ta được

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = r\begin{bmatrix} 4 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} 28 \\ 12 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 37 \\ 16 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + u\begin{bmatrix} -13 \\ -5 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

với $r, s, t, u \in \mathbb{R}$ bất kỳ. Do đó

$$Ker(T) = null(A)$$
$$= span\{(4,2,1,0,0,0), (28,12,0,1,0,0), (37,16,0,0,1,0), (-13,-5,0,0,0,1)\}$$

Hơn nữa, vì hệ các vectơ
$\{(4,2,1,0,0,0), (28,12,0,1,0,0), (37,16,0,0,1,0), (-13,-5,0,0,0,1)\}$

là độc lập tuyến tính nên nó tạo thành một cơ sở cho $Ker(T) = null(A)$ và do đó

$$nullity(A) = 4. \tag{4}$$

Để xác định $col(A) = span\{c_1, c_2, c_3, c_4, c_5, c_6\}$, với

$c_1 = (-1,3,2,4), c_2 = (2,-7,-5,-9), c_3 = (0,2,2,2),$
$c_4 = (4,0,4,-4), c_5 = (5,1,6,-4), c_6 = (-3,4,1,7),$

dùng phép khử Gauss-Jordan,

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \end{bmatrix} = A^T = \begin{bmatrix} -1 & 3 & 2 & 4 \\ 2 & -7 & -5 & -9 \\ 0 & 2 & 2 & 2 \\ 4 & 0 & 4 & -4 \\ 5 & 1 & 6 & -4 \\ -3 & 4 & 1 & 7 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

Ta được $Im(T) = col(A) = span\{(1,0,1,-1), (0,1,1,1)\}$ và vì hệ $\{(1,0,1,-1), (0,1,1,1)\}$ độc lập tuyến tính nên nó tạo thành một cơ sở cho $Im(T) = col(A)$, và do đó

$$dim\big(col(A)\big) = 2. \tag{5}$$

Với không gian dòng $row(A)$, (3) cho $row(A) = span\{(1,0,-4,-28,-37,13), (0,1,-2,-12,-16,5)\}$

và do hệ $\{(1,0,-4,-28,-37,13), (0,1,-2,-12,-16,5)\}$ độc lập tuyến tính nên nó tạo thành một cơ sở của $row(A)$,

$$dim\big(row(A)\big) = 2. \tag{6}$$

Từ (5) và (6) (xem Định lý 1.1.4 và Định nghĩa 1.1.5), ta suy ra

$$rank(A) = dim\big(col(A)\big) = dim\big(row(A)\big) = 2.$$

Từ (4) và (7), ta có $rank(A) + nullity(A) = 6 =$ số cột của ma trận A (xem Định lý 1.1.6).

1.2. CHÉO HÓA MA TRẬN

Xét ánh xạ tuyến tính $T = T_A \in L(\mathbb{R}^n, \mathbb{R}^n)$, với $A \in \mathbb{R}^{n\times n}$, mà ta còn gọi là **toán tử tuyến tính** (linear operator) trên $\mathbb{R}^n$.

Với $S \equiv S_n = \{e_1, e_2, \cdots, e_n\}$ là cơ sở chính tắc của $\mathbb{R}^n$, ta đã có

$$A = [T(e_1) \quad T(e_2) \quad \cdots \quad T(e_n)],$$

trong đó $T(e_i) \equiv [T(e_i)]_S$ là ma trận cột chỉ tọa độ của $T(e_i)$ đối với cơ sở chính tắc S.

Tổng quát, với một cơ sở $B = \{v_1, v_2, \cdots, v_n\}$ bất kỳ của $\mathbb{R}^n$, xét ma trận

$$B = [[T(v_1)]_B \quad [T(v_2)]_B \quad \cdots \quad [T(v_n)]_B] \equiv [b_{ij}],$$

trong đó $[T(v_i)]_B$ chỉ tọa độ của $T(v_i)$ đối với cơ sở B, $i = 1,2,\cdots,n$. Với $x \in \mathbb{R}^n$,

$$[x]_B = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n\times 1},$$

ta có $x = x_1v_1 + x_2v_2 + \cdots + x_nv_n$ và do đó $T(x) = x_1T(v_1) + x_2T(v_2) + \cdots + x_nT(v_n)$. Suy ra

$$\begin{aligned}
[T(x)]_B &= [x_1T(v_1) + x_2T(v_2) + \cdots + x_nT(v_n)]_B \\
&= x_1[T(v_1)]_B + x_2[T(v_2)]_B + \cdots + x_n[T(v_n)]_B \\
&= [[T(v_1)]_B \quad [T(v_2)]_B \quad \cdots \quad [T(v_n)]_B] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= B[x]_B
\end{aligned}$$

1.2.1. Định nghĩa. Xét toán tử tuyến tính $T = T_A \in L(\mathbb{R}^n, \mathbb{R}^n)$. Ứng với mỗi cơ sở $B = \{v_1, v_2, \cdots, v_n\}$ của $\mathbb{R}^n$, ma trận

$$[[T(v_1)]_B \quad [T(v_2)]_B \quad \cdots \quad [T(v_n)]_B] \in \mathbb{R}^{n\times n}$$

được gọi là ma trận của T đối với cơ sở B, ký hiệu $[T]_B$, và khi đó, do (8),

$$[T(x)]_B = [T]_B[x]_B, \text{ với mọi } x \in \mathbb{R}^n. \tag{9}$$

Đặc biệt, A là ma trận của T đối với cơ sở chính tắc S của $\mathbb{R}^n$, $A = [T]_S$,

$$[T(x)]_S = [T]_S[x]_S, \text{ với mọi } x \in \mathbb{R}^n.$$

Ta có

1.2.2. Định lý. Xét toán tử tuyến tính $T = T_A \in L(\mathbb{R}^n, \mathbb{R}^n)$. Với mọi cơ sở $B = \{v_1, v_2, \cdots, v_n\}$ của $\mathbb{R}^n$, ta có

$$[T]_B = P^{-1}AP,$$

với $P = [v_1 \quad v_2 \quad \cdots \quad v_n] \equiv [[v_1]_S \quad [v_2]_S \quad \cdots \quad [v_n]_S]$.

Chứng minh. Chú ý rằng do $rank(P) = n$ nên P là ma trận khả nghịch. Với mọi $x \in \mathbb{R}^n$,

$$[x]_B = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^{n\times 1},$$

ta có $x = x_1v_1 + x_2v_2 + \cdots + x_nv_n$ và do đó

$$[x]_S = x_1[v_1]_S + x_2[v_2]_S + \cdots + x_n[v_n]_S = P[x]_B.$$

Kết hợp với (9), ta được

$$\begin{aligned} T(x) &\equiv [T(x)]_S = P[T(x)]_B = P([T]_B[x]_B) \\ &= (P[T]_B)[x]_B = (P[T]_B)(P^{-1}[x]_S) = (P[T]_BP^{-1})[x]_S \\ &\equiv (P[T]_BP^{-1})x. \end{aligned}$$

Do $T \equiv T_A$, ta suy ra (xem nhận xét 1) $A = P[T]_BP^{-1}$ và do đó $[T]_B = P^{-1}AP$.

Ví dụ 2. Cho toán tử tuyến tính $T: \mathbb{R}^3 \to \mathbb{R}^3$ xác định bởi

$$T(x_1, x_2, x_3) = (4x_1 + 2x_2 - x_3, -6x_1 - 4x_2 + 3x_3, -6x_1 - 6x_2 + 5x_3).$$

Xét cơ sở chính tắc $S_3 = \{e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1)\}$, ta được $T \equiv T_A$, với

$$A = [T(e_1) \quad T(e_2) \quad T(e_3)] = \begin{bmatrix} 4 & 2 & -1 \\ -6 & -4 & 3 \\ -6 & -6 & 5 \end{bmatrix}.$$

Với cơ sở $B = \{v_1 = (1,1,0), v_2 = (1,0,1), v_3 = (0,1,1)\}$, ta tìm $[T]_B$ bằng Định nghĩa 1.2.1 và so sánh với Định lý 1.2.2.

+ Dùng Định nghĩa 1.2.1, ta tính $T(v_1) = (6, -10, -12)$, $T(v_2) = (3, -3, -1)$, $T(v_3) = (1, -1, -1)$, và tìm $[T(v_i)]_B$, $i = 1,2,3$, bằng cách giải các hệ phương trình tuyến tính với ma trận các hệ số $P = [v_1 \quad v_2 \quad v_3]$ và vectơ cột các hệ số vế phải $T(v_i)$, $i = 1,2,3$. Dùng phép khử Gauss-Jordan (cho cả ba hệ phương trình tuyến tính này), ta được

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 6 & 3 & 1 \\ 1 & 0 & 1 & -10 & -3 & -1 \\ 0 & 1 & 1 & -12 & -1 & -1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 & 0.5 & 0.5 \\ 0 & 1 & 0 & 2 & 2.5 & 0.5 \\ 0 & 0 & 1 & -14 & -3.5 & -1.5 \end{array}\right]$$

và do đó

$$[T]_B = [[T(v_1)]_B \quad [T(v_2)]_B \quad [T(v_3)]_B] = \begin{bmatrix} 4 & 0.5 & 0.5 \\ 2 & 2.5 & 0.5 \\ -14 & -3.5 & -1.5 \end{bmatrix}.$$

+ Dùng Định lý 1.2.2,

$$[T]_B = P^{-1}AP = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 4 & 2 & -1 \\ -6 & -4 & 3 \\ -6 & -6 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 4 & 0.5 & 0.5 \\ 2 & 2.5 & 0.5 \\ -14 & -3.5 & -1.5 \end{bmatrix}.$$

1.2.3. Định nghĩa. Hai ma trận $A, B \in \mathbb{R}^{n\times n}$ được gọi là **đồng dạng** (similar), ký hiệu $A \sim B$, khi tồn tại ma trận khả nghịch $P \in \mathbb{R}^{n\times n}$ sao cho $B = P^{-1}AP$.

Quan hệ "đồng dạng" trên $\mathbb{R}^{n\times n}$ là một quan hệ tương đương. Cụ thể, ta có

+ Tính phản xạ: với mọi $A \in \mathbb{R}^{n\times n}$, $A \sim A$ do $A = I^{-1}AI$, với ma trận đơn vị (khả nghịch) $I \in \mathbb{R}^{n\times n}$.

+ Tính đối xứng: với mọi $A, B \in \mathbb{R}^{n\times n}$, nếu $A \sim B$, nghĩa là tồn tại ma trận khả nghịch $P \in \mathbb{R}^{n\times n}$ sao cho $B = P^{-1}AP$, thì với ma trận (khả nghịch) $Q = P^{-1} \in \mathbb{R}^{n\times n}$, ta được $A = PBP^{-1} = Q^{-1}BQ$, nên $B \sim A$.

+ Tính bắc cầu: với mọi $A, B, C \in \mathbb{R}^{n\times n}$, nếu $A \sim B$ và $B \sim C$, nghĩa là tồn tại các ma trận khả nghịch $P, Q \in \mathbb{R}^{n\times n}$ sao cho $B = P^{-1}AP$ và $C = Q^{-1}BQ$, thì do

$$\begin{aligned} C = Q^{-1}BQ &= Q^{-1}(P^{-1}AP)Q = (Q^{-1}P^{-1})A(PQ) \\ &= (PQ)^{-1}A(PQ) \equiv R^{-1}AR \end{aligned}$$

với ma trận (khả nghịch) $R = PQ \in \mathbb{R}^{n\times n}$, ta được $A \sim C$.

Xét ma trận $A \in \mathbb{R}^{n\times n}$ tương ứng với toán tử $T = T_A \in L(\mathbb{R}^n, \mathbb{R}^n)$. Định lý 1.2.2 cho thấy, ma trận của T ứng với một cơ sở bất kỳ B của $\mathbb{R}^n$, $[T]_B$, là ma trận đồng dạng với A. Hơn nữa, chiều ngược lại cũng đúng.

1.2.4. Định lý. Cho $A, B \in \mathbb{R}^{n\times n}$. Ta có B đồng dạng với A nếu và chỉ nếu tồn tại cơ sở $B = \{v_1, v_2, \cdots, v_n\}$ của $\mathbb{R}^n$ sao cho $B = [T]_B$, trong đó $T = T_A \in L(\mathbb{R}^n, \mathbb{R}^n)$.

Chứng minh. Khi B đồng dạng với A, nghĩa là tồn tại ma trận khả nghịch $P \in \mathbb{R}^{n\times n}$ sao cho $B = P^{-1}AP$. Xét hệ các vectơ cột của P

$$B = \{c_1, c_2, \cdots, c_n\}.$$

Do P khả nghịch, B là một cơ sở của $\mathbb{R}^n$, và vì $P = [c_1 \quad c_2 \quad \cdots \quad c_n]$ nên Định lý 1.2.2 cho $[T]_B = P^{-1}AP = B$.

Ví dụ 3. Xét toán tử tuyến tính $T \equiv T_A: \mathbb{R}^3 \to \mathbb{R}^3$ khảo sát trong Ví dụ 2, với

$$A = \begin{bmatrix} 4 & 2 & -1 \\ -6 & -4 & 3 \\ -6 & -6 & 5 \end{bmatrix}.$$

Áp dụng Định lý 1.2.2 với cơ sở $B = \{v_1 = (1,0,2), v_2 = (0,1,2), v_3 = (1,-3,-3)\}$, ta có

$$P = [v_1 \quad v_2 \quad v_3] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -3 \\ 2 & 2 & -3 \end{bmatrix}$$

và

$$[T]_B = P^{-1}AP =$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -3 \\ 2 & 2 & -3 \end{bmatrix}^{-1} \begin{bmatrix} 4 & 2 & -1 \\ -6 & -4 & 3 \\ -6 & -6 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -3 \\ 2 & 2 & -3 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

1.2.5 Định nghĩa. Ma trận $A \in \mathbb{R}^{n\times n}$ được gọi là **chéo hóa được** (diagonalizable) khi nó đồng dạng với một ma trận chéo, nghĩa là, tồn tại ma trận khả nghịch $P \in \mathbb{R}^{n\times n}$ sao cho $D = P^{-1}AP$ là một ma trận chéo. Khi đó, ta nói ma trận P **chéo hóa** (diagonalize) A.

Ví dụ 4. Ma trận A trong Ví dụ 3 là chéo hóa được và

$$P = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -3 \\ 2 & 2 & -3 \end{bmatrix}$$

là ma trận chéo hóa A.

Kết hợp với Định lý 1.2.4, ta được

1.2.6. Định lý. Xét ma trận $A \in \mathbb{R}^{n\times n}$ tương ứng với toán tử tuyến tính $T = T_A \in L(\mathbb{R}^n, \mathbb{R}^n)$. Ma trận A là chéo hóa được nếu và chỉ nếu tồn tại cơ sở $B = \{v_1, v_2, \cdots, v_n\}$ của $\mathbb{R}^n$ sao cho $[T]_B$ là ma trận chéo.

Nhận xét 2. Khi $[T]_B$ là ma trận chéo, chẳng hạn

$$[T]_B = diag(k_1, k_2, \cdots, k_n),$$

do $[T]_B = [[T(v_1)]_B \quad [T(v_2)]_B \quad \cdots \quad [T(v_n)]_B]$, ta suy ra

$$[T(v_1)]_B = \begin{bmatrix} k_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, [T(v_2)]_B = \begin{bmatrix} 0 \\ k_2 \\ \vdots \\ 0 \end{bmatrix}, \cdots, [T(v_n)]_B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ k_n \end{bmatrix}$$

và do đó $T(v_1) = k_1v_1, T(v_2) = k_2v_2, \cdots, T(v_n) = k_nv_n$.

Ví dụ 5. Tiếp tục Ví dụ 4, do

$$[T]_B = [[T(v_1)]_B \quad [T(v_2)]_B \quad [T(v_3)]_B] = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

ta được

$[T(v_1)]_B = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$, nghĩa là $T(v_1) = 2v_1 + 0v_2 + 0v_3 = 2v_1$;

$[T(v_2)]_B = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$, nghĩa là $T(v_2) = 0v_1 + 2v_2 + 0v_3 = 2v_2$; và

$[T(v_3)]_B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, nghĩa là $T(v_3) = 0v_1 + 0v_2 + 1v_3 = 1v_3$.

1.2.7 Định nghĩa. Vectơ $x \in \mathbb{R}^n \backslash \{0\}$ được gọi là một **vectơ riêng** (eigenvector) của A, và cũng là của T_A, khi tồn tại $\lambda \in \mathbb{R}$ sao cho

$$T_A(x) = Ax = \lambda x.$$

Khi đó, λ được gọi là một **trị riêng** (eigenvalue) của A, và cũng là của T_A, và vectơ x được gọi là một **vectơ riêng tương ứng với trị riêng** λ. Cho $I \in \mathbb{R}^{n\times n}$ là ma trận đơn vị, ứng với mỗi trị riêng λ, không gian véc tơ

$$E_\lambda = \{x \in \mathbb{R}^n | Ax = \lambda x\} = \{x \in \mathbb{R}^n | (A - \lambda I)x = 0\}$$
$$= null(A - \lambda I),$$

được gọi là **không gian riêng** (eigenspace) **tương ứng với** trị riêng λ.

Ví dụ 6. Tiếp tục các Ví dụ 3, 4, và 5, ta có $\lambda = 2$ là một trị riêng của A, v_1 và v_2 là hai vectơ riêng của A tương ứng với trị riêng $\lambda = 2$, và không gian riêng tương ứng với trị riêng $\lambda = 2$ là

$$E_2 = null(A - 2I).$$

Tương tự, $\lambda = 1$ cũng là một trị riêng của A, v_3 là vectơ riêng của A tương ứng với trị riêng $\lambda = 1$, và không gian riêng tương ứng với trị riêng $\lambda = 1$ là

$$E_1 = null(A - I).$$ ■

Xuất phát từ nhận xét rằng λ là một trị riêng của A nếu và chỉ nếu hệ phương trình tuyến tính thuần nhất $(A - \lambda I)x = 0$ có ít nhất một **nghiệm không tầm thường** (nontrivial solution), ta được

1.2.8. Định lý. λ là một trị riêng của $A \in \mathbb{R}^{n\times n}$ nếu và chỉ nếu

$$det(A - \lambda I) = 0. \tag{10}$$

Vế trái của (10), $det(A - \lambda I)$, là một đa thức bậc n theo λ mà ta gọi là **đa thức đặc trưng** (characteristic polynomial) của A, và (10) được gọi là **phương trình đặc trưng** (characteristic equation) của A. Giải (10), ta nhận được các trị riêng, nếu có, của A, và ứng với mỗi trị riêng λ, giải hệ phương trình tuyến tính thuần nhất với ma trận các hệ số $A - \lambda I$, ta xác định được không gian riêng E_λ tương ứng.

Ví dụ 7. Với ma trận

$$A = \begin{bmatrix} 4 & 2 & -1 \\ -6 & -4 & 3 \\ -6 & -6 & 5 \end{bmatrix}$$

khảo sát trong các Ví dụ 3-7, ta có đa thức đặc trưng

$$det(A - \lambda I) = \begin{vmatrix} 4-\lambda & 2 & -1 \\ -6 & -4-\lambda & 3 \\ -6 & -6 & 5-\lambda \end{vmatrix} \begin{matrix} 4-\lambda & 2 \\ -6 & -4-\lambda \\ -6 & -6 \end{matrix}$$
$$= (4-\lambda)(-4-\lambda)(5-\lambda) + 2 \cdot 3 \cdot (-6) + (-1) \cdot (-6) \cdot (-6)$$
$$-[(-6)(-4-\lambda)(-1) + (-6) \cdot 3 \cdot (4-\lambda) + (5-\lambda)(-6) \cdot 2]$$
$$= -\lambda^3 + 5\lambda^2 - 8\lambda + 4 = -(\lambda-1)(\lambda-2)^2$$

Giải phương trình đặc trưng, ta được hai trị riêng $\lambda = 1$ và $\lambda = 2$ cho A. Ứng với trị riêng $\lambda = 1$, biến đổi

$$A - I = \begin{bmatrix} 3 & 2 & -1 \\ -6 & -5 & 3 \\ -6 & -6 & 4 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix},$$

ta được nghiệm tổng quát của $(A - I)x = 0$ là $x = \left(-\frac{1}{3}t, t, t\right) = t\left(-\frac{1}{3}, 1, 1\right)$, với $t \in \mathbb{R}$ tùy ý.

Suy ra

$$E_1 = null(A - I) = span\left\{\left(-\frac{1}{3}, 1,1\right)\right\}.$$

Tương tự, với trị riêng $\lambda = 2$,

$$A - 2I = \begin{bmatrix} 2 & 2 & -1 \\ -6 & -6 & 3 \\ -6 & -6 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

cho nghiệm của $(A - 2I)x = 0$ là $x = \left(-s + \frac{1}{2}t, s, t\right) = s(-1,1,0) + t\left(\frac{1}{2}, 0,1\right)$, với $s, t \in \mathbb{R}$ tùy ý. Suy ra

$$E_2 = null(A - 2I) = span\left\{(-1,1,0), \left(\frac{1}{2}, 0,1\right)\right\}.$$

1.2.9. Định lý. Giả sử ma trận $A \in \mathbb{R}^{n\times n}$ có các trị riêng phân biệt $\lambda_1, \lambda_2, \cdots, \lambda_k$. Gọi B_i là một cơ sở của không gian riêng E_{λ_i}, $i = 1,2, \cdots, k$. Ta có

$$B = \bigcup_{i=1}^{k} B_i$$

là tập độc lập tuyến tính.

Mặt khác, kết hợp Định lý 1.2.6, Nhận xét 2, và Định nghĩa 1.2.7, ta có

1.2.10. Hệ quả. Ma trận $A \in \mathbb{R}^{n\times n}$ là chéo hóa được khi và chỉ khi tồn tại một cơ sở $B = \{v_1, v_2, \cdots, v_n\}$ của $\mathbb{R}^n$ gồm toàn các vectơ riêng của A.

Kết hợp Định lý 1.2.9 với Hệ quả 1.2.10, ta được

Giải thuật Chéo hóa ma trận $A \in \mathbb{R}^{n\times n}$

Bước 1. Tìm tất cả các trị riêng $\lambda_1, \lambda_2, \cdots, \lambda_k$của A.

Bước 2. Ứng với mỗi trị riêng λ_i, tìm một cơ sở B_i cho không gian riêng tương ứng E_{λ_i}, $i = 1,2,\cdots,k$.

Bước 3. Xét $B = \cup_{i=1}^{k} B_i$. Ta có 2 khả năng

Khả năng 1. $|B| < n$: Ma trận A không chéo hóa được.

Khả năng 2. $|B| = n$: Ma trận A chéo hóa được. Khi đó, với $B = \{v_1, v_2, \cdots, v_n\}$, ma trận

$P = [v_1 \quad v_2 \quad \cdots \quad v_n]$ là ma trận chéo hóa A, và

$$P^{-1}AP = diag(a_1, a_2, \cdots, a_n),$$

với a_i là trị riêng tương ứng với vectơ riêng v_i, $i = 1,2,\cdots,n$.■

Ví dụ 8. (i) Chéo hóa ma trận

$$A = \begin{bmatrix} 4 & 2 & -1 \\ -6 & -4 & 3 \\ -6 & -6 & 5 \end{bmatrix}$$

cho trong Ví dụ 7. Ta có

Bước 1. A có hai trị riêng $\lambda = 1$ và $\lambda = 2$.

Bước 2. Không gian riêng E_1 có một cơ sở là $B_1 = \left\{\left(-\frac{1}{3}, 1,1\right)\right\}$ và không gian riêng E_2 có một cơ sở là $B_2 = \left\{(-1,1,0), \left(\frac{1}{2}, 0,1\right)\right\}$.

Bước 3. Xét $B = B_1 \cup B_2 = \left\{v_1 = \left(-\frac{1}{3}, 1,1\right), v_2 = \left(-\frac{1}{2}, 1,0\right), v_3 = \left(\frac{1}{2}, 0,1\right)\right\}$. Do $|B| = 3$ (khả năng 2), ta suy ra A là ma trận chéo hóa được,

$$P = [v_1 \quad v_2 \quad v_3] = \begin{bmatrix} -\frac{1}{3} & -1 & \frac{1}{2} \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

là ma trận chéo hóa A, và

$$P^{-1}AP = diag(1,2,2) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

(ii) Chéo hóa ma trận

$$A = \begin{bmatrix} -3 & 1 & -1 \\ -7 & 5 & -1 \\ -6 & 6 & -2 \end{bmatrix}.$$

Bước 1. Tìm các trị riêng của A bằng cách giải phương trình đặc trưng

$$0 = det(A - \lambda I) = \begin{vmatrix} -3-\lambda & 1 & -1 \\ -7 & 5-\lambda & -1 \\ -6 & 6 & -2-\lambda \end{vmatrix} \begin{matrix} -3-\lambda & 1 \\ -7 & 5-\lambda \\ -6 & 6 \end{matrix}$$

$$= (-3-\lambda)(5-\lambda)(-2-\lambda) + 6 + 42 - [6(5-\lambda) - 6(-3-\lambda) - 7(-2-\lambda)]$$

$$= -\lambda^3 + 12\lambda + 16 = -(\lambda+2)^2(\lambda+4)$$

ta được hai trị riêng $\lambda = -2$ và $\lambda = 4$.

Bước 2. Với $\lambda = -2$, do

$$A-(-2)I=\begin{bmatrix}-1 & 1 & -1\\ -7 & 7 & -1\\ -6 & 6 & 0\end{bmatrix}\to\begin{bmatrix}1 & -1 & 1\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix}$$

ta được nghiệm tổng quát của $(A-(-2)I)x=0$ là $x=(t,t,0)=t(1,1,0)$, với $t\in\mathbb{R}$ tùy ý. Suy ra

$E_{-2}=null(A-(-2)I)=span\{(1,1,0)\}$ nhận $B_1=\{(1,1,0)\}$

làm một cơ sở.

Với $\lambda=4$,

$$A-4I=\begin{bmatrix}-7 & 1 & -1\\ -7 & 1 & -1\\ -6 & 6 & -6\end{bmatrix}\to\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & -1\\ 0 & 0 & 0\end{bmatrix}$$

ta được nghiệm tổng quát của $(A-4I)x=0$ là $x=(0,t,t)=t(0,1,1)$, với $t\in\mathbb{R}$ tùy ý. Suy ra

$E_4=null(A-4I)=span\{(0,1,1)\}$ nhận $B_2=\{(0,1,1)\}$
làm một cơ sở.

Bước 3. Xét $B=B_1\cup B_2=\{v_1=(1,1,0),v_2=(0,1,1)\}$. Do $|B|<3$ (khả năng 1), ta kết luận rằng ma trận A không chéo hóa được. ■

Đặc biệt, ta chấp nhận kết quả sau

1.2.11. Định lý. Mọi ma trận đối xứng đều có thể chéo hóa được.

(mà phần chứng minh cần đến các khái niệm về không gian vectơ phức. Bạn đọc quan tâm đến kết quả này có thể tham khảo trong [1, 2, 3]).

Nhắc lại $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ là ma trận đối xứng khi $A^T = A$, nghĩa là $a_{ij} = a_{ji}$, với mọi $i, j = 1,2,\cdots,n$. Ta có một số ma trận đối xứng quan trọng sau:

+ Ma trận hiệp phương sai của một vectơ ngẫu nhiên $X = (X_1, X_2, \cdots, X_n)$ trong xác suất thống kê $cov(X) = [cov(X_i, X_j)]_{i,j=1}^{n}$ do $cov(X_i, X_j) = cov(X_j, X_i)$, với mọi $i, j = 1,2,\cdots,n$.

+ Ma trận Hesse của hàm n-biến $y = f(x_1, x_2, \cdots, x_n)$ thuộc lớp C^2,

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_j \partial x_i}(x)\right] =$$

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix},$$

với $x = (x_1, x_2, \cdots, x_n)$, do

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x), \text{ với mọi } i, j = 1,2,\cdots,n.$$

+ Với $A \in \mathbb{R}^{m\times n}$ bất kỳ, các ma trận

$$A^T A \in \mathbb{R}^{n\times n} \text{ và } AA^T \in \mathbb{R}^{m\times m}$$

là các ma trận đối xứng do

$$(A^T A)^T = A^T (A^T)^T = A^T A \text{ và } (AA^T)^T = (A^T)^T A^T = AA^T. \blacksquare$$

1.3. CHÉO HÓA TRỰC GIAO

Trước hết, ta nhắc lại một số thuật ngữ trên không gian Euclide $\mathbb{R}^n$.

1.3.1. Định nghĩa

Cho $u = (u_1, u_2, \cdots, u_n), v = (v_1, v_2, \cdots, v_n) \in \mathbb{R}^n$.

+ **Tích vô hướng** (dot product/scalar product) của u, v, ký hiệu $\langle u, v\rangle$, xác định bởi

$$\langle u, v\rangle = u_1v_1 + u_2v_2 + \cdots + u_nv_n = \sum_{i=1}^{n} u_iv_i.$$

Khi $\langle u, v\rangle = 0$, ta nói u, v **trực giao** (orthogonal) với nhau, ký hiệu $u \perp v$.

+ **Chuẩn** (norm) của u, ký hiệu $\|u\|$, xác định bởi

$$\|u\| = \sqrt{\langle u, u\rangle} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} = \sqrt{\sum_{i=1}^{n} u_i^2}.$$

Khi $\|u\| = 1$, ta nói u là một **vectơ đơn vị** (unit vector).

+ Xét hệ các vectơ $B = \{u_1, u_2, \cdots, u_k\} \subset \mathbb{R}^n$, với $u_i \neq 0$, $i = 1,2, \cdots, k$. Ta nói B là hệ **trực giao** (orthogonal) khi $\langle u_i, u_j\rangle = 0$, với mọi $i, j = 1,2, \cdots, k$, $i \neq j$, và là hệ **trực chuẩn** (orthonormal) khi nó là hệ trực giao gồm các vectơ đơn vị, nghĩa là

$$\langle u_i, u_j\rangle = \begin{cases} 1, & khi\ i = j, \\ 0, & khi\ i \neq j. \end{cases}$$

Chú ý rằng mọi hệ trực giao đều là hệ độc lập tuyến tính nên hệ trực giao (trực chuẩn) gồm n vectơ, $B = \{u_1, u_2, \cdots, u_n\} \subset \mathbb{R}^n$, được gọi là một **cơ sở trực giao (trực chuẩn)** của $\mathbb{R}^n$.

Ví dụ 9. Cơ sở chính tắc $S = \{e_1 = (1,0,\cdots,0), e_2 = (0,1,\cdots,0), \cdots, e_n = (0,0,\cdots,1)\}$ là một cơ sở trực chuẩn của $\mathbb{R}^n$.

Trước hết, ta có các tính chất cơ bản của tích vô hướng như sau:

1.3.2. Mệnh đề. Với mọi $u, v, u_1, u_2 \in \mathbb{R}^n$ và $h, k \in \mathbb{R}$, ta có

(i) (tính đối xứng) $\langle u, v\rangle = \langle v, u\rangle$.

(ii) (tính tuyến tính) $\langle hu_1 + ku_2, v\rangle = h\langle u_1, v\rangle + k\langle u_2, v\rangle$.

Cho $S = \{u_1, u_2, \cdots, u_k\}$ là một hệ trực giao của $\mathbb{R}^n$. Ta có

1.3.3. Định lý. Với mỗi vectơ $v \in \mathbb{R}^n$, đặt

$$u_{k+1} = v - \frac{\langle v,u_1\rangle}{\|u_1\|^2}u_1 - \frac{\langle v,u_2\rangle}{\|u_2\|^2}u_2 - \cdots - \frac{\langle v,u_k\rangle}{\|u_k\|^2}u_k.$$

Ta có $\langle u_{k+1}, u_i\rangle = 0$, với mọi $i = 1,2,\cdots,n$.

Chứng minh. Do tính tuyến tính của tích vô hướng, ta có

$$\langle u_{k+1}, u_1\rangle = \left\langle v - \frac{\langle v,u_1\rangle}{\|u_1\|^2}u_1 - \frac{\langle v,u_2\rangle}{\|u_2\|^2}u_2 - \cdots - \frac{\langle v,u_k\rangle}{\|u_k\|^2}u_k, u_1\right\rangle$$

$$= \langle v, u_1\rangle - \frac{\langle v,u_1\rangle}{\|u_1\|^2}\langle u_1, u_1\rangle - \frac{\langle v,u_2\rangle}{\|u_2\|^2}\langle u_2, u_1\rangle - \cdots - \frac{\langle v,u_k\rangle}{\|u_k\|^2}\langle u_k, u_1\rangle$$

$$= \langle v, u_1\rangle - \langle v, u_1\rangle = 0$$

vì $\langle u_1, u_1 \rangle = \|u_1\|^2$ và $\langle u_i, u_1 \rangle = 0$, với mọi $i = 2, \cdots, k$. Các đẳng thức $\langle u_{k+1}, u_j \rangle = 0$, $j = 2, \cdots, k$, được chứng minh tương tự.

Chú ý. Khi $u_{k+1} = 0$ trong (11), hệ $\{u_1, u_2, \cdots, u_k, v\}$ là phụ thuộc tuyến tính do v là một tổ hợp tuyến tính các vectơ trong $\{u_1, u_2, \cdots, u_k\}$.

Từ Định lý 1.3.3, ta được giải thuật xây dựng hệ trực giao $\{v_1, v_2, \cdots, v_k\}$ từ một hệ độc lập tuyến tính $S = \{u_1, u_2, \cdots, u_k\}$ sao cho $span\{v_1, v_2, \cdots, v_h\} = span\{u_1, u_2, \cdots, u_h\}$, với mọi $h = 1,2, \cdots, k$.

Giải thuật Gram-Schmidt

Lần lượt thực hiện k bước sau:

Bước 1. Đặt $v_1 = u_1$.

Bước 2. Đặt $v_2 = u_2 - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1$.

Bước 3. Đặt $v_3 = u_3 - \frac{\langle u_3, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle u_2, v_2 \rangle}{\|v_2\|^2} v_2$

Bước 4. Đặt $v_4 = u_4 - \frac{\langle u_4, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle u_4, v_2 \rangle}{\|v_2\|^2} v_2 - \frac{\langle u_4, v_3 \rangle}{\|v_3\|^2} v_3$.⋮

cho tới bước k.

Ngoài ra, ta còn xây dựng được hệ trực chuẩn $\{q_1, q_2, \cdots, q_k\}$ từ hệ S bằng cách đặt

$$q_i = \frac{v_i}{\|v_i\|}, \text{ với } i = 1,2, \cdots, k,$$

Ví dụ 10. Dùng giải thuật Gram-Schmidt, xây dựng hệ trực chuẩn từ các hệ độc lập tuyến tính sau.

(a) Hệ $\{u_1 = (1,0,1), u_2 = (1,1,0)\}$.

Bước 1. $v_1 = u_1 = (1,0,1)$.

Bước 2.

$$v_2 = u_2 - \frac{\langle u_2, v_1\rangle}{\|v_1\|^2} v_1 =$$
$$(1,1,0) - \frac{(1)(1)+(1)(0)+(0)(1)}{1^2+0^2+1^2}(1,0,1) = \left(\frac{1}{2}, 1, -\frac{1}{2}\right).$$

Ta được hệ trực giao $\left\{v_1 = (1,0,1), v_2 = \left(\frac{1}{2}, 1, -\frac{1}{2}\right)\right\}$. Hơn nữa, với

$$q_1 = \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{1^2+0^2+1^2}}(1,0,1) = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right), \quad q_2 = \frac{v_2}{\|v_2\|} = \frac{1}{\sqrt{\left(\frac{1}{2}\right)^2+1^2+\left(-\frac{1}{2}\right)^2}}\left(\frac{1}{2}, 1, -\frac{1}{2}\right) = \left(\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right)$$

ta được hệ trực chuẩn $\left\{q_1 = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right), q_2 = \left(\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right)\right\}$.

(b) Với $\{u_1 = (1,0,1), u_2 = (1,1,0), u_3 = (0,1,1)\}$,

Bước 1. $v_1 = u_1 = (1,0,1)$.

Bước 2. $v_2 = u_2 - \frac{\langle u_2, v_1\rangle}{\|v_1\|^2} v_1$

$$= (1,1,0) - \frac{(1)(1)+(1)(0)+(0)(1)}{1^2+0^2+1^2}(1,0,1)$$

$$= \left(\frac{1}{2}, 1, -\frac{1}{2}\right).$$

Bước. $v_3 = u_3 - \frac{\langle u_3, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle u_3, v_2 \rangle}{\|v_2\|^2} v_2,$

$$v_3 = (0,1,1) - \frac{(0)(1)+(1)(0)+(1)(1)}{1^2+0^2+1^2}(1,0,1) -$$
$$\frac{(0)\left(\frac{1}{2}\right)+(1)(1)+(1)\left(-\frac{1}{2}\right)}{\left(\frac{1}{2}\right)^2+1^2+\left(-\frac{1}{2}\right)^2}\left(\frac{1}{2},1,-\frac{1}{2}\right) = \left(-\frac{2}{3},\frac{2}{3},\frac{2}{3}\right).$$

Ta được hệ trực giao $\left\{v_1 = (1,0,1), v_2 = \left(\frac{1}{2},1,-\frac{1}{2}\right), v_3 = \left(-\frac{2}{3},\frac{2}{3},\frac{2}{3}\right)\right\}$. Hơn nữa, với

$$q_1 = \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{1^2+0^2+1^2}}(1,0,1) = \left(\frac{1}{\sqrt{2}},0,\frac{1}{\sqrt{2}}\right),$$

$$q_2 = \frac{v_2}{\|v_2\|} = \frac{1}{\sqrt{\left(\frac{1}{2}\right)^2+1^2+\left(-\frac{1}{2}\right)^2}}\left(\frac{1}{2},1,-\frac{1}{2}\right) = \left(\frac{1}{\sqrt{6}},\frac{2}{\sqrt{6}},-\frac{1}{\sqrt{6}}\right).$$

$$q_3 = \frac{v_3}{\|v_3\|} = \frac{1}{\sqrt{\left(-\frac{2}{3}\right)^2+\left(\frac{2}{3}\right)^2+\left(\frac{2}{3}\right)^2}}\left(-\frac{2}{3},\frac{2}{3},\frac{2}{3}\right) = \left(-\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right).$$

ta được hệ trực chuẩn $\left\{q_1 = \left(\frac{1}{\sqrt{2}},0,\frac{1}{\sqrt{2}}\right), q_2 = \left(\frac{1}{\sqrt{6}},\frac{2}{\sqrt{6}},-\frac{1}{\sqrt{6}}\right), q_3 = \left(-\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)\right\}$ cho $\mathbb{R}^3$.

Nhận xét 3. Cho $A = [a_{ij}], B = [b_{ij}] \in \mathbb{R}^{n\times n}$. Xét các vectơ dòng của A,

$$r_i^A = (a_{i1}, a_{i2}, \cdots, a_{in}), i = 1,2,\cdots,n,$$

và các vectơ cột của B,

$$c_j^B = (b_{1j}, b_{2j}, \cdots, b_{nj}), j = 1,2,\cdots,n.$$

Ma trận tích AB có thể biểu diễn bằng ma trận các tích vô hướng giữa vectơ dòng của A với vectơ cột của B,

$$AB = \left[\langle r_i^A, c_j^B \rangle\right]_{i,j=1}^{n} = \begin{bmatrix} \langle r_1^A, c_1^B \rangle & \langle r_1^A, c_2^B \rangle & \cdots & \langle r_1^A, c_n^B \rangle \\ \langle r_2^A, c_1^B \rangle & \langle r_2^A, c_2^B \rangle & \cdots & \langle r_2^A, c_n^B \rangle \\ \vdots & \vdots & & \vdots \\ \langle r_n^A, c_1^B \rangle & \langle r_n^A, c_2^B \rangle & \cdots & \langle r_n^A, c_n^B \rangle \end{bmatrix}.$$

1.3.4. Định nghĩa. Ma trận $A \in \mathbb{R}^{n\times n}$ được gọi là **trực giao** (orthogonal) khi

$$A^{-1} = A^T,$$

nghĩa là $AA^T = A^TA = I$.

Từ Nhận xét 3, ta được

1.3.5. Định lý. Cho $A \in \mathbb{R}^{n\times n}$. Các phát biểu sau là tương đương

(i) A là ma trận trực giao.

(ii) Các vectơ dòng của A tạo thành một cơ sở trực chuẩn cho $\mathbb{R}^n$.

(iii) Các vectơ cột của A tạo thành một cơ sở trực chuẩn cho $\mathbb{R}^n$.

Bằng cách thay ma trận khả nghịch P chéo hóa A, bằng ma trận trực giao P, đẳng thức $P^{-1}AP = D$ trở thành $P^TAP = D$. Ta có khái niệm "chéo hóa trực giao" sau

1.3.6. Định nghĩa

(i) Hai ma trận vuông $A, B \in \mathbb{R}^{n\times n}$ được gọi là **đồng dạng trực giao** (orthogonally similar) khi tồn tại ma trận trực giao $P \in \mathbb{R}^{n\times n}$ sao cho $B = P^TAP$.

(ii) Ma trận $A \in \mathbb{R}^{n\times n}$ được gọi là **chéo hóa trực giao được** (orthogonally diagonalizable) khi nó đồng dạng trực giao với một ma trận chéo, nghĩa là tồn tại ma trận trực giao $P \in \mathbb{R}^{n\times n}$ sao cho $P^T AP = D$ là một ma trận chéo. Khi đó, ta nói P **chéo hóa trực giao** (orthogonally diagonalizes) A.

Nhận xét 4. (i) Nếu $A \in \mathbb{R}^{n\times n}$ chéo hóa trực giao được thì nó cũng chéo hóa được nên do phần 1.2, điều này chỉ xảy ra khi tồn tại cơ sở $B = \{v_1, v_2, \cdots, v_n\}$ của $\mathbb{R}^n$ gồm toàn các vectơ riêng của A và ma trận P chéo hóa trực giao A cho bởi

$$P = [v_1 \quad v_2 \quad \cdots \quad v_n].$$

Do P là ma trận trực giao, Định lý 1.3.5 cho thấy $B = \{v_1, v_2, \cdots, v_n\}$ phải là một cơ sở trực chuẩn gồm các vectơ riêng của $\mathbb{R}^n$.

(ii) Đẳng thức $P^T AP = D$ cho $A = PDP^T$ (do $P^T = P^{-1}$). Khi đó, ta có

$$A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A.$$

Điều này cho thấy nếu A là ma trận chéo hóa trực giao được thì nó phải là ma trận đối xứng.

Ngược lại, chú ý rằng do Định lý 1.2.11, mọi ma trận đối xứng A đều chéo hóa được, nghĩa là tồn tại một cơ sở gồm toàn các vectơ riêng của A.

Bằng cách đồng nhất một vectơ $u = (u_1, u_2, \cdots, u_n)$ trong $\mathbb{R}^n$ với ma trận cột,

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^{n\times 1},$$

ta có mối liên hệ giữa tích vô hướng và phép toán ma trận sau.

1.3.7. Mệnh đề. Với mọi vectơ $u, v \in \mathbb{R}^n$, và ma trận $A \in \mathbb{R}^{n\times n}$,

(i) $\langle u, v\rangle = u^T \cdot v = v^T \cdot u$.

(ii) $\langle Au, v\rangle = \langle u, A^T v\rangle$.

Chứng minh. (i) Với $u = (u_1, u_2, \cdots, u_n), v = (v_1, v_2, \cdots, v_n)$ biểu diễn bằng ma trận cột, ta có

$$u^T \cdot v = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = [\textstyle\sum_{i=1}^{n} u_i v_i] \equiv \displaystyle\sum_{i=1}^{n} u_i v_i = \langle u, v\rangle.$$

(ii) Do (i) $\langle Au, v\rangle = (Au)^T \cdot v = (u^T A^T) \cdot v = u^T \cdot (A^T v) = \langle u, A^T v\rangle$.

1.3.8. Định lý. Các vectơ riêng tương ứng với các trị riêng khác nhau của một ma trận đối xứng đều trực giao với nhau.

Chứng minh. Cho $A \in \mathbb{R}^{n\times n}$ là một ma trận đối xứng với các trị riêng λ_i và không gian riêng tương ứng E_i, với $i = 1,2,\cdots,k$. Với $v_i \in E_i$, $v_j \in E_j$, $i \neq j$, ta có $Av_i = \lambda_i v_i$, $Av_j = \lambda_j v_j$, và $\lambda_i \neq \lambda_j$. Do mệnh đề 1.3.7,

$$\langle Av_i, v_j \rangle = \langle v_i, A^T v_j \rangle = \langle v_i, Av_j \rangle,$$

với

$$\langle Av_i, v_j \rangle = \langle \lambda_i v_i, v_j \rangle = \lambda_i \langle v_i, v_j \rangle$$

và

$$\langle v_i, Av_j \rangle = \langle v_i, \lambda_j v_j \rangle = \lambda_j \langle v_i, v_j \rangle.$$

Do đó $(\lambda_i - \lambda_j)\langle v_i, v_j \rangle = 0$ và vì $\lambda_i \neq \lambda_j$ nên ta phải có $\langle v_i, v_j \rangle = 0$, nghĩa là $v_i \perp v_j$.

Kết hợp Nhận xét 4, Định lý 1.3.8, và Giải thuật Chéo hóa ma trận $A \in \mathbb{R}^{n\times n}$ trong phần 1.2, ta được

Giải thuật Chéo hóa trực giao ma trận đối xứng $A \in \mathbb{R}^{n\times n}$

Bước 1. Tìm tất cả các trị riêng $\lambda_1, \lambda_2, \cdots, \lambda_k$ của A.

Bước 2. Ứng với mỗi trị riêng λ_i, tìm một cơ sở trực chuẩn B_i cho không gian riêng tương ứng E_{λ_i}, $i = 1,2,\cdots,k$.

Bước 3. Xét $B = \cup_{i=1}^{k} B_i = \{v_1, v_2, \cdots, v_n\}$. Ma trận $P = [v_1 \quad v_2 \quad \cdots \quad v_n]$ chính là ma trận chéo hóa trực giao ma trận A.

Ví dụ 11. Chéo hóa trực giao ma trận

$$A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix}.$$

Bước 1. Giải phương trình đặc trưng

$$0 = det(A - \lambda I) = \begin{vmatrix} 4-\lambda & 2 & 2 \\ 2 & 4-\lambda & 2 \\ 2 & 2 & 4-\lambda \end{vmatrix} \begin{matrix} 4-\lambda & 2 \\ 2 & 4-\lambda \\ 2 & 2 \end{matrix}$$

$$= (4-\lambda)^3 + 16 - 12(4-\lambda) = -\lambda^3 + 12\lambda^2 - 36\lambda + 32$$
$$= -(\lambda - 2)^2(\lambda - 8),$$

ta được hai trị riêng $\lambda = 2$ và $\lambda = 8$.

Bước 2. Ứng với trị riêng $\lambda = 2$, biến đổi

$$A - 2I = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

ta được nghiệm tổng quát của $(A - 2I)x = 0$ là $x = (-s - t, s, t) = s(-1,1,0) + t(-1,0,1)$, với $t \in \mathbb{R}$ tùy ý. Suy ra $E_2 = null(A - 2I) = span\{u_1 = (-1,1,0), u_2 = (-1,0,1)\}$.

Dùng giải thuật Gram-Schmidt cho hệ $\{u_1, u_2\}$,

$$v_1 = u_1 = (-1,1,0), v_2 = u_2 - \frac{\langle u_2, v_1\rangle}{\|v_1\|^2} v_1$$

$$= (-1,0,1) - \frac{(-1)(-1) + (0)(1) + (1)(0)}{(-1)^2 + 1^2 + 0^2}(-1,1,0)$$

$$= \left(-\frac{1}{2}, -\frac{1}{2}, 1\right),$$

ta được hệ trực giao $\{v_1, v_2\}$, và với

$$q_1 = \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{(-1)^2+1^2+0^2}}(-1,1,0) = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right),$$

$$q_2 = \frac{v_2}{\|v_2\|} = \frac{1}{\sqrt{\left(-\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 + 1^2}}\left(-\frac{1}{2}, -\frac{1}{2}, 1\right) = \left(-\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right),$$

ta được một cơ sở trực chuẩn $B_1 = \left\{q_1 = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), q_2 = \left(-\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right)\right\}$ cho không gian riêng E_2.

Với trị riêng $\lambda = 8$,

$$A - 8I = \begin{bmatrix} -4 & 2 & 2 \\ 2 & -4 & 2 \\ 2 & 2 & -4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

ta được nghiệm tổng quát của $(A - 8I)x = 0$ là $x = (t, t, t) = t(1,1,1)$, với $t \in \mathbb{R}$ tùy ý.

Suy ra:

$$E_8 = null(A - 8I) = span\{u_3 = (1,1,1)\}$$

$$= span\left\{q_3 = \frac{u_3}{\|u_3\|} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)\right\}.$$

Ta nhận được cơ sở trực chuẩn $B_2 = \left\{q_3 = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)\right\}$ cho không gian riêng E_8.

Bước 3. Với $B = B_1 \cup B_2 = \{q_1, q_2, q_n\}$, ma trận

$$P = [q_1 \quad q_2 \quad q_3] = \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

là ma trận chéo hóa trực giao ma trận A,

$$P^TAP = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{bmatrix}.$$

1.4. PHÂN TÍCH GIÁ TRỊ KỲ DỊ

Trước hết, chú ý rằng xuất phát từ một hệ trực chuẩn

$$S = \{u_1, u_2, \cdots, u_n\} \subset \mathbb{R}^m, n < m,$$

bằng cách xây dựng một cơ sở trực chuẩn $\{u_{n+1}, u_{n+2}, \cdots, u_m\}$ cho không gian nghiệm của hệ phương trình tuyến tính thuần nhất

$$\begin{cases} \langle u_1, x\rangle & = & 0 \\ \langle u_2, x\rangle & = & 0 \\ \vdots & \vdots & \vdots \\ \langle u_n, x\rangle & = & 0 \end{cases}$$

ta được một cơ sở trực chuẩn $B = \{u_1, u_2, \cdots, u_n, u_{n+1}, u_{n+2}, \cdots, u_m\}$ của $\mathbb{R}^m$.

Ví dụ 12. Với hệ trực chuẩn $S = \left\{u_1 = \left(\frac{\sqrt{6}}{3}, \frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6}\right), u_2 = \left(0, \frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)\right\}$ trong $\mathbb{R}^3$, hệ phương trình tuyến tính thần nhất theo ẩn $x = (x_1, x_2, x_3) \in \mathbb{R}^3$,

$$\begin{cases}\langle u_1, x\rangle &= 0\\ \langle u_2, x\rangle &= 0,\end{cases}$$

có ma trận các hệ số

$$\begin{bmatrix}\frac{\sqrt{6}}{3} & \frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{6}\\ 0 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2}\end{bmatrix} \to \begin{bmatrix}1 & 0 & 1\\ 0 & 1 & -1\end{bmatrix}$$

cho không gian nghiệm $span\{(-1,1,1)\}$. Do đó, với $u_3 = \frac{1}{\|(-1,1,1)\|}(-1,1,1) = \left(-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$, ta được một cơ sở trực chuẩn $B = \{u_1, u_2, u_3\}$ của $\mathbb{R}^3$.

Ta chấp nhận kết quả sau cho **phân tích giá trị kỳ dị** (Singular Value Decomposition – SVD) của một ma trận bất kỳ. Bạn đọc quan tâm tới chứng minh có thể tham khảo trong [1, 2, 4].

1.4.1. Định lý Cho $A \in \mathbb{R}^{m\times n}$. Tồn tại các ma trận trực giao

$$U = [u_1 \quad u_2 \quad \cdots \quad u_m] \in \mathbb{R}^{m\times m} \text{ và}$$
$$V = [v_1 \quad v_2 \quad \cdots \quad v_n] \in \mathbb{R}^{n\times n}$$

sao cho

$$U^T AV = \Sigma = diag(\sigma_1, \sigma_2, \cdots, \sigma_p) \in \mathbb{R}^{m\times n}, \qquad (12)$$

với $p = min\{m, n\}$ và $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$.

Đẳng thức (12) còn có thể viết lại thành

$$A = U\Sigma V^T$$

và cả hai cách viết cùng được gọi là phân tích giá trị kỳ dị (SVD) của A. Khi đó, các σ_i được gọi là các **giá trị kỳ dị** (singular values) của A, các u_i là các **vectơ kỳ dị trái** (left singular vectors) của A, và v_i **vectơ kỳ dị phải** (right singular vectors) của A.

Nhận xét. Không mất tính tổng quát, ta có thể giả sử $A \in \mathbb{R}^{m\times n}$, với $m \geq n$. Khi đó, $p = n$ và $\Sigma \in \mathbb{R}^{m\times n}$ có dạng

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{m\times n}. \tag{13}$$

Bấy giờ, SVD của $B \in \mathbb{R}^{n\times m}$ có thể nhận được từ SVD của $A = B^T \in \mathbb{R}^{m\times n}$ bởi

$$B = A^T = (U\Sigma V^T)^T = V\Sigma^T U^T,$$

với

$$\Sigma^T = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n\times m}.$$

Để có thể xây dựng giải thuật SVD cho một ma trận $A \in \mathbb{R}^{m\times n}$, ta có các nhận xét sau:

1.4.2. Hệ quả. Cho $U^TAV = \Sigma$ là SVD của $A \in \mathbb{R}^{m\times n}$, với $m \geq n$. Ta có

(i) $Av_i = \sigma_i u_i$ và $A^T u_i = \sigma_i v_i$, với mọi $i = 1,2,\cdots,n$.

(ii) $A^T A v_i = \sigma_i^2 v_i$ và $AA^T u_i = \sigma_i^2 u_i$, với mọi $i = 1,2,\cdots,n$.

Chứng minh. (i) suy ra từ đẳng thức $AV = U\Sigma$ và $A^T U = V\Sigma^T$ và (ii) suy ra từ (i).

Nhận xét. Hệ quả 1.4.2 cho thấy (σ_i^2, v_i) là n cặp trị riêng/vectơ riêng của ma trận $A^T A \in \mathbb{R}^{n\times n}$. Với $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > \sigma_{k+1} = \cdots = \sigma_n = 0$, các vectơ riêng $u_1, u_2, \cdots, u_k$ của $AA^T \in \mathbb{R}^{m\times m}$, nhận được từ $v_1, v_2, \cdots, v_k$ thông qua đẳng thức

$$u_i = \frac{1}{\sigma_i} A v_i, \text{ với } i = 1,2,\cdots,k,$$

Hơn nữa, (1) và (2) cho thấy hệ trực chuẩn $\{u_{k+1}, u_{k+2}, \cdots, u_m\}$ có thể chọn tùy ý sao cho $\{u_1, u_2, \cdots, u_m\}$ là một cơ sở trực chuẩn của $\mathbb{R}^m$.

Từ nhận xét trên, ta được

Giải Thuật SVD cho $A \in \mathbb{R}^{m\times n}$, $m \geq n$.

Bước 1. Tìm tất cả các cặp trị riêng/vectơ riêng, (λ_i, v_i), $i = 1,2,\cdots,n$, của ma trận đối xứng $A^T A \in \mathbb{R}^{n\times n}$, sắp xếp theo thứ tự giảm dần,

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0,$$

và tính các giá trị kỳ dị $\sigma_i = \sqrt{\lambda_i}$, $i = 1,2,\cdots,n$.

Bước 2. Với $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > \sigma_{k+1} = \cdots = \sigma_n = 0$, tính k vectơ riêng $u_1, u_2, \cdots, u_k$ của $AA^T \in \mathbb{R}^{m\times m}$ tương ứng bằng đẳng thức

$$u_i = \frac{1}{\sigma_i} A v_i, \text{ với } i = 1,2,\cdots,k,$$

Bước 3. Nới rộng hệ trực chuẩn $\{u_1, u_2, \cdots, u_k\}$ thành một cơ sở trực chuẩn $\{u_1, u_2, \cdots, u_m\}$ cho $\mathbb{R}^m$.

Bước 4. Thành lập các ma trận

$$U = [u_1 \quad u_2 \quad \cdots \quad u_m] \in \mathbb{R}^{m\times m},$$

$$V = [v_1 \quad v_2 \quad \cdots \quad v_n] \in \mathbb{R}^{n\times n},$$

$$\text{và } \Sigma = diag(\sigma_1, \sigma_2, \cdots, \sigma_n).$$

Ta được SVD

$$U^T A V = \Sigma \text{ và } A = U \Sigma V^T.$$

Ví dụ 13. (i) Tìm SVD của ma trận

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 2}.$$

Bước 1. Tìm các cặp trị riêng/vectơ riêng của

$$A^T A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Phương trình đặc trưng $0 = det(A^T A - \lambda I) = (2 - \lambda)^2 - 1$ cho hai trị riêng $\lambda_1 = 3 \geq \lambda_2 = 1$ và ta được các giá trị kỳ dị

$$\sigma_1 = \sqrt{3}, \text{ và } \sigma_2 = 1.$$

Với $\lambda_1 = 3$, biến đổi

$$A^T A - 3I = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$

ta được không gian riêng $E_3 = span\{(1,1)\}$. Suy ra

$$v_1 = \frac{1}{\|(1,1)\|} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$

Với $\lambda_2 = 1$, biến đổi

$$A^T A - I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix},$$

ta được không gian riêng $E_3 = span\{(-1,1)\}$. Suy ra

$$v_2 = \frac{1}{\|(-1,1)\|} = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$

Bước 2. Tính các vectơ riêng u_1, u_2 của $AA^T \in \mathbb{R}^{3\times3}$:

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{6} \end{bmatrix}$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{1} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$$

Bước 3. Với hệ trực chuẩn $\{u_1, u_2\}$ trong $\mathbb{R}^3$, ta tìm được (xem **Ví dụ 12**) $u_3 = \left(-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$ sao cho $B = \{u_1, u_2, u_3\}$ là một cơ sở trực chuẩn của $\mathbb{R}^3$.

Bước 4. Thành lập các ma trận

$$U = [u_1 \quad u_2 \quad u_3] = \begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \end{bmatrix} \in \mathbb{R}^{3\times 3},$$

$$V = [v_1 \quad v_2] = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \in \mathbb{R}^{2\times 2}, \text{ và}$$

$$\Sigma = diag(\sigma_1, \sigma_2) = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 2},$$

ta được SVD

$$U\Sigma V^T = \begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}^T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = A.$$

(ii) Với ma trận

$$B = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 3},$$

do $B = A^T$ nên từ (i), ta được SVD của B

$$B = (U\Sigma V^T)^T = V\Sigma^T U^T =$$

$$\begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{3}}{3} \end{bmatrix}^T.$$

BÀI TẬP

1) Ánh xạ T có là ánh xạ tuyến tính không?

a) $T: \mathbb{R}^2 \to \mathbb{P}^2$ với $\mathbb{P}^2$ là không gian vectơ các đa thức bậc hai, cho bởi biểu thức $T(a_1, a_2) = a_1x^2 + (a_1, a_2)x - a_2$.

b) $T: \mathbb{R}^2 \to \mathbb{R}^2$, với

(i) $T(a, b) = -(a, b) + (3,2)$.

(ii) $T(a, b) = (b, a)$.

c) $T: \mathbb{R}^3 \to \mathbb{R}^2$, với

(i) $T(x_1, x_2, x_3) = (x_2, x_3)$.

(ii) $T(x_1, x_2, x_3) = (0,0)$.

(iii) $T(x_1, x_2, x_3) = (1 + x_1, x_2)$.

(iv) $T(x_1, x_2, x_3) = (x_3, x_1 + x_2)$.

d) $T: \mathbb{R}^2 \to \mathbb{R}^3$, với

(i) $T(x_1, x_2) = (x_1, x_2, 1)$.

(ii) $T(x_1, x_2) = (x_1, x_2, x_1 + 2x_2)$.

(iii) $T(x_1, x_2) = (x_1, 0,0)$.

(iv) $T(x_1, x_2) = (x_1, x_2, x_1^2 + x_2^2)$.

2) Cho ánh xạ tuyến tính $T: \mathbb{R}^2 \to \mathbb{R}^2$ thỏa $T(1,2) = (-2,3)$ và $T(1, -1) = (5,2)$. Hãy tính giá trị $T(7,5)$.

3) Cho $S = \{v_1 = (1,1), v_2 = (1,0)\}$ là một cơ sở của $\mathbb{R}^2$. Xác định ánh xạ tuyến tính $T: \mathbb{R}^2 \to \mathbb{R}^2$ sao cho $T(v_1) = (1,-2)$, và $T(v_2) = (-4,1)$.

Tìm $Ker(T)$ và $Im(T)$.

4) Cho $S = \{v_1 = (1,1,1), v_2 = (1,1,0), v_3 = (1,0,0)\}$ là một cơ sở của $\mathbb{R}^3$. Xác định ánh xạ tuyến tính $T: \mathbb{R}^3 \to \mathbb{R}^2$ sao cho $T(v_1) = (2,-1,4), T(v_2) = (3,0,1)$, và $T(v_3) = (-1,5,1)$.

Tìm $Ker(T)$ và $Im(T)$.

5) Cho $S = \{v_1 = (1,1,1), v_2 = (1,1,0), v_3 = (1,0,0)\}$ là một cơ sở của $\mathbb{R}^3$. Xác định ánh xạ tuyến tính $T: \mathbb{R}^3 \to \mathbb{R}^2$ sao cho $T(v_1) = (1,0), T(v_2) = (2,-1)$, và $T(v_3) = (4,3)$.

Tìm $Ker(T)$ và $Im(T)$.

6) Cho $S = \{v_1 = (-2,1), v_2 = (1,3)\}$ là một cơ sở của $\mathbb{R}^2$. Xác định ánh xạ tuyến tính $T: \mathbb{R}^2 \to \mathbb{R}^3$ sao cho $T(v_1) = (-1,2,0)$, và $T(v_2) = (0,-3,5)$.

Tìm $Ker(T)$ và $Im(T)$.

7) Chéo hóa các ma trận sau

a) $A = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$

b) $A = \begin{bmatrix} -2 & -7 \\ 1 & 2 \end{bmatrix}$

c) $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

d) $A = \begin{bmatrix} 4 & 0 & 1 \\ -2 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}$

e) $A = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 0 & 0 \\ -2 & 0 & 4 \end{bmatrix}$

f) $A = \begin{bmatrix} 6 & 3 & -8 \\ 0 & -2 & 0 \\ 1 & 0 & -3 \end{bmatrix}$

8) Giải thích sự phụ thuộc tuyến tính của các vectơ sau:

a) $u_1 = (-1,2,4), u_2 = (5,-10,-20)$.

b) $u_1 = (3,-1), u_2 = (4,5), u_3 = (-4,7)$.

9) Xác định hệ sau độc lập hay phụ thuộc tuyến tính trong $\mathbb{R}^3$.

a) $(-3,0,4), (5,-1,2), (1,1,3)$.

b) $(-2,0,1), (3,2,5), (6,-1,1), (7,0,-2)$.

10) Xác định sự phụ thuộc hay độc lập tuyến tính trong $\mathbb{R}^4$ của hệ các vectơ sau:

a) $(3,8,7,-3), (1,5,3,-1), (2,-1,2,6), (4,2,6,4)$.

b) $(3,0,-3,6), (0,2,3,1), (0,-2,-2,0), (-2,1,2,1)$.

11) Cho hệ các vectơ $S = \left\{v_1 = (0,1,0), v_2 = \left(-\frac{4}{5}, 0, \frac{3}{5}\right), v_3 = \left(\frac{3}{5}, 0, \frac{4}{5}\right)\right\}$.

a) Kiểm chứng rằng S là một cơ sở trực chuẩn của $\mathbb{R}^3$.

b) Tìm tọa độ của $u = (1,1,1)$ đối với cơ sở S.

12) Cho hệ các vectơ $S = \{u_1 = (0,2,0), u_2 = (3,0,3), u_3 = (-4,0,4)\}$.

a) Kiểm chứng rằng S là một cơ sở trực giao của $\mathbb{R}^3$.

b) Dùng giải thuật Gram-Schmidt, xây dựng hệ trực chuẩn $\{v_1, v_2, v_3\}$ từ hệ S.

c) Tìm tọa độ của $u = (1,2,4)$ đối với cơ sở $\{v_1, v_2, v_3\}$.

13) Cho hệ các vectơ $S = \{u_1 = (1,1,1), u_2 = (0,1,0), u_3 = (0,0,1)\}$.

a) Kiểm chứng rằng S là hệ độc lập tuyến tính.

b) Dùng giải thuật Gram-Schmidt, xây dựng hệ trực giao $\{v_1, v_2, v_3\}$, và hệ trực chuẩn $\{w_1, w_2, w_3\}$ từ hệ S.

14) Kiểm chứng tính độc lập tuyến tính, xây dựng hệ trực giao và hệ trực chuẩn của các hệ sau

a) $S = \{u_1 = (1,-3), u_2 = (2,2)\}$.

b) $S = \{u_1 = (1,0,0), u_2 = (3,7,-2), u_3 = (0,4,1)\}$.

c) $S = \{u_1 = (0,2,1,0), u_2 = (1,-1,0,0), u_3 = (1,2,0,-1), u_4 = (1,0,0,1)\}$.

15) Chéo hóa trực giao các ma trận sau

a) $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$.

b) $A = \begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix}$.

c) $A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$.

d) $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

16) Phân tích giá trị kỳ dị các ma trận sau

a) $A = \begin{bmatrix} 1 & 2 \\ 2 & 0 \\ 0 & 2 \end{bmatrix}$.

b) $A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \\ 1 & 0 \end{bmatrix}$.

c) $A = \begin{bmatrix} -1 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix}$.

d) $A = \begin{bmatrix} 2 & 2 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}$.

e) $A = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 2 & 0 & 1 & 1 \end{bmatrix}$.

Chương 2

TỐI ƯU LỒI

Cho $f: \Omega \to \mathbb{R}$ là một hàm **trơn** (smooth) xác định trên một tập mở không rỗng Ω của $\mathbb{R}^n$. Trong phần này, ta xét bài toán tìm $x_0 \in \Omega$ sao cho $f(x_0) = \max_{x \in \Omega} f(x)$ hay $f(x_0) = \min_{x \in \Omega} f(x)$, ký hiệu

$$x_0 = \underset{x \in \Omega}{argmax} f(x) \text{ hay } x_0 = \underset{x \in \Omega}{argmin} f(x).$$

Khi đó, x_0 được gọi là một điểm **cực đại toàn cục** (global maximum) hay **cực tiểu toàn cục** (global minimum) của f trên Ω. Với hàm khả vi f bất kỳ, ta có điều kiện cần cho x_0,

$$\nabla f(x_0) = 0,$$

và khi f là hàm lồi/lõm, nó cũng là điều kiện đủ mà ta sẽ khảo sát trong phần còn lại.

2.1. HÀM MỘT BIẾN

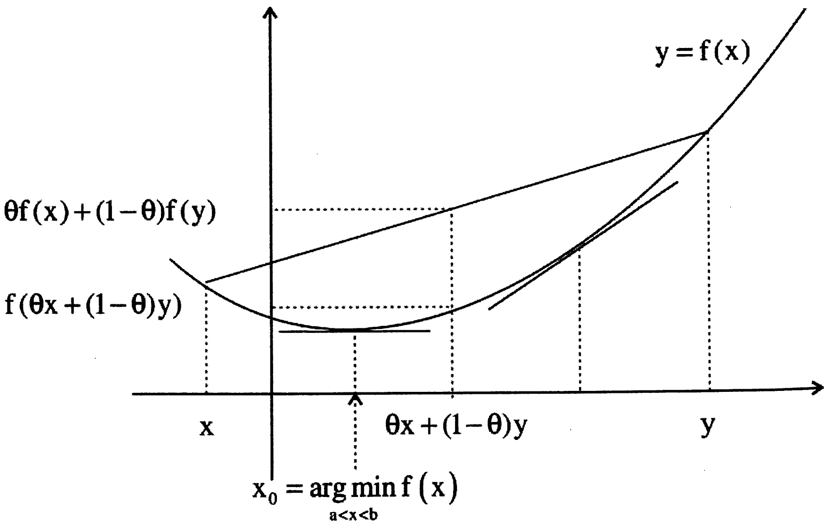
Với $\Omega = (a, b) \subset \mathbb{R}$, $-\infty \leq a < b \leq \infty$, ta có

2.1.1. Định nghĩa. Hàm số $f: (a, b) \to \mathbb{R}$ được gọi là **lồi** (convex) khi với mọi $x, y \in (a, b)$ và $0 \leq \theta \leq 1$, ta có

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y). \tag{1}$$

Khi $-f$ là hàm lồi, ta nói f là hàm **lõm** (concave).

Biểu diễn trong hệ trục tọa độ Oxy, bất đẳng thức (1) cho thấy đoạn thẳng nối $(x, f(x))$ và $(y, f(y))$, còn gọi là **dây cung** (chord) nối hai điểm, nằm phía bên trên đồ thị của f.



2.1.2. Định lý. Cho $f: (a, b) \to \mathbb{R}$, với $-\infty \leq a < b \leq \infty$, là một hàm thuộc lớp C^2, nghĩa là đạo hàm bậc hai f'' tồn tại và là hàm liên tục. Ta có các phát biểu tương đương sau

(i) f là hàm lồi.

(ii) $f(y) \geq f(x) + f'(x)(y - x)$, với mọi $x, y \in (a, b)$.

(iii) $f''(x) \geq 0$, với mọi $x \in (a, b)$.

Để chứng minh, ta cần dùng công thức Taylor đến cấp 2 cho hàm một biến sau.

2.1.3. Mệnh đề. Cho $f: (a, b) \to \mathbb{R}$ là hàm thuộc lớp C^2. Với mọi $x, y \in (a, b)$, $x < y$, tồn tại $x \leq z \leq y$ sao cho

$$f(y) = f(x) + f'(x)(y - x) + f''(z)\frac{(y-x)^2}{2!}.$$

Chứng minh định lý 2.1.2. Ta sẽ chứng minh $(i) \Leftrightarrow (ii)$ và $(ii) \Leftrightarrow (iii)$. Khi (i) thỏa, với $x, y \in (a, b)$ và $0 \leq \theta \leq 1$, (1) cho

$$f\big(y + \theta(x - y)\big) \leq f(y) + \theta\big(f(x) - f(y)\big).$$

Do đó, với $0 < \theta \leq 1$, vì

$$f(x) - f(y) \geq \frac{f\big(y + \theta(x - y)\big) - f(y)}{\theta} \to f'(y)(x - y)$$

khi $\theta \downarrow 0$, phát biểu (ii) thỏa. Ngược lại, khi (ii) đúng, với $z = \theta x + (1 - \theta)y$, ta có

$$f(x) \geq f(z) + f'(z)(x - z), \tag{2}$$

và

$$f(y) \geq f(z) + f'(z)(y - z). \tag{3}$$

Nhân hai vế của (2) với θ, (3) với $1 - \theta$, và cộng lại, vế theo vế, ta được

$$\theta f(x) + (1 - \theta)f(y) \geq f(z) = f(\theta x + (1 - \theta)y),$$

và (i) được chứng minh.

Bây giờ, khi (ii) đúng, với $x, y \in (a, b)$, $x < y$, ta có

$f(y) \geq f(x) + f'(x)(y-x)$ và $f(x) \geq f(y) + f'(y)(x-y)$.

Do đó

$$f'(x)(y-x) \leq f(y) - f(x) \leq f'(y)(y-x).$$

Chia hai vế cho $(y-x)^2$, ta được

$$\frac{f'(y)-f'(x)}{y-x} \geq 0,$$

với mọi $x, y \in (a,b)$, $x < y$. Cho $y \to x$, ta được (iii). Ngược lại, do Mệnh đề 2.1.3, tồn tại $z \in [x,y]$ sao cho

$$f(y) = f(x) + f'(x)(y-x) + \frac{1}{2}f''(z)(y-x)^2.$$

Do $f''(z) \geq 0$, ta được $f(y) \geq f(x) + f'(x)(y-x)$, i.e., (ii) thỏa.

Áp dụng Định lý 2.1.2 cho $-f$ khi f là hàm lõm, ta có

2.1.4. Hệ quả. Cho $f\colon (a,b) \to \mathbb{R}$, với $-\infty \leq a < b \leq \infty$, là hàm thuộc lớp C^2. Các phát biểu sau là tương đương

(i) f là hàm lõm.

(ii) $f(y) \leq f(x) + f'(x)(y-x)$, với mọi $x, y \in (a,b)$.

(iii) $f''(x) \leq 0$, với mọi $x \in (a,b)$.

Ví dụ 1. Các hàm $y = x^n$, với $n \in \mathbb{N}$ chẵn, và $y = e^x$ là các hàm lồi. Hàm $y = ln\,|x|$ là hàm lõm.

Xét hàm lồi (hay lõm) f trên (a,b) và xét điểm $x_0 \in (a,b)$ sao cho $f'(x_0) = 0$. Bất đẳng thức (ii) của Định lý 2.1.2 (hay

Hệ quả 2.1.4), với $x = x_0$, cho $f(x) \geq f(x_0)$ (hay $f(x) \leq f(x_0)$), với mọi $x \in (a, b)$. Điều này cho thấy x_0 là điểm cực tiểu toàn cục (hay cực đại toàn cục) của f trên (a, b), nghĩa là, $f'(x_0) = 0$ cũng là điều kiện đủ cho cực trị toàn cục.

2.1.5. Hệ quả. Xét $f: (a, b) \to \mathbb{R}$, với $-\infty \leq a < b \leq \infty$, là hàm thuộc lớp C^1, nghĩa là đạo hàm bậc nhất f' tồn tại và là hàm liên tục, và xét điểm $x_0 \in (a, b)$ sao cho $f'(x_0) = 0$.

(i) Khi f là hàm lồi, ta có $x_0 = \underset{x \in (a,b)}{argmin} f(x)$.

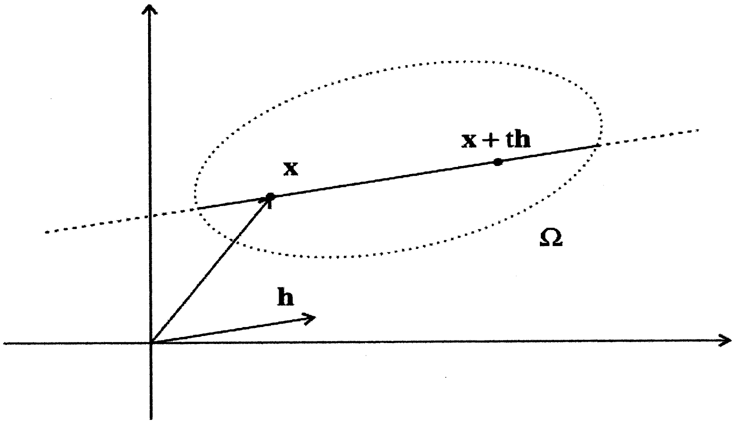
(ii) Khi f là hàm lõm, ta có $x_0 = \underset{x \in (a,b)}{argmax} f(x)$.

Ví dụ 2. Hàm $y = x \ln x$, xác định trên $(0, +\infty)$, có $y' = \ln x + 1$, $y'' = \frac{1}{x}$. Do $y'' > 0$, với mọi $x \in (0, +\infty)$, $y = x \ln x$ là hàm lồi trên $(0, +\infty)$. Do $y' = 0 \Leftrightarrow x = e^{-1}$, ta suy ra $\underset{x \in (0,+\infty)}{argmax} (x \ln x) = e^{-1}$.

2.2. HÀM NHIỀU BIẾN

Xét hàm $f: \Omega \to \mathbb{R}$, trong đó Ω là một tập con không rỗng của $\mathbb{R}^n$. Trước hết, ta cần một số ký hiệu và thuật ngữ.

2.2.1. Định nghĩa. Một tập con không rỗng $\Omega \subset \mathbb{R}^n$ được gọi là một tập **mở lồi** trong $\mathbb{R}^n$ khi với mọi $x \in \Omega$ và $h \in \mathbb{R}^n$, tập $\{t \in \mathbb{R} | x + th \in \Omega\}$ là một khoảng mở (a, b) trong $\mathbb{R}$, với $-\infty \leq a < b \leq \infty$.



Ví dụ 3. Mọi tập mở lồi Ω trong $\mathbb{R}$ đều là một khoảng mở. Bản thân $\mathbb{R}^n$ là một tập con mở lồi của chính nó. Nửa không gian trên $\mathbb{R} \times (0, \infty) = \{(x, y) \in \mathbb{R}^2 | y > 0\}$ và góc phần tư thứ nhất $(0, \infty) \times (0, \infty) = \{(x, y) \in \mathbb{R}^2 | x, y > 0\}$ là hai tập mở lồi trong $\mathbb{R}^2$.

2.2.2. Định nghĩa. Ma trận vuông $A \in \mathbb{R}^{n \times n}$ được gọi là **nửa xác định dương** (positive semidefinite), ký hiệu $A \geq 0$, khi $h^T A h \geq 0$, với mọi $h \in \mathbb{R}^n$. Khi $h^T A h \leq 0$, với mọi $h \in \mathbb{R}^n$, ta nói A là ma trận **nửa xác định âm** (negative semidefinite), ký hiệu $A \leq 0$.

Ví dụ 4. Xét ma trận

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}.$$

Ta có, với mọi $\boldsymbol{h} = (h, k) \in \mathbb{R}^2$,

$$\boldsymbol{h}^T A \boldsymbol{h} = [h \quad k] \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix} = [h \quad k] \begin{bmatrix} h + 2k \\ 2h + 5k \end{bmatrix} =$$
$$h(h + 2k) + k(2h + 5k) = h^2 + 4hk + 5k^2.$$

Bằng cách viết $h^2 + 4hk = (h + 2k)^2 - 4k^2$, vế phải trở thành

$$\begin{aligned} h^2 + 4hk + 5k^2 &= (h + 2k)^2 - 4k^2 + 5k^2 \\ &= (h + 2k)^2 + k^2 \geq 0, \end{aligned}$$

với mọi $\boldsymbol{h} = (h, k) \in \mathbb{R}^2$. Do vậy, A là ma trận nửa xác định dương.

Cho $A \in \mathbb{R}^{n \times n}$ là ma trận đối xứng với các trị riêng λ_i, $i = 1, \cdots, n$. Ta có

2.2.3. Định lý. Ma trận đối xứng A là nửa xác định dương khi $\lambda_i \geq 0$, với mọi $i = 1, \cdots, n$, và là ma trận nửa xác định âm khi $\lambda_i \leq 0$, với mọi $i = 1, \cdots, n$.

Chứng minh. Gọi Q là ma trận chéo hóa trực giao A, $Q^T A Q = D$ với $D = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$. Do $Q^T = Q^{-1}$, $QDQ^T = A$, và với mọi $h \in \mathbb{R}^n$, ta có

$$h^T A h = h^T (QDQ^T) h = (Q^T h)^T D (Q^T h).$$

Đặt $Q^T h = (y_1, y_2, \cdots, y_n) \in \mathbb{R}^n$, vế phải được viết lại thành

$$\begin{aligned} (Q^T h)^T D (Q^T h) &= [y_1 \quad y_2 \quad \cdots \quad y_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ &= \sum_{i=1}^{n} \lambda_i y_i^2 \end{aligned}$$

và điều này cho thấy $h^TAh \geq 0$ khi $\lambda_i \geq 0$, với mọi $i = 1, \cdots, n$, và $h^TAh \leq 0$ khi $\lambda_i \leq 0$, với mọi $i = 1, \cdots, n$.

Ví dụ 5. Phương trình đặc trưng của A, cho trong Ví dụ 4,

$$\begin{aligned} det(A - \lambda I_2) &= \begin{vmatrix} 1-\lambda & 2 \\ 2 & 5-\lambda \end{vmatrix} = (1-\lambda)(5-\lambda) - 4 \\ &= \lambda^2 - 6\lambda + 5 = (\lambda - 1)(\lambda - 5) = 0 \end{aligned}$$

có hai nghiệm dương, 1 và 5. Do đó, A là ma trận nửa xác định dương.

2.2.4. Định nghĩa. Cho $f: \Omega \to \mathbb{R}$, với Ω là một tập mở không rỗng của $\mathbb{R}^n$.

(i) f được gọi là một **hàm lồi** khi với mọi $x, y \in \Omega$ và $0 \leq \theta \leq 1$,

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y).$$

(ii) f được gọi là một **hàm lõm** khi $-f$ là một hàm lồi, nghĩa là

$$f(\theta x + (1-\theta)y) \geq \theta f(x) + (1-\theta)f(y).$$

Với hàm $f: \Omega \to \mathbb{R}$ xác định trên một tập mở lồi không rỗng Ω của $\mathbb{R}^n$, ứng với mỗi $x \in \Omega$ và $h \in \mathbb{R}^n$, xét tập $I = \{t \in \mathbb{R} | x + th \in \Omega\}$, là một khoảng mở trong $\mathbb{R}$, và xét hàm $g_{x,h}: I \to \mathbb{R}$ xác định bởi

$$g_{x,h}(t) = f(x + th), \text{ với } t \in I. \tag{4}$$

Ta nhận được mối liên hệ quan trọng giữa hàm nhiều biến f và họ các hàm một biến $g_{x,h}$ như sau.

2.2.5 Mệnh đề. Cho $f\colon \Omega \to \mathbb{R}$ là hàm xác định trên một tập mở lồi Ω trong $\mathbb{R}^n$ và gọi $\{g_{x,h}\}$ là họ các hàm một biến, xác định bởi (4). Ta có

f là hàm lồi nếu và chỉ nếu $g_{x,h}$ là các hàm lồi, với mọi $x \in \Omega$ và $h \in \mathbb{R}^n$,

và

f là hàm lõm nếu và chỉ nếu $g_{x,h}$ là các hàm lõm, với mọi $x \in \Omega$ và $h \in \mathbb{R}^n$.

Cho $f\colon \mathbb{R}^n \to \mathbb{R}$ là hàm thuộc lớp C^2. **Vectơ gradient** và **ma trận Hesse** của f tại $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$, ký hiệu lần lượt là $\nabla f(x)$ và $\nabla^2 f(x)$, xác định bởi

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix} \text{ và}$$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}.$$

Khi f thuộc lớp C^2, ta có $\frac{\partial^2 f}{\partial x_j \partial x_i}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$, với mọi $i, j = 1, \cdots, n$, và do đó $\nabla^2 f(x)$ là một ma trận đối xứng, $\nabla^2 f(x)^T = \nabla^2 f(x)$.

Ví dụ 6. (a) Xét $f: \mathbb{R}^3 \to \mathbb{R}$ xác định bởi $f(x_1, x_2, x_3) = 2x_1 - x_2 + 3x_3$. Ta có

$$\nabla f(x_1, x_2, x_3) = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \text{ và } \nabla^2 f(x_1, x_2, x_3) = \mathbf{0},$$

với **0** là **ma trận không** (zero matrix) trong $\mathbb{R}^{3\times 3}$.

(b) Xét $f: \mathbb{R}^3 \to \mathbb{R}$ xác định bởi $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - x_1x_2 + x_1x_3$. Ta có

$$\nabla f(x_1, x_2, x_3) = \begin{bmatrix} 2x_1 - x_2 + x_3 \\ 2x_2 - x_1 \\ 2x_3 + x_1 \end{bmatrix}$$

$$\text{và } \nabla^2 f(x_1, x_2, x_3) = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}.$$

2.2.6. Mệnh đề. Cho $f: \Omega \to \mathbb{R}$ thuộc lớp C^2, trong đó Ω là một tập mở lồi không rỗng của $\mathbb{R}^n$. Ứng với mỗi $x \in \Omega$ và $h = (h_1, h_2, \cdots, h_n) \in \mathbb{R}^n$, đặt $I = \{t \in \mathbb{R} | x + th \in \Omega\}$ và xét hàm $g_{x,h}: I \to \mathbb{R}$ xác định bởi (4). Ta có

$$g'_{x,h}(t) = \nabla f(x + th)h \text{ và } g''_{x,h}(t) = h^T \nabla^2 f(x + th)h.$$

Chứng minh. Do f khả vi, tồn tại hàm $\varepsilon(h)$, xác định với $||h|| < r$, sao cho $\varepsilon(h) \to 0$ khi $h \to 0$, và

$$f(x + h) - f(x) = \nabla f(x)h + ||h||\varepsilon(h).$$

Cố định $t \in I$. Với mọi $k \in \mathbb{R}$ sao cho $t + k \in I$, ta có

$$\begin{aligned} g_{x,h}(t + k) - g_{x,h}(t) &= f(x + (t + k)h) - f(x + th) \\ &= \nabla f(x + th)(kh) + ||kh||\varepsilon(kh) \\ &= k\nabla f(x + th)h + |k|||h||\varepsilon(kh) \end{aligned}$$

và do đó

$$\frac{g_{x,h}(t+k)-g_{x,h}(t)}{k} = \nabla f(x+th)h + \frac{|k|}{k}\|h\|\varepsilon(kh).$$

Cho $k \to 0$, ta có $\|kh\| \to 0$, nên

$$\left|\frac{|k|}{k}\|h\|\varepsilon(kh)\right| = \|h\||\varepsilon(kh)| \to 0.$$

Suy ra

$$g'_{x,h}(t) = \lim_{k\to 0}\frac{g_{x,h}(t+k)-g_{x,h}(t)}{k} = \nabla f(x+th)h$$
$$= \textstyle\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(x+th)h_i.$$

Bây giờ, với mỗi $i = 1,2,\cdots,n$, xét hàm $g_i(t) = \frac{\partial f}{\partial x_i}(x+th)$ và áp dụng kết quả vừa nhận được cho $\frac{\partial f}{\partial x_i}$ thay vì cho f, ta được

$$g'_i(t) = \sum_{j=1}^{n}\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x+th)h_j = \sum_{j=1}^{n}\frac{\partial^2}{\partial x_j\partial x_i}(x+th)h_j$$

và do đó

$$g''_{x,h}(t) = \sum_{i=1}^{n}h_i g'_i(t) = \sum_{i=1}^{n}h_i\sum_{j=1}^{n}\frac{\partial^2}{\partial x_j\partial x_i}(x+th)h_j$$

$$= [h_1 \quad h_2 \quad \cdots \quad h_n]\begin{bmatrix} \sum_{j=1}^{n}\frac{\partial^2}{\partial x_j\partial x_1}(x+th)h_j \\ \sum_{j=1}^{n}\frac{\partial^2}{\partial x_j\partial x_2}(x+th)h_j \\ \vdots \\ \sum_{j=1}^{n}\frac{\partial^2}{\partial x_j\partial x_n}(x+th)h_j \end{bmatrix}$$

$$= \begin{bmatrix} h_1 & h_2 & \cdots & h_n \end{bmatrix} \begin{bmatrix} \frac{\partial^2}{\partial x_1^2}(x+th) & \frac{\partial^2}{\partial x_2 \partial x_1}(x+th) & \cdots & \frac{\partial^2}{\partial x_n \partial x_1}(x+th) \\ \frac{\partial^2}{\partial x_1 \partial x_2}(x+th) & \frac{\partial^2}{\partial x_2^2}(x+th) & \cdots & \frac{\partial^2}{\partial x_n \partial x_2}(x+th) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2}{\partial x_1 \partial x_n}(x+th) & \frac{\partial^2}{\partial x_2 \partial x_n}(x+th) & \cdots & \frac{\partial^2}{\partial x_n^2}(x+th) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix}$$
$$= h^T \nabla^2 f(x+th)h.$$

Mệnh đề được chứng minh.

Ta nhận được phiên bản nhiều chiều cho Định lý 2.1.2 như sau.

2.2.7. Định lý. Cho $f: \Omega \to \mathbb{R}$ là hàm thuộc lớp C^2 xác định trên một tập mở lồi Ω của $\mathbb{R}^n$. Các phát biểu sau là tương đương.

(i) f là hàm lồi.

(ii) $f(y) \geq f(x) + \nabla f(x)^T(y-x)$, với mọi $x, y \in \Omega$.

(iii) $\nabla^2 f(x) \geq 0$, với mọi $x \in \Omega$.

Chứng minh. Với mỗi $z \in \Omega$ và $h \in \mathbb{R}^n$, xét $I = \{t \in \mathbb{R} | z + th \in \Omega\}$, và xét hàm $g_{z,h}: I \to \mathbb{R}$ xác định bởi $g_{z,h}(t) = f(z+th)$, với mọi $t \in I$. Do Mệnh đề 2.2.5, (i) thỏa khi và chỉ khi mọi $g_{z,h}$ đều là các hàm lồi. Do Định lý 2.1.2, điều này tương đương với hai phát biểu sau

(a) $g_{z,h}(s) \geq g_{z,h}(t) + g'_{z,h}(t)(s-t)$, với mọi $s, t \in I$.

(b) $g''_{z,h}(t) \geq 0$, với mọi $t \in I$.

Do Mệnh đề 2.2.6,

$$g'_{z,h}(t) = \nabla f(z+th)h \text{ và } g''_{z,h}(t) = h^T \nabla^2 f(z+th)h.$$

Với $z = x$, $h = y - x$, $s = 1$, và $t = 0$ trong (a), ta được

$$f(y) = g_{x,h}(1) \geq g_{z,h}(0) + g'_{z,h}(0) = f(x) + \nabla f(x)h,$$

và (ii) thỏa. Khi $z = x$, do

$$h^T \nabla^2 f(x) h = g''_{x,h}(0) \geq 0, \text{ với mọi } h \in \mathbb{R}^n,$$

ta suy ra $\nabla^2 f(x) \gtrsim 0$, nghĩa là, (iii) được chứng minh. Ngược lại, xét các hàm $g_{z,h}(t) = f(z + th)$, $z \in \Omega$ và $h \in \mathbb{R}^n$. Với mỗi $s, t \in I$, chọn $x = z + th$ và $y = z + sh$, ta được

$$g_{z,h}(s) = f(z + sh) \geq f(z + th) + \nabla f(z + th)^T\big((s - t)h\big) = g_{z,h}(t) + g'_{z,h}(t)(s - t),$$

khi (ii) thỏa, và $g''_{z,h}(t) = h^T \nabla^2 f(z + th) h \geq 0$ khi (iii) thỏa. Do đó, nếu (ii) hay (iii) thỏa, thì do Định lý 2.1.2, $g_{z,h}$ là một hàm lồi và định lý được chứng minh.

Điều kiện (iii) được dùng để kiểm chứng tính lồi của hàm f và khi f là hàm lồi, điều kiện (ii) cho thấy rằng nghiệm của phương trình $\nabla f(x) = 0$ chính là điểm cực tiểu toàn cục của f trên Ω.

Tương tự như cho trường hợp hàm một biến, bằng cách thay f bằng $-f$ trong Định lý 2.2.7, ta được kết quả sau cho hàm lõm.

2.2.8. Hệ quả. Cho $f: \Omega \to \mathbb{R}$ là hàm thuộc lớp C^2 xác định trên một tập mở lồi Ω của $\mathbb{R}^n$. Các phát biểu sau là tương đương.

(i) f là hàm lõm.

(ii) $f(y) \leq f(x) + \nabla f(x)^T(y - x)$, với mọi $x, y \in \Omega$.

(iii) $\nabla^2 f(x) \leq 0$, với mọi $x \in \Omega$.

Ta tổng hợp các kết quả nêu trên để tiện sử dụng trong các ứng dụng sau này.

2.2.9. Hệ quả. Cho Ω là một tập mở lồi của $\mathbb{R}^n$. Xét hàm $f: \Omega \to \mathbb{R}$ thuộc lớp C^1, nghĩa là mọi đạo hàm riêng của f tồn tại và là hàm liên tục, và xét điểm $x_0 \in \Omega$ thỏa $\nabla f(x_0) = 0$.

(i) Nếu f là hàm lồi, thì $x_0 = \underset{x \in \Omega}{argmin} f(x)$.

(ii) Nếu f là hàm lõm, thì $x_0 = \underset{x \in \Omega}{argmax} f(x)$.

Ví dụ 7. Trở lại ví dụ 6 (b), ta có

$f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - x_1x_2 + x_1x_3$,
$(x_1, x_2, x_3) \in \mathbb{R}^3$,

$$\nabla f(x_1, x_2, x_3) = \begin{bmatrix} 2x_1 - x_2 + x_3 \\ 2x_2 - x_1 \\ 2x_3 + x_1 \end{bmatrix} \text{ và } \nabla^2 f(x_1, x_2, x_3) = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}.$$

Do $\nabla^2 f(x_1, x_2, x_3)$ có các trị riêng không âm 2, $2 + \sqrt{2}$, và $2 - \sqrt{2}$, ta suy ra $\nabla^2 f(x_1, x_2, x_3) \geq 0$. Suy ra f là hàm lồi và do đó nghiệm (duy nhất) của phương trình $\nabla f(x_1, x_2, x_3) = (0,0,0)$ là điểm cực tiểu toàn cục của f trên $\mathbb{R}^3$.

2.3. DẠNG TOÀN PHƯƠNG

Tổng quát hóa các hàm một biến, hàm tuyến tính $f(x) = ax$ và hàm bậc hai $f(x) = ax^2$, cho hàm nhiều biến, ta có

2.3.1. Định nghĩa. Cho vectơ $q = (q_1, q_2, \cdots, q_n) \in \mathbb{R}^n$ và ma trận $A = [a_{ij}] \in \mathbb{R}^{n \times n}$.

(i) Hàm $f: \mathbb{R}^n \to \mathbb{R}$, xác định bởi

$$f_q(x) = \langle q, x \rangle = \sum_{i=1}^n q_i x_i \equiv q^T x,$$

với $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$, được gọi là một **dạng tuyến tính** (linear form).

(ii) Hàm $f: \mathbb{R}^n \to \mathbb{R}$, xác định bởi

$$f_A(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = x^T A x,$$

với $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$, được gọi là một **dạng toàn phương** (quadratic form).

Ví dụ 8. (a) Hàm $f(x_1, x_2, x_3) = 2x_1 - x_2 + 3x_3$ trong Ví dụ 6 (a) là một dạng tuyến tính, $f = f_q$, với $q = (2, -1, 3)$.

(b) Hàm $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - x_1 x_2 + x_1 x_3$ trong Ví dụ 6 (b) là một dạng toàn phương, $f = f_A$, với

$$A = \begin{bmatrix} 1 & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 0 & 1 \end{bmatrix}.$$ ■

Ghi chú. Với dạng toàn phương $f_A(x) = x^T A x$ trên $\mathbb{R}^n$, đặt $B = \frac{1}{2}(A + A^T)$. Ta có B là một ma trận đối xứng và

$$x^T B x = \frac{1}{2}(x^T A x + x^T A^T x) = x^T A x,$$

do $x^T A^T x = (x^T A^T x)^T = x^T (A^T)^T x = x^T A x \in \mathbb{R}^{1\times 1}$. Vì vậy, ta có thể giả sử ma trận A xác định dạng toàn phương $f_A(x) = x^T A x$ là một ma trận đối xứng.

2.3.2. Mệnh đề. Cho $f_q(x)$ là một dạng tuyến tính, với $q = (q_1, q_2, \cdots, q_n) \in \mathbb{R}^n$, và cho $f_A(x) = x^T A x$ là một dạng toàn phương xác định bởi ma trận đối xứng $A = [a_{ij}] \in \mathbb{R}^{n\times n}$. Ta có

(i) $\nabla f_q(x) = q$ và $\nabla^2 f_q(x) = 0$, với mọi $x \in \mathbb{R}^n$.

(ii) $\nabla f_A(x) = 2Ax$ và $\nabla^2 f_A(x) = 2A$, với mọi $x \in \mathbb{R}^n$.

Chứng minh. (i) Đẳng thức $f_q(x) = q_1x_1 + \cdots + q_ix_i + \cdots + q_nx_n$ cho $\frac{\partial f_q}{\partial x_i}(x) = q_i$, với mọi $i = 1,2,\cdots,n$, và $\frac{\partial^2 f_q}{\partial x_j \partial x_i}(x) = 0$, với mọi $i,j = 1,2,\cdots,n$.

(ii) Do $\frac{\partial}{\partial x_k}(x_ix_j) = 0$, khi $k \neq i,j$, $\frac{\partial}{\partial x_k}(x_kx_j) = x_j$, khi $k = i \neq j$, và $\frac{\partial}{\partial x_k}(x_kx_k) = 2x_k$, nên từ đẳng thức $f_A(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_ix_j$, ta được, với mỗi $k = 1,2,\cdots,n$,

$$\frac{\partial f_A}{\partial x_k}(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}\frac{\partial}{\partial x_k}(x_ix_j)$$

$$= \sum_{i=1}^n \left(a_{ik}\frac{\partial}{\partial x_k}(x_ix_k) + \sum_{j\neq k} a_{ij}\frac{\partial}{\partial x_k}(x_ix_j) \right)$$

$$= \sum_{i=1}^n a_{ik}\frac{\partial}{\partial x_k}(x_ix_k)$$

$$+ \sum_{i=1}^n \left(\sum_{j\neq k} a_{ij}\frac{\partial}{\partial x_k}(x_ix_j) \right)$$

$$= 2a_{kk}x_k + \sum_{i \neq k} a_{ik}x_i + \sum_{j \neq k}\sum_{i=1}^{n} a_{ij}\frac{\partial}{\partial x_k}(x_i x_j)$$

$$= 2a_{kk}x_k + \sum_{i \neq k} a_{ik}x_i + \sum_{j \neq k} a_{kj}x_j$$

$$= 2\sum_{i=1}^{n} a_{ki}x_i \qquad (5)$$

chính là dòng thứ k của ma trận $2Ax$. Suy ra $\nabla f_A(x) = 2Ax$.

Bây giờ, do (5), $\frac{\partial f_A}{\partial x_k}(x)$ là một dạng tuyến tính, với $q = (2a_{k1}, 2a_{k2}, \cdots, 2a_{kn})$, dòng thứ k của ma trận $2A$, nên áp dụng (i), ta được $\frac{\partial^2 f_A}{\partial x_l \partial x_k}(x) = 2a_{lk}$, với mọi $l, k = 1,2,\cdots,n$, nghĩa là $\nabla^2 f_A(x) = 2A$và mệnh đề được chứng minh.

Cuối cùng, cho ma trận $A \in \mathbb{R}^{N\times n}$, với $n \leq N$, $rank(A) = n$, và cho vectơ $b \in \mathbb{R}^N$. Xét hàm số $f: \mathbb{R}^n \to \mathbb{R}$ xác định bởi $f(x) = \|Ax - b\|^2$. Ta có

$$\begin{aligned} f(x) &= \langle Ax - b, Ax - b\rangle = (Ax - b)^T(Ax - b) \\ &= (x^T A^T - b^T)(Ax - b) \\ &= x^T A^T Ax - b^T Ax - x^T A^T b + b^T \\ &= x^T(A^T A)x - (2b^T A)x + b^T b, \end{aligned}$$

do $b^T Ax = x^T A^T b$. Áp dụng Mệnh đề 2.3.2, ta được

$$\nabla f(x) = 2(A^T A)x - 2A^T b \text{ và } \nabla^2 f(x) = 2(A^T A).$$

Vì $h^T \nabla^2 f(x)h = 2[h^T(A^T A)h] = 2[(h^T A^T)(Ah)] = 2\|Ah\|^2 \geq 0$, với mọi $h \in \mathbb{R}^N$, nên ma trận Hesse $\nabla^2 f(x)$ là nửa xác định dương. Hơn nữa, vì $rank(A^T A) = n$ nên $A^T A$ là ma trận khả nghịch. Do đó, nghiệm $x = (A^T A)^{-1}A^T b$ của phương trình $\nabla f(x) = 0$ là điểm cực tiểu toàn cục của hàm f.

Ta được

2.3.3. Định lý. Cho ma trận $A \in \mathbb{R}^{N\times n}$, với $n \le N$, sao cho $rank(A) = n = min(N, n)$, và cho vectơ $b \in \mathbb{R}^N$. Ta có

$$(A^TA)^{-1}A^Tb = \underset{x\in\mathbb{R}^n}{argmin}\|Ax - b\|^2. \qquad \blacksquare$$

Ví dụ 9. Cho vectơ $b = (2,2,5,8) \in \mathbb{R}^4$ và cho ma trận

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \in \mathbb{R}^{4\times 2}.$$

Dễ thấy $dim\big(col(A)\big) = 2 = rank(A) = min(4,2)$, $A^TA = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}$ là ma trận khả nghịch, và $(A^TA)^{-1}A^Tb = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}^{-1}\begin{bmatrix} 17 \\ 53 \end{bmatrix} = \begin{bmatrix} -1 \\ 2.1 \end{bmatrix}$. Do đó

$$(-1,2.1) = \underset{x\in\mathbb{R}^2}{argmin}\|Ax - b\|^2.$$

Với

$$A = \begin{bmatrix} 1 & 1 & 1^2 \\ 1 & 2 & 2^2 \\ 1 & 3 & 3^2 \\ 1 & 4 & 4^2 \end{bmatrix} \in \mathbb{R}^{4\times 3},$$

ta có $dim\big(col(A)\big) = 3 = rank(A) = min(4,3)$, $A^TA = \begin{bmatrix} 4 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix}$ là ma trận khả nghịch, và $(A^TA)^{-1}A^Tb = \begin{bmatrix} 4 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{bmatrix}^{-1}\begin{bmatrix} 17 \\ 53 \\ 183 \end{bmatrix} = \begin{bmatrix} 2.75 \\ -1.65 \\ 0.75 \end{bmatrix}$.

Do đó $(2.75, -1.65, 0.75) = \underset{x\in\mathbb{R}^3}{argmin}\|Ax - b\|^2.$

BÀI TẬP

1) Khảo sát tính lồi/lõm của hàm số $f: O \subset \mathbb{R} \to \mathbb{R}$ xác định bởi

a) $f(x) = e^{ax}, O = \mathbb{R}$.

b) $f(x) = -\log(x), O = (0, \infty)$.

c) $f(x) = x^a, a \geq 0, a \leq 1, O = (0, \infty)$.

d) $f(x) = -x^a, 0 \leq a \leq 1, O = (0, \infty)$.

e) $f(x) = x\log(x), O = (0, \infty)$.

2) Khảo sát tính lồi/lõm của hàm số $f: \mathbb{R}^2 \to \mathbb{R}$ xác định bởi

a) $f(x_1, x_2) = (x_1 - 3x_2)^2$.

b) $f(x_1, x_2) = (x_1 - 3x_2)^2 + (x_1 - 2x_2)^2$.

c) $f(x_1, x_2) = (x_1 - 3x_2)^2 + (x_1 - 2x_2)^2 + x_1^3$.

3) Khảo sát tính lồi/lõm của hàm số $f: O \subset \mathbb{R}^2 \to \mathbb{R}$ xác định bởi

a) $f(x_1, x_2) = x_1x_2$, với $O = (0, \infty) \times (0, \infty)$.

b) $f(x_1, x_2) = \frac{1}{x_1x_2}$, với $O = (0, \infty) \times (0, \infty)$.

c) $f(x_1, x_2) = \frac{x_1^2}{x_2}$, với $O = \mathbb{R} \times (0, \infty)$.

d) $f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$, với $O = (0, \infty) \times (0, \infty), 0 \leq \alpha \leq 1$.

4) Khảo sát tính lồi/lõm và tìm cực trị toàn cục, nếu có, của hàm $f: \mathbb{R}^2 \to \mathbb{R}$ xác định bởi

a) $f(x_1, x_2) = \frac{1}{2}x_1^2 + x_2^2 + x_1 - 2x_2 + 5.$

b) $f(x_1, x_2) = \frac{3}{4}x_1^2 + \frac{3}{4}x_2^2 - \frac{1}{2}x_1x_2 - 2x_1 + x_2 + 3.$

5) Khảo sát tính lồi/lõm và tìm cực trị toàn cục, nếu có, của hàm $f: \mathbb{R}^3 \to \mathbb{R}$, xác định bởi $f(x) = x^TAx + q^Tx + 1$, với $x = (x_1, x_2, x_3) \in \mathbb{R}^3$, và

a) $A = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 1 & -2 \\ 1 & -2 & 1 \end{bmatrix}, q = (1,2,-1).$

b) $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 3 \\ 0 & 2 & 1 \end{bmatrix}, q = (1,0,1).$

Chương 3

PHƯƠNG PHÁP BÌNH PHƯƠNG NHỎ NHẤT

3.1. ĐẶT VẤN ĐỀ

Một trong những bài toán căn bản trong các ngành khoa học, lý thuyết cũng như ứng dụng, là khảo sát sự liên hệ giữa các đại lượng. Chẳng hạn như sự liên hệ giữa chiều dài thanh kim loại và nhiệt độ của nó, chiều dài lò xo với khối lượng vật treo vào lò xo đó, v.v... Đặc biệt, các kết quả đạt được thông qua bộ dữ liệu quan sát hay thực nghiệm xuất hiện từ thời cổ đại, kéo dài cho tới nay, và chắc là mãi mãi!

Vào khoảng năm 1490, Leonardo da Vinci công bố bức họa Vitruvian Man miêu tả sự liên hệ giữa các đại lượng liên quan đến cơ thể người mà qua đó, các họa sĩ, các nhà điêu khắc có thể tạo ra những tác phẩm có giá trị. Chẳng hạn, bức họa cho biết chiều dài hai cánh tay dang rộng bằng với chiều cao, chiều dài bàn chân bằng với 1/6 chiều cao cơ thể, v.v... Ngày nay, các bạn trẻ còn bổ sung thêm sự liên hệ giữa vòng bụng và vòng cổ còn thiếu trong bức họa này nhằm ứng dụng trong cuộc sống hàng ngày!

Vào thế kỷ 19, Francis Galton khảo sát các đặc trưng di truyền, chẳng hạn như chiều cao, từ đời cha sang đời con và công bố trong công trình

F. Galton. *Regression towards mediocrity in hereditary stature*. The Journal of the Anthropological Institute of Great Britain and Ireland. 15: 246–263, 1886.

Thuật ngữ "Regression", xuất phát từ di truyền học, được phổ biến rộng rãi qua công trình này của Galton và được dùng để đặt tên cho một mảng thống kê suy diễn quan trọng mà ta gọi là **phân tích hồi quy** (regression analysis) trong đó người ta khảo sát sự liên hệ giữa một biến số ngẫu nhiên theo một biến số ngẫu nhiên khác (hồi quy đơn – simple regression), hay theo một số biến ngẫu nhiên khác (hồi quy bội – multiple regression).

Kỹ thuật căn bản được sử dụng là phương pháp "bình phương nhỏ nhất" – "least squares" mà Gauss đề xuất nhằm giải quyết bài toán xác định vị trí của hành tinh Ceres giới thiệu trong Lời tựa. Ngày nay, người ta gọi phương pháp này là OLS (Ordinary Least Squares) nhằm phân biệt với các phát triển nới rộng của nó.

3.2. KHỚP DỮ LIỆU (DATA FITTING)

Nhằm tìm sự liên hệ giữa một đại lượng, ký hiệu y, theo một số đại lượng khác, ký hiệu $x_1, x_2, \cdots, x_n$, tiếp cận đơn giản nhất là tìm một hàm số $f: \mathbb{R}^n \to \mathbb{R}$ sao cho

$$y \approx f(x), \tag{1}$$

trong đó $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ được gọi là vectơ các **biến độc lập** (independent variables), và $y \in \mathbb{R}$ được gọi là **biến phụ thuộc** (response variable).

Một cách tự nhiên, việc đầu tiên mà người ta cần làm là thu thập số liệu của các biến độc lập,

$$x^{(1)} = \left(x_1^{(1)}, \cdots, x_n^{(1)}\right), \cdots, x^{(N)} = \left(x_1^{(N)}, \cdots, x_n^{(N)}\right)$$

và của biến phụ thuộc tương ứng,

$$y^{(1)}, \cdots, y^{(N)}.$$

Ta nhận được bộ **dữ liệu** (data) cho các biến, có thể sắp xếp dưới dạng "ma trận" như sau

	x_1	x_2	$\cdots$	x_n	y
Số liệu (1)	$x_1^{(1)}$	$x_2^{(1)}$	$\cdots$	$x_n^{(1)}$	$y^{(1)}$
Số liệu (2)	$x_1^{(2)}$	$x_2^{(2)}$	$\cdots$	$x_n^{(2)}$	$y^{(2)}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
Số liệu (N)	$x_1^{(N)}$	$x_2^{(N)}$	$\cdots$	$x_n^{(N)}$	$y^{(N)}$

Kế tiếp, ta đưa ra một **mô hình** (model) cho mối quan hệ giữa biến phụ thuộc và các biến độc lập thông qua một hàm $\hat{f}: \mathbb{R}^n \to \mathbb{R}$ sao cho

$$y \approx \hat{f}(x),$$

và $\hat{y} = \hat{f}(x)$ được gọi là giá trị **dự báo** (prediction) của biến phụ thuộc y, khi biết giá trị của các biến độc lập $x_1, x_2, \cdots, x_n$. Hàm $\hat{f}$ được gọi là **hàm dự báo** (prediction function), hay vắn tắt là **mô hình** (model), của bài toán.

Trong phần này, ta tập trung khảo sát mô hình **tuyến tính theo tham số** (linear in the parameters model) cho lớp hàm dự báo,

$$\hat{f}(x) = \theta_1 f_1(x) + \cdots + \theta_p f_p(x), \tag{3}$$

trong đó $f_i: \mathbb{R}^n \to \mathbb{R}$ là các **hàm số cơ sở** (basis functions) được xác định trước, các θ_i là các **tham số mô hình** (model parameters) cần tìm sao cho hàm dự báo tương ứng là **phù hợp** (consistent) với dữ liệu thu thập được.

Phương pháp bình phương nhỏ nhất (Least squares)

Để đánh giá sự phù hợp của mô hình (3) với dữ liệu, ta tính các giá trị dự báo của mô hình tương ứng với giá trị của các biến độc lập của dữ liệu (cho trong (2)),

$$\hat{y}^{(i)} = \hat{f}\left(x^{(i)}\right) = \hat{f}\left(x_1^{(i)}, \cdots, x_n^{(i)}\right), \text{ với } i = 1, \cdots, N,$$

và đánh giá các **sai số dự báo** (prediction error) hay còn gọi là **phần dư** (residual),

$$r^{(i)} = y^{(i)} - \hat{y}^{(i)}, \text{ với } i = 1, \cdots, N. \tag{4}$$

Phương pháp bình phương nhỏ nhất tìm cách xác định các tham số θ_i của mô hình (3) sao cho vectơ các phần dư

$$r = \left(r^{(1)}, \cdots, r^{(N)}\right) \in \mathbb{R}^N$$

có chuẩn nhỏ nhất.

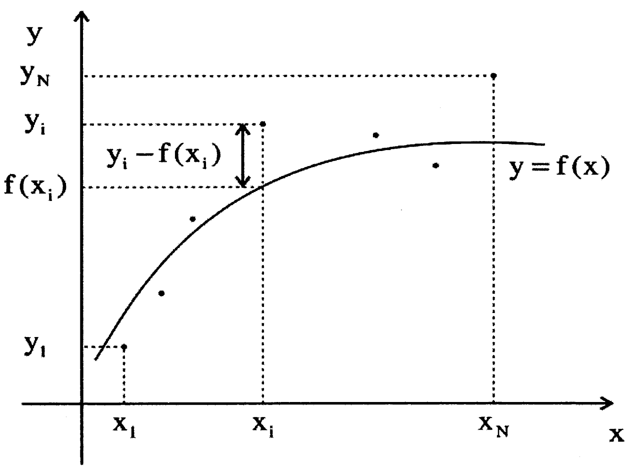
Để có thể minh họa bằng hình học, ta xét trường hợp biến y chỉ phụ thuộc vào một biến độc lập x, nghĩa là xét trường hợp $n = 1$. Với dữ liệu gồm N số liệu thu thập cho trong bảng sau

x	x_1	x_2	$\cdots$	x_N
y	y_1	y_2	$\cdots$	y_N

ta có thể biểu diễn mỗi bộ số liệu (x_i, y_i), $i = 1,2,\cdots,N$, bằng một điểm trong mặt phẳng tọa độ Oxy. Mỗi "mô hình" $y = f(x)$ được biểu diễn bằng đồ thị của nó, và mức độ "khớp" của mô hình với dữ liệu được đánh giá thông qua khoảng cách từ các điểm dữ liệu, (x_i, y_i), đến các điểm "dự báo", $(x_i, \hat{y}_i) = (x_i, f(x_i))$, của mô hình, mà ta gọi là phần dư,

$$r_i = y_i - \hat{y}_i = y_i - f(x_i),$$

với $i = 1,2,\cdots,N$.



Ví dụ 1. (i) Để khảo sát sự liên hệ giữa chiều dài lò xo, y, tính bằng cm, với khối lượng vật treo vào lò xo, x, tính bằng kg, ta thu thập các bộ số liệu (x, y) tương ứng. Chẳng hạn,

x (kg)	1	2	3	4
y (cm)	6.3	6.7	7	7.4

và với "mô hình" hàm bậc nhất, $y = f(x) = a + bx$, $a, b \in \mathbb{R}$, ta được các phần dư

r (cm)	$6.3 - (a + b \cdot 1)$	$6.7 - (a + b \cdot 2)$	$7 - (a + b \cdot 3)$	$7.4 - (a + b \cdot 5)$

(ii) Để khảo sát vật thể rơi tự do bằng cách tìm sự liên hệ giữa vị trí vật thể, x, tính bằng mét, với thời gian, t, tính bằng giây, ta dùng "mô hình" hàm bậc hai, $y = f(x) = a + bx + cx^2$, $a, b, c \in \mathbb{R}$. Ứng với mỗi bộ số liệu, ta tính được các phần dư tương ứng. Chẳng hạn, với

t (s)	0.1	0.2	0.3	0.4
x (m)	10.1	10.4	10.7	11.2

ta được các phần dư

$r_1 = 10.1 - (a + b \cdot 0.1 + c \cdot 0.1^2)$, $r_2 = 10.4 - (a + b \cdot 0.2 + c \cdot 0.2^2)$,

$r_3 = 10.7 - (a + b \cdot 0.3 + c \cdot 0.3^2)$, và $r_4 = 11.2 - (a + b \cdot 0.4 + c \cdot 0.4^2)$.

Chuẩn Euclide của vectơ các phần dư

$$RSS(f) \equiv \|r\|^2 = \sum_{i=1}^{N} r_i^2$$

được dùng để đánh giá mức độ "khớp" của hàm f đối với bộ dữ liệu, trong đó RSS chỉ "**tổng bình phương các phần dư**" (Residual Sum of Squares). Hàm nào có RSS nhỏ hơn được đánh giá là "khớp" với bộ dữ liệu hơn. Do đó, khi ta giới hạn xem xét hàm f trong một lớp L những hàm số nào đó, phương pháp OLS quy về bài toán tối ưu

Tìm $f_0 \in L$ sao cho $RSS(f_0) = \min_{f \in L} RSS(f)$.

Đặc biệt, khi L là lớp hàm sinh bởi hữu hạn các hàm cho trước f_i, $i = 1,2,\cdots,p$, mỗi hàm $f \in L$ được xác định bằng p tham số $(\theta_1, \theta_2, \cdots, \theta_p) \in \mathbb{R}^p$ sao cho $f = \theta_1 f_1 + \theta_2 f_2 + \cdots + \theta_p f_p$. Bằng cách đặt

$$RSS(\theta_1, \theta_2, \cdots, \theta_p) = RSS(f),$$

bài toán (5) trở thành

Tìm $(\theta_1^0, \theta_2^0, \cdots, \theta_p^0) \in \mathbb{R}^p$ sao cho $RSS(\theta_1^0, \theta_2^0, \cdots, \theta_p^0) = \min_{(\theta_1, \theta_2, \cdots, \theta_p) \in \mathbb{R}^p} RSS(\theta_1, \theta_2, \cdots, \theta_p)$.

Ví dụ 2. Xét bộ dữ liệu

x	x_1	x_2	$\cdots$	x_N
y	y_1	y_2	$\cdots$	y_N

(i) Với lớp L các hàm bậc nhất, $y = a + bx$, tổ hợp tuyến tính của hai đơn thức

$$f_1(x) = 1, \text{ và } f_2(x) = x,$$

nghĩa là mỗi hàm $f \in L$ được đồng nhất với bộ thứ tự $(a, b) \in \mathbb{R}^2$,

$$RSS(a,b) = \sum_{i=1}^{N}\left(y_i - (a + bx_i)\right)^2,$$

và ta nhận được bài toán tối ưu

Tìm $(a_0, b_0) \in \mathbb{R}^2$ sao cho $RSS(a_0, b_0) = \min_{(a,b)\in\mathbb{R}^2} RSS(a, b)$.

(ii) Với lớp L các hàm bậc hai, $y = a + bx + cx^2$, tổ hợp tuyến tính của ba đơn thức

$$f_1(x) = 1, f_2(x) = x, \text{ và } f_3(x) = x^2,$$

nghĩa là mỗi hàm $f \in L$ được đồng nhất với bộ thứ tự $(a, b, c) \in \mathbb{R}^3$,

$$RSS(a,b,c) = \sum_{i=1}^{N}\left(y_i - (a + bx_i + cx_i^2)\right)^2,$$

và ta nhận được bài toán tối ưu

Tìm $(a_0, b_0, c_0) \in \mathbb{R}^3$ sao cho $RSS(a_0, b_0, c_0) = \min_{(a,b,c)\in\mathbb{R}^3} RSS(a, b, c)$.

Hơn nữa, ứng với $f = \theta_1 f_1 + \theta_2 f_2 + \cdots + \theta_p f_p$ và ứng với bộ dữ liệu thứ i, (x_i, y_i), ta được phần dư thứ i,

$$r_i = y_i - f(x_i) = y_i - \left(\theta_1 f_1(x_i) + \theta_2 f_2(x_i) + \cdots + \theta_p f_p(x_i)\right),$$

$i = 1,2,\cdots,N$. Do đó

$$r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix} = \begin{bmatrix} y_1 - [\theta_1 f_1(x_1) + \theta_2 f_2(x_1) + \cdots + \theta_p f_p(x_1)] \\ y_2 - [\theta_1 f_1(x_2) + \theta_2 f_2(x_2) + \cdots + \theta_p f_p(x_2)] \\ \vdots \\ y_N - [\theta_1 f_1(x_N) + \theta_2 f_2(x_N) + \cdots + \theta_p f_p(x_N)] \end{bmatrix}$$

$$= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} - \begin{bmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_p(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_N) & f_2(x_N) & \cdots & f_p(x_N) \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix} \equiv y - X\theta,$$

với

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^N, X = \begin{bmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_p(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_N) & f_2(x_N) & \cdots & f_p(x_N) \end{bmatrix} \in \mathbb{R}^{N\times p}, \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix} \in \mathbb{R}^p.$$

Suy ra

$$RSS(\theta) = \|r\|^2 = \|X\theta - y\|^2.$$

Áp dụng Định lý 2.3.3 cho $A = X \in \mathbb{R}^{N\times p}$, với $p \leq N$, sao cho $rank(X) = p = min(N,p)$, và vectơ $b = y \in \mathbb{R}^N$, ta được

$$\theta_0 = (X^T X)^{-1} X^T y = \underset{\theta\in\mathbb{R}^p}{argmin}\|X\theta - y\|^2.$$

Ví dụ 3. Tiếp tục dữ liệu và mô hình trong Ví dụ 1. Với (i), đặt

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, y = \begin{bmatrix} 6.3 \\ 6.7 \\ 7.0 \\ 7.4 \end{bmatrix},$$

ta được

$$\theta_0 = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = (X^T X)^{-1} X^T y = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}^{-1} \begin{bmatrix} 27.4 \\ 70.3 \end{bmatrix} = \begin{bmatrix} 5.95 \\ 0.36 \end{bmatrix}.$$

Điều này có nghĩa là hàm

$$y = 5.95 + 0.36x$$

là hàm bậc nhất "khớp" với dữ liệu đã cho nhất. Với (ii), đặt

$$X = \begin{bmatrix} 1 & 0.1 & 0.1^2 \\ 1 & 0.2 & 0.2^2 \\ 1 & 0.3 & 0.3^2 \\ 1 & 0.4 & 0.4^2 \end{bmatrix}, y = \begin{bmatrix} 10.1 \\ 10.4 \\ 10.7 \\ 11.2 \end{bmatrix},$$

ta được

$$\theta_0 = \begin{bmatrix} a_0 \\ b_0 \\ c_0 \end{bmatrix} = (X^T X)^{-1} X^T y = \begin{bmatrix} 4 & 1 & 0.3 \\ 1 & 0.3 & 0.1 \\ 0.3 & 0.1 & 0.0354 \end{bmatrix}^{-1} \begin{bmatrix} 42.4 \\ 10.78 \\ 3.272 \end{bmatrix} = \begin{bmatrix} 9.95 \\ 1.1 \\ 5.0 \end{bmatrix}$$

và hàm

$$x = 9.95 + 1.1t + 5t^2$$

là hàm bậc hai "khớp" với dữ liệu đã cho nhất.

Tổng quát cho trường hợp với $n > 1$ biến độc lập, với dữ liệu cho $(x_1, x_2, \cdots, x_n, y)$,

	x_1	x_2	$\cdots$	x_n	y
Số liệu (1)	$x_1^{(1)}$	$x_2^{(1)}$	$\cdots$	$x_n^{(1)}$	$y^{(1)}$
Số liệu (2)	$x_1^{(2)}$	$x_2^{(2)}$	$\cdots$	$x_n^{(2)}$	$y^{(2)}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
Số liệu (N)	$x_1^{(N)}$	$x_2^{(N)}$	$\cdots$	$x_n^{(N)}$	$y^{(N)}$

ta tìm hàm n-biến $y = f(x_1, x_2, \cdots, x_n) \equiv f(x)$, với $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$, trong lớp L sinh bởi p hàm cơ sở, $y = f_i(x)$, $i = 1,2, \cdots, p$, nghĩa là mỗi một hàm $f \in L$ được xác định bởi vectơ $\theta = (\theta_1, \theta_2, \cdots, \theta_p) \in \mathbb{R}^p$, với

$$f(x) = \theta_1 f_1(x) + \theta_2 f_2(x) + \cdots + \theta_p f_p(x), x = (x_1, x_2, \cdots, x_n).$$

Với $\left(x^{(i)}, y^{(i)}\right) \equiv \left(x_1^{(i)}, x_2^{(i)}, \cdots, x_n^{(i)}, y^{(i)}\right)$ chỉ bộ dữ liệu thứ i, bảng số liệu nêu trên có thể viết lại thành

	x	y
Số liệu (1)	$x^{(1)}$	$y^{(1)}$
Số liệu (2)	$x^{(2)}$	$y^{(2)}$
$\vdots$	$\vdots$	$\vdots$
Số liệu (N)	$x^{(N)}$	$y^{(N)}$

Hoàn toàn tương tự như cho trường hợp $n = 1$ biến, ta được phần dư thứ i,

$$r^{(i)} = y^{(i)} - f\left(x^{(i)}\right) = y_i - \left(\theta_1 f_1\left(x^{(i)}\right) + \theta_2 f_2\left(x^{(i)}\right) + \cdots + \theta_p f_p\left(x^{(i)}\right)\right),$$

$i = 1,2,\cdots,N$. Do đó

$$r = \begin{bmatrix} r^{(1)} \\ r^{(2)} \\ \vdots \\ r^{(N)} \end{bmatrix} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{bmatrix} - \begin{bmatrix} f_1\left(x^{(1)}\right) & f_2\left(x^{(1)}\right) & \cdots & f_p\left(x^{(1)}\right) \\ f_1\left(x^{(2)}\right) & f_2\left(x^{(2)}\right) & \cdots & f_p\left(x^{(2)}\right) \\ \vdots & \vdots & \ddots & \vdots \\ f_1\left(x^{(N)}\right) & f_2\left(x^{(N)}\right) & \cdots & f_p\left(x^{(N)}\right) \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix}$$
$$\equiv y - X\theta,$$

với

$$y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{bmatrix} \in \mathbb{R}^N, X = \begin{bmatrix} f_1\left(x^{(1)}\right) & f_2\left(x^{(1)}\right) & \cdots & f_p\left(x^{(1)}\right) \\ f_1\left(x^{(2)}\right) & f_2\left(x^{(2)}\right) & \cdots & f_p\left(x^{(2)}\right) \\ \vdots & \vdots & \ddots & \vdots \\ f_1\left(x^{(N)}\right) & f_2\left(x^{(N)}\right) & \cdots & f_p\left(x^{(N)}\right) \end{bmatrix} \in$$
$$\mathbb{R}^{N\times p}, \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix} \in \mathbb{R}^p.$$

Suy ra

$$RSS(\theta) = \|r\|^2 = \|X\theta - y\|^2.$$

Áp dụng Định lý 2.3.3 cho $A = X \in \mathbb{R}^{N\times p}$, với $p \leq N$, sao cho $rank(X) = p = min(N,p)$, và vectơ $b = y \in \mathbb{R}^N$, ta được

$$\theta_0 = (X^T X)^{-1} X^T y = \underset{\theta\in\mathbb{R}^p}{argmin}\|X\theta - y\|^2.$$

Ví dụ 4. Giả sử ta cần tìm mối liên hệ giữa giá bán nhà, y, đơn vị tính tỷ đồng, với diện tích, x_1, tính bằng m^2, và số phòng ngủ, x_2, của căn nhà đó, từ bộ dữ liệu

x_1	x_2	y
50	1	1.5
60	2	3.2
70	3	4.5
80	3	5.8
100	3	6.5

(i) Với mô hình tuyến tính $y = a + bx_1 + cx_2$, tổ hợp tuyến tính của ba hàm cơ sở $f_1(x_1, x_2) = 1$, $f_2(x_1, x_2) = x_1$, và $f_3(x_1, x_2) = x_2$. Đặt

$$X = \begin{bmatrix} f_1(50,1) & f_2(50,1) & f_3(50,1) \\ f_1(60,2) & f_2(60,2) & f_3(60,2) \\ f_1(70,3) & f_2(70,3) & f_3(70,3) \\ f_1(80,3) & f_2(80,3) & f_3(80,3) \\ f_1(100,3) & f_2(100,3) & f_3(100,3) \end{bmatrix} = \begin{bmatrix} 1 & 50 & 1 \\ 1 & 60 & 2 \\ 1 & 70 & 3 \\ 1 & 80 & 3 \\ 1 & 100 & 3 \end{bmatrix},$$

$$y = \begin{bmatrix} 1.5 \\ 3.2 \\ 4.5 \\ 5.8 \\ 6.5 \end{bmatrix}, \theta = \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

Hàm "khớp" với dữ liệu đã cho nhất trong mô hình này xác định bởi

$$\theta = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = (X^T X)^{-1} X^T y =$$

$$\begin{bmatrix} 5 & 360 & 12 \\ 360 & 27400 & 920 \\ 12 & 920 & 32 \end{bmatrix}^{-1} \begin{bmatrix} 21.5 \\ 1696 \\ 58.3 \end{bmatrix} = \begin{bmatrix} -2.57 \\ 0.0615 \\ 1.0175 \end{bmatrix},$$

nghĩa là hàm $y = -2.57 + 0.0615x_1 + 1.0175x_2$.

(ii) Với mô hình đa thức $y = a + bx_1 + cx_1^2 + dx_2$, tổ hợp tuyến tính của 4 hàm cơ sở $f_1(x_1, x_2) = 1$, $f_2(x_1, x_2) = x_1$, $f_3(x_1, x_2) = x_1^2$, và $f_4(x_1, x_2) = x_2$. Đặt

$$X = \begin{bmatrix} f_1(50,1) & f_2(50,1) & f_3(50,1) & f_4(50,1) \\ f_1(60,2) & f_2(60,2) & f_3(60,2) & f_4(60,2) \\ f_1(70,3) & f_2(70,3) & f_3(70,3) & f_4(70,3) \\ f_1(80,3) & f_2(80,3) & f_3(80,3) & f_4(80,3) \\ f_1(100,3) & f_2(100,3) & f_3(100,3) & f_4(100,3) \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 50 & 50^2 & 1 \\ 1 & 60 & 60^2 & 2 \\ 1 & 70 & 70^2 & 3 \\ 1 & 80 & 80^2 & 3 \\ 1 & 100 & 100^2 & 3 \end{bmatrix},$$

$$y = \begin{bmatrix} 1.5 \\ 3.2 \\ 4.5 \\ 5.8 \\ 6.5 \end{bmatrix}, \theta = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}.$$

Hàm "khớp" với dữ liệu đã cho nhất trong mô hình này xác định bởi

$$\theta = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = (X^T X)^{-1} X^T y =$$

$$\begin{bmatrix} 5 & 360 & 27400 & 12 \\ 360 & 27400 & 2196000 & 920 \\ 27400 & 2196000 & 184180000 & 73600 \\ 12 & 920 & 73600 & 32 \end{bmatrix}^{-1} \begin{bmatrix} 21.5 \\ 1696 \\ 139440 \\ 58.3 \end{bmatrix} = \begin{bmatrix} -19.25 \\ 0.5775 \\ -0.003 \\ -0.6625 \end{bmatrix},$$

i.e., nghĩa là hàm $y = -19.25 + 0.5775x_1 - 0.003x_1^2 - 0.6625x_2$.

(iii) Xét mô hình tuyến tính có thêm **thừa số tương tác** (interaction term) $y = a + bx_1 + cx_2 + dx_1x_2$, tổ hợp tuyến tính của 4 hàm cơ sở, $f_1(x_1, x_2) = 1$, $f_2(x_1, x_2) = x_1$, $f_3(x_1, x_2) = x_2$, và $f_4(x_1, x_2) = x_1x_2$. Đặt

$$X = \begin{bmatrix} f_1(50,1) & f_2(50,1) & f_3(50,1) & f_4(50,1) \\ f_1(60,2) & f_2(60,2) & f_3(60,2) & f_4(60,2) \\ f_1(70,3) & f_2(70,3) & f_3(70,3) & f_4(70,3) \\ f_1(80,3) & f_2(80,3) & f_3(80,3) & f_4(80,3) \\ f_1(100,3) & f_2(100,3) & f_3(100,3) & f_4(100,3) \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 50 & 1 & 1 \cdot 50 \\ 1 & 60 & 2 & 2 \cdot 60 \\ 1 & 70 & 3 & 3 \cdot 70 \\ 1 & 80 & 3 & 3 \cdot 80 \\ 1 & 100 & 3 & 3 \cdot 100 \end{bmatrix}, y = \begin{bmatrix} 1.5 \\ 3.2 \\ 4.5 \\ 5.8 \\ 6.5 \end{bmatrix}, \theta = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}.$$

Hàm "khớp" với dữ liệu đã cho nhất trong mô hình này xác định bởi

$$\theta = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = (X^T X)^{-1} X^T y =$$

$$\begin{bmatrix} 5 & 360 & 12 & 920 \\ 360 & 27400 & 920 & 73600 \\ 12 & 920 & 32 & 2540 \\ 920 & 73600 & 2540 & 208600 \end{bmatrix}^{-1} \begin{bmatrix} 21.5 \\ 1696 \\ 58.3 \\ 4746 \end{bmatrix} = \begin{bmatrix} -3.58571 \\ 0.08143 \\ 1.33571 \\ -0.00643 \end{bmatrix},$$

nghĩa là hàm $y = -3.58571 + 0.08143x_1 + 1.33571x_2 - 0.00643x_1x_2$.

BÀI TẬP

1) Trong bài báo "Residual Stresses and Adhesion of Thermal Spray Coatings" (Surface Engineering, 2005: 35–40) tác giả nghiên cứu mối quan hệ giữa độ dày x (mm) của lớp phủ NiCrAl lắng đọng trên nền thép không gỉ và độ bền liên kết y tương ứng (MPa) với dữ liệu như sau

Độ dày	220	370	440	680	860
Độ bền	24.0	26.3	25.2	17.0	12.2

Giả sử x và y thỏa mô hình $y = \beta_0 + \beta_1 x + \beta_2 x^2$. Hãy sử dụng phương pháp bình phương nhỏ nhất ước lượng các tham số β_0, β_1 và β_2.

2) Giả sử rằng tám mẫu thử của một loại hợp kim nhất định được sản xuất ở các nhiệt độ khác nhau, và độ bền của mỗi mẫu vật sau đó được quan sát. Các giá trị quan sát được cho trong bảng

x_i	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
y_i	40	41	43	42	44	42	43	42

Trong đó x_i là nhiệt độ của mẫu thứ i và y_i là độ bền tương ứng. Hãy sử dụng phương pháp bình phương nhỏ nhất tìm các tham số chưa biết nếu x và y thỏa

a) $y = \beta_0 + \beta_1 x$

b) $y = \beta_0 + \beta_1 x + \beta_2 x^2$

3) Giả sử rằng mỗi người trong số 10 bệnh nhân được điều trị với cùng một lượng của hai loại thuốc khác nhau có thể ảnh hưởng đến huyết áp. Cụ thể, mỗi bệnh nhân đầu tiên được điều trị bằng loại thuốc tiêu chuẩn A, và sự thay đổi huyết áp x_1 của họ được đo. Sau khi tác dụng của thuốc hết tác dụng, bệnh nhân được điều trị bằng một lượng thuốc B mới tương đương, và sự thay đổi huyết áp y của họ được đo lại. Những thay đổi về huyết áp này sẽ được gọi là phản ứng của người bệnh với từng loại thuốc. Hơn nữa ta cũng đo nhịp tim x_2 của các bệnh nhân. Ta có bảng dữ liệu như sau:

x_1	1.9	0.8	1.1	0.1	−0.1	4.4
x_2	66	62	64	61	63	70
y	0.7	−1.0	−0.2	−1.2	−0.1	3.4

Giả sử x_1, x_2 và y thỏa mô hình $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. Hãy sử dụng phương pháp bình phương nhỏ nhất ước lượng các tham số β_0, β_1 và β_2.

4) Với dữ liệu nhận được cho (x_1, x_2, y),

x_1	x_2	y
2.04	3.55	3.11
2.04	6.07	3.26
3.06	3.55	3.89
3.06	6.97	10.25
4.08	3.55	3.11
4.08	6.16	13.48
2.06	3.62	3.94
2.06	6.16	3.53

dùng phương pháp bình phương nhỏ nhất ước lượng các tham số β_k trong mô hình

a) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$.

b) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$.

Chương 4

GIẢI SỐ PHƯƠNG TRÌNH VI PHÂN

4.1. ĐẶT VẤN ĐỀ

Trong Chương 3, ta xét sự liên hệ của biến phụ thuộc y theo biến độc lập x bằng mô hình của hàm một biến

$$y = f(x). \tag{1}$$

Trong nhiều ứng dụng quan trọng, mối quan hệ giữa x và y được xác lập dựa trên **tỷ lệ thay đổi** (rate of changes) của y theo x, nghĩa là theo đạo hàm của y theo x,

$$y' = f(x, y), \tag{2}$$

và (2) được gọi là một **mô hình động** (dynamic model). Về mặt toán học, ta nói (2) là một **phương trình vi phân** (ordinary differential equation – ODE).

Chẳng hạn, trong các mô hình dân số (population models) được khảo sát bởi T. R. Malthus, với $p(t)$ chỉ dân số* một loài

* Dù rằng dân số một loài luôn là số nguyên không âm nhưng người ta thường giả định rằng nó đủ lớn để có thể xấp xỉ bằng một hàm số có giá trị thực với sai số không đáng kể.

tại thời điểm t, người ta giả định rằng độ thay đổi dân số thì tỷ lệ với dân số, nghĩa là tồn tại hằng số tỷ lệ $k > 0$ sao cho $\frac{dp}{dt} = kp$, ta nhận được một phương trình vi phân. Bấy giờ, khi biết dân số tại một thời điểm, chẳng hạn $p(0) = p_0$, ta được **bài toán giá trị đầu** (initial value problem),

$$\frac{dp}{dt} = kp, p(0) = p_0, \tag{3}$$

và mô hình toán học này được biết dưới tên là **luật Malthus** (Malthusian law), hay **luật mũ** (exponential law), với nghiệm

$$p(t) = p_0 e^{kt}.$$

Mô hình này được tiếp tục phát triển bởi P. F. Verhust bằng cách xét thêm **đại lượng tương tác** (interaction term) của dân số kích thước p, $p(p-1)/2$, dẫn đến **mô hình logistic** (logistic model),

$$\frac{dp}{dt} = -Ap(p - p_1), p(0) = p_0, \tag{4}$$

với $A > 0, p_1 > 1$. Nghiệm của (4),

$$p(t) = \frac{p_0 p_1}{p_0 + (p_1 - p_0)e^{-Ap_1 t}},$$

được gọi là **hàm logistic** (logistic function).

Khi áp dụng các mô hình dân số nêu trên cho một loài cụ thể, các tham số k (của mô hình mũ) cũng như các tham số A và p_1 (của mô hình logistic) cần được ước lượng từ dữ liệu thu thập của dân số trong quá khứ. Chẳng hạn, (4) có thể viết lại thành

$$\frac{1}{p}\frac{dp}{dt} = Ap_1 - Ap$$

mà ta xấp xỉ vế trái

$$y = \frac{1}{p}\frac{dp}{dt}$$

như là hàm bậc nhất theo p,

$$y = a + bp.$$

Chẳng hạn, với số liệu thu thập được trên p, xấp xỉ đạo hàm p' bằng một cách nào đó, chẳng hạn

$$\frac{dp}{dt}(t) \approx \frac{p(t+h)-p(t-h)}{2h}, \text{ với } \square > 0,$$

ta xấp xỉ được bộ số liệu cho (p, y). Dùng phương pháp OLS trình bày trong Chương 3 cho bộ số liệu này, ta nhận được giá trị $a = Ap_1$, và $b = A$. Từ đó, suy ra các tham số A và p_1 của mô hình logistic.

Hơn nữa, không phải lúc nào nghiệm của bài toán giá trị đầu

$$\frac{dy}{dx} = f(x, y), y(x_0) = y_0, \tag{5}$$

cũng có thể viết dưới dạng tường minh như trong các mô hình (3) và (4). Tuy nhiên, dù nghiệm của (5) có dạng tường minh hay không, ta cũng cần phải tính xấp xỉ giá trị của nó tại một số điểm cần thiết $x \neq x_0$.

4.2. PHƯƠNG PHÁP EULER

Trước hết, ta chấp nhận kết quả sau về sự tồn tại và duy nhất nghiệm của (5).

4.2.1. Định lý. Nếu f và $\frac{\partial f}{\partial y}$ xác định và liên tục trên một hình chữ nhật

$$R = \{(x, y) | a < x < b, c < y < d\}$$

chứa điểm (x_0, y_0), thì (5) có duy nhất nghiệm $y = \phi(x)$ xác định trên một khoảng mở chứa x_0, $(x_0 - \delta, x_0 + \delta)$, $\delta > 0$, nghĩa là

$\phi'(x) = f\big(x, \phi(x)\big)$, với mọi $x \in (x_0 - \delta, x_0 + \delta)$, và $\phi(x_0) = y_0$.

Ví dụ 1. Với bài toán giá trị đầu

$$\frac{dy}{dx} = x\sqrt{y}, \quad y(1) = 4, \tag{6}$$

ta có

$$f(x, y) = x\sqrt{y} \text{ và } \frac{\partial f}{\partial y}(x, y) = \frac{x}{2\sqrt{y}}$$

xác định và liên tục trên mọi hình chữ nhật chứa điểm (1,4) không chứa điểm (1,0) nên do Định lý 4.2.1, tồn tại duy nhất hàm $y = \phi(x)$ xác định trên một lân cận của điểm 1, $(1 - \delta, 1 + \delta)$, với $\delta > 0$, thỏa (6), nghĩa là

$$\phi(1) = 4, \text{ và } \phi'(x) = x\sqrt{\phi(x)},$$

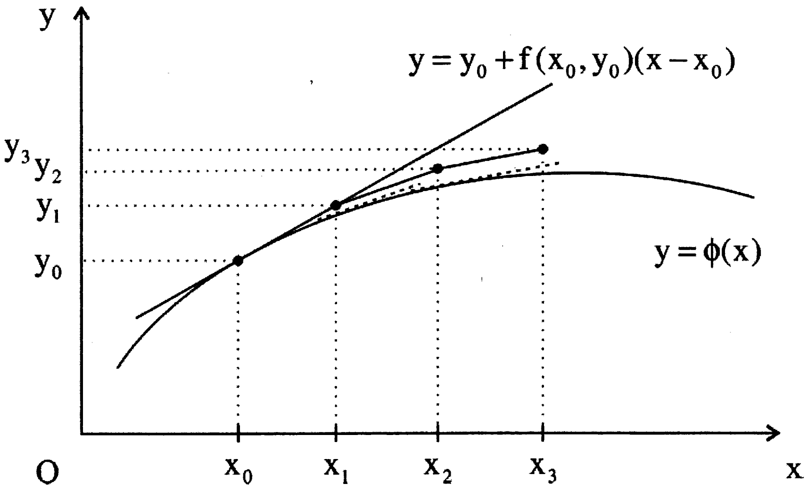
với mọi $x \in (1 - \delta, 1 + \delta)$.

Khi (5) có nghiệm duy nhất, phương pháp Euler (còn gọi là phương pháp tiếp tuyến) tìm cách xấp xỉ giá trị của $y = \phi(x)$ tại các điểm $x \neq x_0$ bằng công thức khai triển Taylor,

$$\phi(x) \approx \phi(x_0) + \phi'(x_0)(x - x_0),$$

nghĩa là

$$y \approx y_0 + f(x_0, y_0)(x - x_0).$$



Ví dụ 2. Với bài toán giá trị đầu của Ví dụ 1,

$$\frac{dy}{dx} = x\sqrt{y}, \quad y(1) = 4,$$

ta có $x_0 = 1$, $y_0 = y(x_0) = 4$, nên tại $x = x_1 = 1.1$, ta có xấp xỉ

$$y(1.1) = y(x_1) = y_1 \approx y_0 + f(x_0, y_0)(x_1 - x_0)$$

$$= 4 + 1 \cdot \sqrt{4} \cdot (1.1 - 1) = 4.2$$

Tại điểm $x = x_2 = 1.2$, ta xấp xỉ

$$y_2 \approx y_1 + f(x_1, y_1)(x_2 - x_1),$$

với chú ý rằng do y_1 chỉ là một xấp xỉ của $\phi(x_1)$ nên đường

$$y \approx y_1 + f(x_1, y_1)(x - x_1)$$

không hẳn là tiếp tuyến với đồ thị hàm $y = \phi(x)$. Tương tự cho các điểm $x = 1.3,\ 1.4, \ldots$ Các giá trị $x_1 - x_0$, $x_2 - x_1$, $x_3 - x_2$, ... gọi là các **bước** (step) của xấp xỉ. Với **kích thước bước** (step size) chung

$$h = x_{n+1} - x_n, n = 0,1,2, \cdots,$$

giải thuật Euler tính xấp xỉ $y(x_n) = y_n$, $n = 0,1,2, \cdots$, bằng công thức

$$y_{n+1} = y_n + hf(x_n, y_n), n = 0,1,2, \cdots$$

trong đó $x_{n+1} = x_n + h = x_0 + nh$, với $n = 0,1,2, \cdots$.

Ví dụ 3. Với bài toán giá trị đầu của Ví dụ 1, và với kích thước bước $h = 0.1$, giá trị xấp xỉ của y tại các điểm $x_n = x_0 + nh$, $n = 0,1,2, \cdots$ như sau

$$y_0 = y(x_0) = 4,$$

$$y_1 = y_0 + (0.1)x_0\sqrt{y_0} = 4 + (0.1)(1)\sqrt{4} = 4.2,$$

$$y_2 = y_1 + (0.1)x_1\sqrt{y_1} = 4.2 + (0.1)(1.1)\sqrt{4.2} \approx 4.42543, \ldots \blacksquare$$

4.3. PHƯƠNG PHÁP TAYLOR VÀ RUNGE-KUTTA

Trong phần này, ta điểm qua một số giải thuật tính số cho nghiệm y của bài toán giá trị đầu

$$\frac{dy}{dx} = f(x, y), y(x_0) = y_0,$$

trên đoạn $[x_0, c]$ bằng cách chia đoạn này thành N đoạn đều nhau chiều dài

$$h = \frac{c - x_0}{N}$$

và xấp xỉ các giá trị $y_n = y(x_n)$ tại các điểm $x_n = x_0 + nh, n = 0,1,2,\cdots,N$.

Trước hết, giải thuật Euler xây dựng dựa trên công thức Taylor

$$\phi(x) \approx \phi(x_0) + \phi'(x_0)(x - x_0)$$

cho nghiệm $y = \phi(x)$ để được công thức xấp xỉ

$$y \approx y_0 + f(x_0, y_0)(x - x_0).$$

Từ đó, ta được

$$y_{n+1} = y_n + hf(x_n, y_n), n = 0,1,2,\cdots,N. \quad (7)$$

Nhằm nâng cao độ chính xác cho xấp xỉ, người ta dùng công thức Taylor cấp cao hơn 1. Chẳng hạn, với xấp xỉ đến cấp 2,

$$\phi(x) = \phi(x_0) + \phi'(x_0)(x - x_0) + \frac{(x - x_0)^2}{2}\phi''(x_0)$$

Do $\phi'(x) = f\big(x, \phi(x)\big)$, ta suy ra

$$\begin{aligned}\phi''(x) &= \frac{\partial f}{\partial x}\big(x, \phi(x)\big) + \frac{\partial f}{\partial y}\big(x, \phi(x)\big)\phi'(x) \\ &= \frac{\partial f}{\partial x}\big(x, \phi(x)\big) + \frac{\partial f}{\partial y}\big(x, \phi(x)\big)f\big(x, \phi(x)\big)\end{aligned}$$

và do đó

$$y \approx y_0 + f(x_0, y_0)(x - x_0) + \frac{(x-x_0)^2}{2}\left[\frac{\partial f}{\partial x}(x_0, y_0) + \frac{\partial f}{\partial y}(x_0, y_0) f(x_0, y_0)\right].$$

Công thức (7) trở thành

$$y_{n+1} = y_n + h\left[f(x_n, y_n) + \frac{h}{2}\left(\frac{\partial f}{\partial x}(x_n, y_n) + \frac{\partial f}{\partial y}(x_n, y_n) f(x_n, y_n)\right)\right].$$

Giải thuật Taylor cấp 2

1. Khởi tạo $x = x_0$, $y = y_0$, $h = (c - x_0)/N$.

2. Với $i = 1,2,\cdots,N$, thực hiện các bước sau

3. Tính $F = f(x,y) + \frac{h}{2}\left(\frac{\partial f}{\partial x}(x,y) + \frac{\partial f}{\partial y}(x,y) f(x,y)\right)$.

4. Đặt $x = x + h$, $y = y + hF$.

Giải thuật Runge-Kutta cấp 2 tiếp tục xấp xỉ biểu thức F trong bước 3. Chẳng hạn, với xấp xỉ

$$F \approx f\left(x + \frac{h}{2}, y + \frac{h}{2} f(x,y)\right),$$

ta được

$$y_{n+1} = y_n + hf\left(x_n + \frac{h}{2}, y_n + \frac{h}{2} f(x_n, y_n)\right).$$

Giải thuật Runge-Kutta cấp 2

1. Khởi tạo $x = x_0, y = y_0, h = (c - x_0)/N$.

2. Với $i = 1,2,\cdots,N$, thực hiện các bước sau

3. Tính $F = f(x,y)$ và $G = f\left(x + \frac{h}{2}, y + \frac{hF}{2}\right)$.

4. Đặt $x = x + h, y = y + hG$.

BÀI TẬP

1) Dùng phương pháp Euler với $h = 0.25$, tính xấp xỉ giá trị $y(1)$, trong đó $y(x)$ là nghiệm của bài toán giá trị đầu

a) $\frac{dy}{dx} = xy - y^2, y(0) = 1$.

b) $\frac{dy}{dx} = x - y, y(0) = 0$.

c) $\frac{dy}{dx} = x^2 + y, y(0) = 0$.

2) Xét bài toán giá trị đầu

$$\frac{dy}{dx} = 2y - 6, y(0) = 1.$$

a) Dùng phương pháp Euler với $h = 0.25$, tính xấp xỉ giá trị $y(1)$.

b) Dùng phương pháp Taylor cấp 2 với $h = 0.25$, tính xấp xỉ giá trị $y(1)$.

c) Dùng phương pháp Runge-Kutta cấp 2 với $h = 0.25$, tính xấp xỉ giá trị $y(1)$.

d) So sánh kết quả nhận được với nghiệm chính xác $y = 3 - 2e^{2x}$.

3) Xét bài toán giá trị đầu

$$\frac{dy}{dx} = 1 - y, y(0) = 0.$$

a) Dùng phương pháp Euler với $h = 0.25$, tính xấp xỉ giá trị $y(1)$.

b) Dùng phương pháp Taylor cấp 2 với $h = 0.25$, tính xấp xỉ giá trị $y(1)$.

c) Dùng phương pháp Runge-Kutta cấp 2 với $h = 0.25$, tính xấp xỉ giá trị $y(1)$.

d) So sánh kết quả nhận được với nghiệm chính xác $y = 1 - e^{-x}$.

Chương 5

PHƯƠNG PHÁP HỢP LÝ CỰC ĐẠI

5.1. ĐẶT VẤN ĐỀ

Trong chương này, ta khảo sát biến số ngẫu nhiên, mà theo quan điểm ứng dụng, chúng là đại lượng thay đổi theo một số đại lượng khác mà tạm thời ta chưa biết hoặc không thể kiểm soát. Ngay cả với mô hình khảo sát sự thay đổi của biến y theo một số biến khác, $x_1, x_2, \cdots, x_n$, bằng mô hình hàm số

$$y = f(x_1, x_2, \cdots, x_n),$$

khảo sát trong Chương 3, thống kê suy diễn xem y như là giá trị của một biến ngẫu nhiên Y,

$$Y = f(x_1, x_2, \cdots, x_n) + \varepsilon,$$

với ε dành cho tác động của các biến không xuất hiện trong mô hình.

Để khảo sát một biến số ngẫu nhiên X, ta tìm cách xác định hàm mật độ xác suất (probability density function – pdf), $f_X(x)$, của nó. Khi đó, các đại lượng liên quan đến X như trung bình, $E(X) = \mu_X$, phương sai, $Var(X) = \sigma_X^2$, cũng như các phân vị $Q(p)$, với $0 < p < 1$, đều có thể suy ra từ hàm $f_X(x)$.

5.2. PHƯƠNG PHÁP HỢP LÝ CỰC ĐẠI

Cho X là một biến số ngẫu nhiên. Hàm $F: \mathbb{R} \to \mathbb{R}$, xác định bởi

$$F(x) = P(X \leq x), \text{ với } x \in \mathbb{R},$$

được gọi là **hàm phân phối tích lũy** (cumulative distribution function – cdf) của X, với các tính chất căn bản sau (xem [10]).

5.2.1. Mệnh đề. (i) $0 \leq F(x) \leq 1$, với mọi $x \in \mathbb{R}$.

(ii) F là hàm tăng, i.e., $F(x) \leq F(y)$ khi $x < y$.

(iii) F liên tục bên phải tại mọi điểm, i.e., $lim_{t \to x^+} F(t) = F(x)$, với mọi $x \in \mathbb{R}$.*

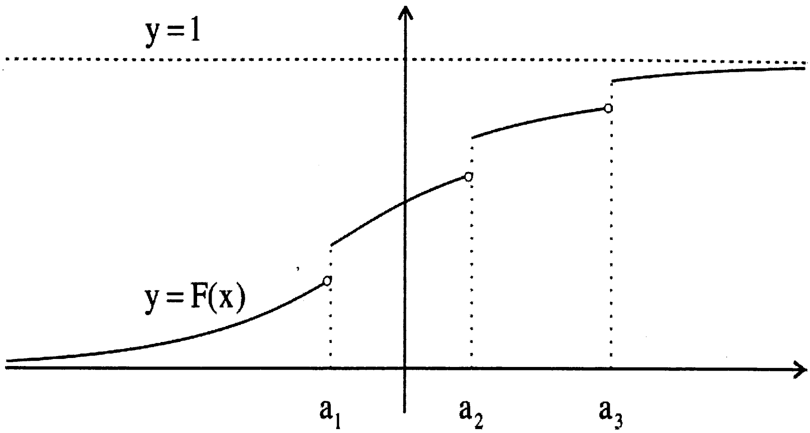
(iv) $lim_{x \to -\infty} F(x) = 0$ và $lim_{x \to \infty} F(x) = 1$.

Khi đó, tập các điểm không liên tục của F là tập rời rạc† (discrete) và do đó nó hoặc là tập hữu hạn, hoặc là tập đếm được.

Đồ thị cho cdf của một biến số ngẫu nhiên tổng quát, $y = F(x)$, với ba điểm không liên tục, a_1, a_2, a_3, cho trong Hình 1. Ký hiệu o trên một nhánh ám chỉ rằng giá trị của hàm $F(x)$ không được xác định bằng nhánh đồ thị đó.

* Dễ dàng chứng minh rằng F có giới hạn bên trái tại mọi điểm, i.e., $\lim_{t \to x^-} F(t)$ tồn tại, với mọi $x \in \mathbb{R}$. Ký hiệu **càdlàg** (tiếng Pháp "continue à droite, limite à gauche") được dùng phổ biến trong giới xác suất thống kê để chỉ tính chất đặc biệt này của cdf.

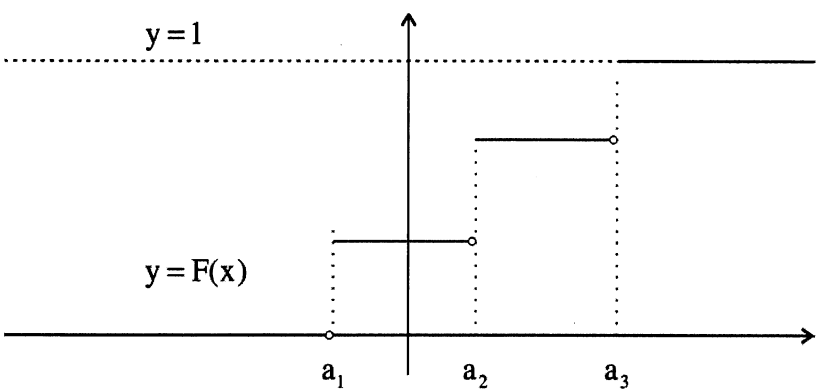
† Tập $A \subset \mathbb{R}$ được gọi là rời rạc khi với mọi $x \in A$, tồn tại $\delta > 0$ sao cho $(x-\delta, x+\delta) \cap A = \{x\}$. Nói khác đi, mọi điểm khác của A phải nằm xa điểm x. Điểm x này còn được gọi là **điểm cô lập** (isolated point) của A.



Hình 1. Đồ thị hàm phân phối tích lũy của một biến số ngẫu nhiên tổng quát

5.2.2. Biến ngẫu nhiên rời rạc

Khi cdf của biến số ngẫu nhiên X là hàm hằng giữa hai điểm không liên tục của nó,



Hình 2. Đồ thị hàm phân phối tích lũy của một biến số ngẫu nhiên rời rạc

X được gọi là **biến số ngẫu nhiên rời rạc** (discrete random variable). Với các điểm không liên tục của $y = F(x)$

$$a_1 < a_2 < \cdots < a_k < \cdots,$$

ta có

(i) $P(X = x) = 0$, tại các điểm $x \neq a_k$, với mọi k, và

(ii) $P(X = a_k) = F(a_k) - lim_{x\uparrow a_k} F\,(x) \equiv p_k$,

và hàm số $f: \mathbb{R} \to \mathbb{R}$, xác định bởi

$$f(x) = \begin{cases} p_k, & x = a_k, \\ 0, & x \neq a_k, \forall k \end{cases}$$

được gọi là **hàm mật độ xác suất** (probability density function – pdf, hay probability mass function – pmf) của X.

Bây giờ, xét số liệu quan sát $x_1, x_2, \cdots, x_n$ nhận được từ biến số ngẫu nhiên rời rạc X với pdf $f(x)$ mà ta coi như chúng là giá trị của dãy các biến số ngẫu nhiên $X_1, X_2, \cdots, X_n$, độc lập và có cùng phân phối (independent, identical distribution – iid) với X, mà ta còn gọi là **mẫu ngẫu nhiên** (random sample). Khi đó, giá trị của **hàm mật độ xác suất đồng thời** (joint-pdf) của vectơ ngẫu nhiên $V = (X_1, X_2, \cdots, X_n)$ **tại điểm** $(x_1, x_2, \cdots, x_n)$,

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = f_V(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} f_X(x_i),$$

cho ta số đo đánh giá mức độ phù hợp của mẫu ngẫu nhiên với số liệu quan sát. Ứng với hai pdf khác nhau, pdf nào có số đo này lớn hơn được xem là tốt hơn.

Bằng cách tìm hàm mật độ xác suất của X trong một lớp các pdf $f(x;\theta)$, với $\theta \in \Theta \subset \mathbb{R}^p$, phương pháp hợp lý cực đại tìm cách xác định $\theta_0 \in \Theta$ sao cho

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) \tag{1}$$

đạt giá trị lớn nhất tại $\theta = \theta_0$, nghĩa là

$$\theta_0 = \underset{\theta\in\Theta}{argmax} L(\theta).$$

Hàm $L(\theta)$ xác định bởi (1) được gọi là **hàm hợp lý** (likelihood function) của số liệu mẫu $x_1, x_2, \cdots, x_n$ và θ_0 được gọi là **ước lượng hợp lý cực đại** (maximum likelihood estimator) của θ.

Ví dụ 1. Xét số liệu quan sát nhận được của một biến X,

$$1, 0, 1, 1.$$

Để tìm pdf tốt nhất cho X trong lớp các pdf $f(x;p)$, $0 < p < 1$, xác định bởi

$$f(x;p) = \begin{cases} p^x(1-p)^{1-x}, & x = 0{,}1, \\ 0, & x \neq 0{,}1, \end{cases}$$

Ta khảo sát hàm hợp lý

$$L(p) = f(1;p)\cdot f(0;p)\cdot f(1;p)\cdot f(1;p) = p^3(1-p).$$

Lập bảng tính một số giá trị của $L(p)$,

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$L(p)$	0.0009	0.0064	0.0189	0.0384	0.0625	0.0864	0.1029	0.1024	0.0729

ta thấy $f(x; 0.7)$ là pdf tốt nhất trong số các pdf được tính. Để tìm pdf tốt nhất, ta tìm

$$p_0 = \underset{0<p<1}{argmax} L(p).$$

Ta có $L'(p) = 3p^2 - 4p^3 = p^2(3 - 4p)$ và $L'(p) = 0 \Leftrightarrow p = \frac{3}{4}$ và

p	0		$\frac{3}{4}$		1
$L'(p)$		+	0	−	
$L(p)$	0	↗	$\frac{27}{256}$	↘	0

Suy ra $p_0 = \frac{3}{4}$ và pdf $f\left(x; \frac{3}{4}\right)$ là pdf tốt nhất cho bộ số liệu 1, 0, 1, 1.

Do $y = ln\, x$ là hàm đơn điệu tăng, ta được

$$\theta_0 = \underset{\theta\in\Theta}{argmax} L(\theta) \Leftrightarrow \theta_0 = \underset{\theta\in\Theta}{argmax}\, ln\big(L(\theta)\big)$$

và do đó, thay vì tìm điểm cực đại toàn cục cho hàm hợp lý, $L(\theta)$, ta có thể tìm điểm cực đại toàn cục của hàm $L(\theta) = ln\big(L(\theta)\big)$, hàm **log-hợp lý** (log-likelihood function).

Ví dụ 2. Hàm log-hợp lý của Ví dụ 1 cho bởi

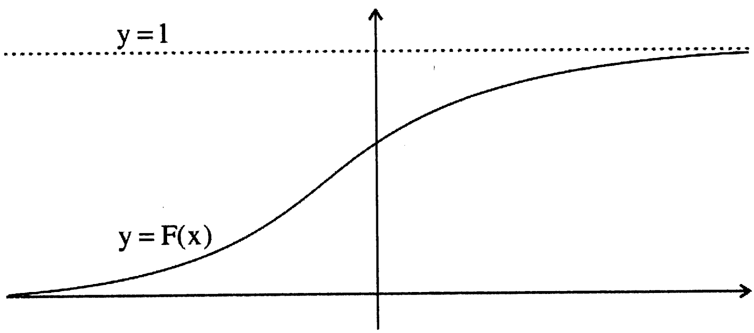
$$L(p) = ln\big(L(p)\big) = 3\, ln\, p + ln(1 - p).$$

Ta có $L'(p) = \frac{3}{p} - \frac{1}{1-p}$, và $L''(p) = -\frac{3}{p^2} - \frac{1}{(1-p)^2} < 0$, với mọi $p \in (0,1)$. Suy ra $L(p)$ là hàm lõm và vì $L'(p) = 0 \Leftrightarrow p = \frac{3}{4}$, ta cũng nhận được

$$\frac{3}{4} = \underset{0<p<1}{argmax} L(p) = \underset{0<p<1}{argmax} L(p). \quad \blacksquare$$

5.2.3 Biến ngẫu nhiên liên tục

Khi cdf của biến số ngẫu nhiên X là hàm số liên tục trên $\mathbb{R}$,



Hình 3. Đồ thị hàm phân phối tích lũy của một biến số ngẫu nhiên liên tục

X được gọi là **biến số ngẫu nhiên liên tục** (continuous random variable). Khi đó, người ta chứng minh được rằng tồn tại một hàm số $y = f(x)$ sao cho (xem [10])

$$F(x) = \int_{-\infty}^{x} f(t)dt, \text{ với mọi } x \in \mathbb{R},$$

và hàm $y = f(x)$ được gọi là hàm mật độ xác suất – pdf của X. Chú ý rằng khi đó, ta có

$$P(X \in I) = \int_I f(x)dx,$$

với mọi khoảng $I \subset \mathbb{R}$. Đặc biệt,

(i) $P(X = x) = 0$, với mọi $x \in \mathbb{R}$, và

(ii) $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$, với mọi $a, b \in \mathbb{R}$, $a \leq b$.

Khác với biến số ngẫu nhiên rời rạc, giá trị của pdf tại một điểm x, $f(x)$, không phải là xác suất để X lấy giá trị x đó, $P(X = x)$, nhưng giá trị này cho ta số đo đánh giá khả năng X lấy giá trị "xung quanh" điểm x. Chẳng hạn, khi pdf $f(x)$ là một hàm số liên tục, ta có

$$P(x - \delta < X < x + \delta) = \int_{x-\delta}^{x+\delta} f(t)dt \approx 2\delta \cdot f(x).$$

Chính vì vậy, giải thuật tìm ước lượng hợp lý cực đại bằng hàm hợp lý, hay hàm log-hợp lý cho các biến số ngẫu nhiên rời rạc cũng được nới rộng sử dụng cho các biến số ngẫu nhiên liên tục.

Ví dụ 3. Hàm mật độ xác suất của một biến số ngẫu nhiên X có phân phối mũ với tham số λ cho bởi

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Với số liệu quan sát nhận được là 1, 2, 4, ta có hàm hợp lý

$$L(\lambda) = f(1, \lambda) \cdot f(2, \lambda) \cdot f(4, \lambda) = \lambda e^{-\lambda} \cdot \lambda e^{-2\lambda} \cdot \lambda e^{-4\lambda} = \lambda^3 e^{-7\lambda}$$

và hàm log-hợp lý

$$L(\lambda) = ln\big(L(\lambda)\big) = 3\, ln\, \lambda - 7\lambda.$$

Do

$$L'(\lambda) = \frac{3}{\lambda} - 7, \text{ và } L''(\lambda) = -\frac{3}{\lambda^2} < 0, \text{ với mọi } \lambda \in (0,\infty),$$

Ta suy ra $L(\lambda)$ là hàm lõm trên $(0,\infty)$ và vì $L'(\lambda) = 0 \Leftrightarrow \lambda = \frac{3}{7}$, ta suy ra

$$\frac{3}{7} = \underset{\lambda\in(0,\infty)}{argmax} L(\lambda),$$

và do đó pdf $f\left(x; \frac{3}{7}\right)$ là phân phối mũ phù hợp nhất với dữ liệu 1, 2, 4.

Chú ý. Các biến số ngẫu nhiên có hàm phân phối tích lũy tổng quát cho trong Hình 1 được gọi là các **biến số ngẫu nhiên hỗn hợp** (mixed random variable). Loại biến số ngẫu nhiên này được sử dụng trong một số chuyên ngành, chẳng hạn như trong Tài chính Định lượng, Tính toán Rủi ro Bảo hiểm. Loại biến số ngẫu nhiên này được khảo sát riêng biệt dựa vào yêu cầu của từng chuyên ngành nên thường không được đề cập đến trong các giáo trình xác suất thống kê.

BÀI TẬP

1) Cho X là biến ngẫu nhiên có phân phối chuẩn với trung bình μ chưa biết và phương sai σ^2 đã biết. Từ bộ dữ liệu nhận được của X, $x_1, x_2, \cdots, x_n$, ước lượng tham số μ bằng phương pháp MLE.

2) Cho X là biến ngẫu nhiên có phân phối Weibull như sau

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Từ bộ dữ liệu nhận được của X, $x_1, x_2, \cdots, x_n$, ước lượng tham số α, β bằng phương pháp MLE và nhận xét dạng ẩn của nghiệm.

3) Cho X là biến ngẫu nhiên có phân phối mũ với tham số λ chưa biết. Từ bộ dữ liệu nhận được của X, $x_1, x_2, \cdots, x_n$, ước lượng tham số λ bằng phương pháp MLE.

4) Độ bền cắt của mỗi mối hàn trong số mười mối hàn thử nghiệm được quan trắc và thu được dữ liệu sau (psi): 392; 376; 401; 367; 389; 362; 409; 415; 358; 375. Giả sử độ bền có phân phối chuẩn. Hãy xây dựng ước lượng MLE cho trung bình và phương sai cho độ bền mối hàn.

5) Gọi X là tỷ lệ thời gian một sinh viên được chọn ngẫu nhiên phân bổ để làm bài kiểm tra năng khiếu nhất định. Giả sử X có hàm mật độ xác suất như sau

$$f(x;\theta) = \begin{cases} (\theta+1)x^\theta, & x \in [0,1], \\ 0, & x \notin [0,1], \end{cases}$$

với $\theta > -1$. Một mẫu 10 sinh viên được khảo sát với thời gian như sau: $x_1 = 0.92$; $x_2 = 0.79$; $x_3 = 0.90$; $x_4 = 0.65$; $x_5 = 0.86$; $x_6 = 0.47$; $x_7 = 0.73$; $x_8 = 0.97$; $x_9 = 0.94$; và $x_{10} = 0.77$.

Sử dụng dữ liệu trên ước lượng tham số θ bằng phương pháp MLE.

6) Giả sử X là một biến ngẫu nhiên có hàm mật độ

$$f(x;\lambda,\theta) = \begin{cases} \lambda e^{-\lambda(x-\theta)}, & x \geq \theta, \\ 0, & x < \theta. \end{cases}$$

Với một mẫu 10 giá trị quan sát sau: 3.11, 0.64, 2.55, 2.20, 5.44, 3.42, 10.39, 8.93, 17.82 và 1.30, hãy ước lượng hai tham số λ, θ bằng phương pháp MLE.

7) Giả sử X là biến ngẫu nhiên có hàm mật độ

$$f(x) = \begin{cases} \frac{1}{2}(1+\theta x), & x \in [-1,1], \\ 0, & x \notin [-1,1]. \end{cases}$$

Hãy ước lượng tham số θ bằng phương pháp MLE từ các dữ liệu của X nhận được $x_1, x_1, \cdots, x_n$.

8) Giả sử X là biến ngẫu nhiên có hàm mật độ $f(x) = \frac{1}{2\theta^3}x^2e^{-x/\theta}$, $\theta, x > 0$. Hãy ước lượng tham số θ bằng phương pháp MLE từ các dữ liệu của X nhận được $x_1, x_1, \cdots, x_n$.

9) Giả sử X là biến ngẫu nhiên có hàm mật độ $f(x) = \theta x^{\theta-1}$, $\theta > 0, 0 < x < 1$. Hãy ước lượng tham số θ bằng phương pháp MLE từ các dữ liệu của X nhận được $x_1, x_1, \cdots, x_n$.

10) Giả sử X là biến ngẫu nhiên có hàm mật độ $f(x) = \frac{1}{\theta}x^{\frac{1-\theta}{\theta}}$, $\theta > 0, 0 < x < 1$. Hãy ước lượng tham số θ bằng phương pháp MLE từ các dữ liệu của X nhận được $x_1, x_1, \cdots, x_n$.

PHỤ LỤC

A. KIẾN THỨC CƠ BẢN CỦA ĐẠI SỐ TUYẾN TÍNH

♦ **Vectơ (n-chiều)**, $n \in \mathbb{N}$, là một **bộ n-thứ tự** (ordered n-tuple) các số thực,

$$v = (v_1, v_2, \cdots, v_n).$$

Tập hợp tất cả các bộ n-thứ tự được ký hiệu là $\mathbb{R}^n$,

$$\mathbb{R}^n = \{v = (v_1, v_2, \cdots, v_n) | v_1, v_2, \cdots, v_n \in \mathbb{R}\}.$$

Đặc biệt,

$$e_1 = (1,0,0, \cdots, 0), e_2 = (0,1,0, \cdots, 0), \cdots, e_n = (0,0,0, \cdots, 1) \in \mathbb{R}^n$$

được gọi là các **vectơ đơn vị chính tắc** (standard unit vector).

♦ **Ma trận cấp** $m \times n$, với $m, n \in \mathbb{N}$, được biểu diễn bằng bảng chữ nhật gồm $m \times n$ số thực, sắp xếp thành m hàng, n cột,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \equiv [a_{ij}],$$

trong đó a_{ij} chỉ **số hạng** (entry) nằm ở dòng i, cột j, $i = 1,2, \cdots, m$; $j = 1,2, \cdots, n$. Tập hợp tất cả các ma trận cấp $m \times n$ được ký hiệu là $\mathbb{R}^{m \times n}$.

♦ **Quan hệ** trên $\mathbb{R}^n$ và $\mathbb{R}^{m\times n}$.

Với $u = (u_1, u_2, \cdots, u_n), v = (v_1, v_2, \cdots, v_n) \in \mathbb{R}^n$,

$u = v$ nếu và chỉ nếu $u_i = v_i$, với mọi $i = 1,2,\cdots,n$,

Với $A = [a_{ij}], B = [b_{ij}] \in \mathbb{R}^{m\times n}$, $A = B$ nếu và chỉ nếu $a_{ij} = b_{ij}$, với mọi $i = 1,2,\cdots,m, j = 1,2,\cdots,n$.

♦ **Các phép toán** trên $\mathbb{R}^n$ và $\mathbb{R}^{m\times n}$.

Cho $u = (u_1, u_2, \cdots, u_n), v = (v_1, v_2, \cdots, v_n) \in \mathbb{R}^n$, và $k \in \mathbb{R}$. Các vectơ $u + v, ku \in \mathbb{R}^n$ xác định bởi

$$u + v = (u_1 + v_1, u_2 + v_2, \cdots, u_n + v_n), ku = (ku_1, ku_2, \cdots, ku_n).$$

Cho $A = [a_{ij}], B = [b_{ij}] \in \mathbb{R}^{m\times n}$, và $k \in \mathbb{R}$. Các ma trận $A + B, kA \in \mathbb{R}^{m\times n}$ xác định bởi

$$A + B = [a_{ij} + b_{ij}], kA = [ka_{ij}].$$

Tính chất. $(\mathbb{R}^n, +, \cdot)$ và $(\mathbb{R}^{m\times n}, +, \cdot)$ có cấu trúc của một **không gian vectơ** (vector space), nghĩa là, với mọi $u, v, w \in \mathbb{R}^n$, $A, B, C \in \mathbb{R}^{m\times n}$, và $h, k \in \mathbb{R}$,

(i) Tính giao hoán : $u + v = v + u$, và $A + B = B + A$.

(ii) Tính kết hợp : $u + (v + w) = (u + v) + w$, và $A + (B + C) = (A + B) + C$.

(iii) Phần tử trung hòa : $u + 0 = u$, và $A + 0 = A$, với **vectơ không** $0 = (0,0,\cdots,0) \in \mathbb{R}^n$ và **ma trận không** $0 \in \mathbb{R}^{m\times n}$, ma trận mà mọi số hạng đều là số 0.

(iv) Phần tử đối : $u + (-u) = 0$, và $A + (-A) = 0$, với $-u = (-1)u$, và $-A = (-1)A$.

(v) $h(ku) = (hk)u$, và $h(kA) = (hk)A$.

(vi) $k(u + v) = ku + kv$, và $k(A + B) = kA + kB$.

(vii) $(h + k)u = hu + ku$, và $(h + k)A = hA + kA$.

(viii) $1 \cdot u = u$, và $1 \cdot A = A$.

Cho $A = [a_{ij}] \in \mathbb{R}^{m\times n}, B = [b_{ij}] \in \mathbb{R}^{n\times p}$. Ma trận tích $C = AB = [c_{ij}] \in \mathbb{R}^{m\times p}$ xác định bởi

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj},$$

với $i = 1,2,\cdots,m$, $j = 1,2,\cdots,p$.

$$\text{row } i \rightarrow \begin{bmatrix} & & & \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ & & & \end{bmatrix} \underset{\substack{\uparrow \\ \text{column } j}}{\begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}} = \underset{\substack{\uparrow \\ \text{column } j}}{\begin{bmatrix} a_{i1}b_{ij} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} \end{bmatrix}} \leftarrow \text{row } i$$

Tính chất. Với mọi ma trận A, B, C có kích thước thích hợp trong từng biểu thức, và với mọi $k \in \mathbb{R}$, ta có

(i) Tính kết hợp : $A(BC) = (AB)C$.

(ii) Tính phân bố: $A(B + C) = AB + AC$, và $(A + B)C = AC + BC$.

(iii) $k(AB) = (kA)B = A(kB)$.

Chú ý rằng tích hai ma trận không có tính giao hoán, nghĩa là tồn tại các ma trận A, B sao cho $AB \neq BA$. Khi đẳng thức $AB = BA$ xảy ra, ta nói A và B giao hoán với nhau.

Định nghĩa. (i) Ma trận **chuyển vị** (transpose) của A, $A^T = \left[a_{ji}^T\right] \in \mathbb{R}^{n\times m}$, xác định bởi

$$a_{ji}^T = a_{ij}, \text{ với mọi } i = 1,2,\cdots,m, j = 1,2,\cdots,n.$$

(ii) Ma trận $A \in \mathbb{R}^{n\times n}$ được gọi là **khả nghịch** (invertible) khi tồn tại (duy nhất) ma trận $B \in \mathbb{R}^{n\times n}$ sao cho $AB = I = BA$, trong đó $I \in \mathbb{R}^{n\times n}$ là **ma trận đơn vị**, ma trận chéo với mọi số hạng trên đường chéo bằng 1. Khi đó, B được gọi là ma trận nghịch đảo (inverse matrix) của A, ký hiệu $B = A^{-1}$. Hiển nhiên B cũng là ma trận khả nghịch và $B^{-1} = A$.

Tính chất. (i) Với các ma trận A, B có kích thước thích hợp, ta có $(AB)^T = B^T A^T$.

(ii) Với $A, B \in \mathbb{R}^{n\times n}$ khả nghịch, ta có AB, A^T cũng là các ma trận khả nghịch, $(AB)^{-1} = B^{-1}A^{-1}$, và $(A^T)^{-1} = (A^{-1})^T$.

Bằng cách đồng nhất các vectơ $u = (u_1, u_2, \cdots, u_n) \in \mathbb{R}^n$ với **ma trận cột**,

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^{n\times 1},$$

ta định nghĩa tích của một ma trận với một vectơ, $Au \in \mathbb{R}^{m\times 1} \equiv \mathbb{R}^m$, bởi

$$Au = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} =$$

$$\begin{bmatrix} a_{11}u_1 + a_{12}u_2 + \cdots + a_{1n}u_n \\ a_{21}u_1 + a_{22}u_2 + \cdots + a_{2n}u_n \\ \vdots \\ a_{m1}u_1 + a_{m2}u_2 + \cdots + a_{mn}u_n \end{bmatrix} \in \mathbb{R}^{m\times 1}.$$

Hơn nữa, với các vectơ cột $a^j = (a_{1j}, a_{2j}, \cdots, a_{mj}) \in \mathbb{R}^m \equiv \mathbb{R}^{m\times 1}$, với $j = 1,2,\cdots,n$, ta có

$$Au = u_1a^1 + u_2a^2 + \cdots + u_na^n.$$

♦ **Hệ phương trình tuyến tính**

Phương trình tuyến tính (linear equation) theo các biến $x_1, x_2, \cdots, x_n$ có dạng

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b,$$

trong đó $a_1, a_2, \cdots, a_n$ và b là các hằng số.

Tập hữu hạn các phương trình tuyến tính được gọi là **hệ phương trình tuyến tính** (system of linear equations),

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{cases} \qquad (1)$$

trong đó a_{ij}, b_i, $i = 1,2,\cdots,m$, $j = 1,2,\cdots,n$, là các hằng số. Khi $b_i = 0$, với mọi $i = 1,2,\cdots,m$, hệ được gọi là **thuần nhất** (homogeneous),

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & 0 \end{cases}$$

Một **nghiệm** (solution) của (1) là một bộ n-thứ tự $(s_1, s_2, \cdots, s_n)$ sao cho khi thay thế $x_i = s_i$, $i = 1,2,\cdots,n$, vào hệ (1), ta nhận được n đẳng thức đúng. Tập tất cả các nghiệm của (1) được gọi là **tập nghiệm** (solution set) của nó.

Mỗi hệ phương trình tuyến tính có thể biểu diễn bằng **ma trận bổ sung** (augmented matrix) chứa đầy đủ các thông tin của hệ, các a_{ij} cũng như các b_i. Chẳng hạn, ma trận bổ sung của hệ (1) cho bởi

$$\bar{A} = [A \quad b] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix} \in \mathbb{R}^{m\times(n+1)}.$$

Ngoài ra, với

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \equiv [a^1 \quad a^2 \quad \cdots \quad a^n],$$

trong đó

$$a^j = (a_{1j}, a_{2j}, \cdots, a_{mj}) \equiv \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}, j = 1,2,\cdots,n,$$

là các vec tơ cột của A, vectơ

$$b = (b_1, b_2, \cdots, b_m) \equiv \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

các hệ số vế phải, và vectơ các ẩn số,

$$x = (x_1, x_2, \cdots, x_n) \equiv \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

hệ phương trình tuyến tính

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{cases}$$

tương đương với phương trình ma trận

$$Ax = b$$

cũng như với phương trình vectơ

$$x_1a^1 + x_2a^2 + \cdots + x_na^n = b.$$

♦ **Không gian con**

Tập con không rỗng $V \subset \mathbb{R}^n$ được gọi là một **không gian con** của $\mathbb{R}^n$ khi với mọi $u, v \in V$, và $k \in \mathbb{R}$, ta có $u + v, ku \in V$. Hiển nhiên $\mathbb{R}^n$ là một không gian con của chính nó. Ngoài ra, ta có hai không gian con quan trọng sau:

(i) Không gian nghiệm của một hệ phương trình tuyến tính thuần nhất, $Ax = 0$, với $A \in \mathbb{R}^{m\times n}$.

(ii) Không gian sinh bởi hệ các vectơ $B = \{u_1, u_2, \cdots, u_r\} \subset \mathbb{R}^n$,

$$span(B) \equiv span\{u_1, u_2, \cdots, u_r\} = \{k_1u_1 + k_2u_2 + \cdots + k_ru_r | k_1, k_2, \cdots, k_r \in \mathbb{R}\},$$

tập hợp các **tổ hợp tuyến tính** (linear combination) các vectơ của B.

♦ **Hệ độc lập tuyến tính – hệ sinh – cơ sở**

Cho V là một không gian con của $\mathbb{R}^n$. Xét hệ các vectơ $B = \{u_1, u_2, \cdots, u_r\} \subset V$

B được gọi là **độc lập tuyến tính** khi với mọi $x_1, x_2, \cdots, x_r \in \mathbb{R}$, phương trình vectơ $x_1u_1 + x_2u_2 + \cdots + x_ru_r = 0$ chỉ có nghiệm tầm thường $x_1 = x_2 = \cdots = x_r = 0$.

B được gọi là **sinh** ra V khi $span(B) = V$, i.e., ứng với mọi $v \in V$, tồn tại $x_1, x_2, \cdots, x_r \in \mathbb{R}$ sao cho $v = x_1u_1 + x_2u_2 + \cdots + x_ru_r$.

B được gọi là một **cơ sở** của V khi nó độc lập tuyến tính và sinh ra V. Bấy giờ, ta có

(i) Mọi cơ sở khác của V đều phải có đúng r vectơ. Ta nói V là một không gian vectơ r-chiều, ký hiệu $dim(V) = r$.

(ii) Ứng với mỗi $v \in V$, tồn tại duy nhất $x_1, x_2, \cdots, x_r \in \mathbb{R}$ sao cho $v = x_1u_1 + x_2u_2 + \cdots + x_ru_r$. Các x_i được gọi là **tọa độ** của v đối với cơ sở B, ký hiệu

$$[v]_B = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix} \in \mathbb{R}^{r\times 1}.$$

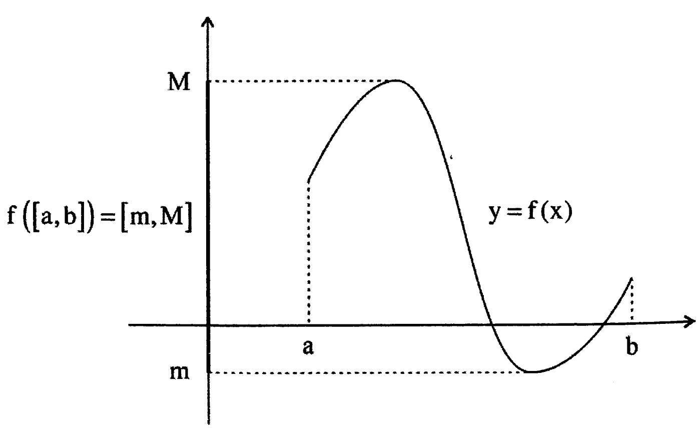
Đặc biệt, hệ $S = \{e_1 = (1,0,\cdots,0), e_2 = (0,1,\cdots,0),\cdots, e_n = (0,0,\cdots,1)\}$ là một cơ sở của $\mathbb{R}^n$, mà ta gọi là cơ sở chính tắc (standard basis). Do đó, $dim(\mathbb{R}^n) = n$, và đối với cơ sở chính tắc của $\mathbb{R}^n$, tọa độ của vectơ $u = (u_1, u_2, \cdots, u_n)$ là

$$[u]_S \equiv u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}.$$

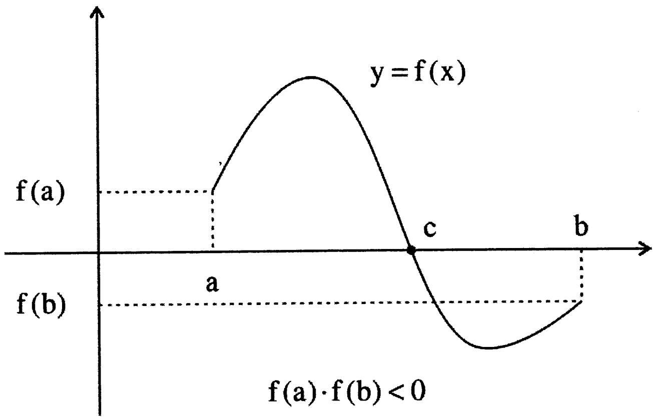
B. PHÉP TÍNH VI TÍCH PHÂN

♦ **Định lý**. Với mọi hàm liên tục $f: [a,b] \to \mathbb{R}$, ta có $f([a,b]) = [m, M]$, với

$$m = \min_{x\in[a,b]} f(x) \text{ và } M = \min_{x\in[a,b]} f(x).$$



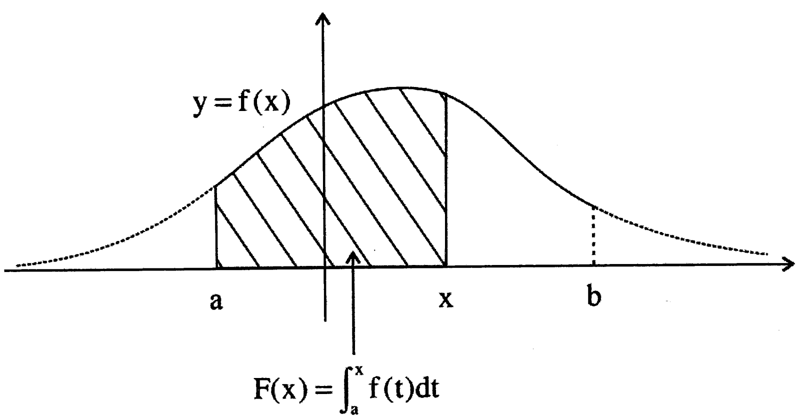
♦ **Hệ quả**. Cho $f:[a,b] \to \mathbb{R}$ là một hàm liên tục. Nếu $f(a)\cdot f(b) < 0$, tồn tại $c \in (a,b)$ sao cho $f(c) = 0$.



♦ **Định lý cơ bản của phép tính vi tích phân**. Với hàm liên tục $f:\mathbb{R} \to \mathbb{R}$ và với $a \in \mathbb{R}$, xét hàm số $F:\mathbb{R} \to \mathbb{R}$ xác định bởi

$$F(x) = \int_a^x f(t)dt.$$

Ta có F là hàm khả vi và $F'(x) = f(x)$, với mọi $x \in \mathbb{R}$.



♦ **Công thức Taylor**. Cho $f: \mathbb{R} \to \mathbb{R}$ là hàm thuộc lớp C^n, nghĩa là f có đạo hàm liên tục tới cấp n. Với mọi $x, y \in \mathbb{R}$, ta có

$$f(y) = f(x) + f'(x)(y-x) + f''(x)\frac{(y-x)^2}{2!} + \cdots +$$
$$f^{(n-1)}(x)\frac{(y-x)^{n-1}}{(n-1)!} + \int_x^y \frac{(y-t)^{n-1}}{(n-1)!} f^{(n)}(t)dt.$$

Chứng minh. Với $n = 1$, (1) được viết lại thành

$$f(y) = f(x) + \int_x^y \frac{(y-t)^{1-1}}{(1-1)!} f^{(1)}(t)dt = f(x) + \int_x^y f'(t)dt.$$

Do định lý cơ bản của phép tính vi tích phân, với mỗi $x \in \mathbb{R}$, và với hàm

$$F(y) = \int_x^y f'(t)dt, y \in \mathbb{R},$$

ta có $F'(y) = f'(y)$, với mọi $y \in \mathbb{R}$. Do đó, tồn tại hằng số $C \in \mathbb{R}$ sao cho $F(y) = f(y) + C$, với mọi $y \in \mathbb{R}$. Mặt khác, do $F(x) = 0$ nên ta được $C = -f(x)$ vá do đó (2) được chứng minh.

Bây giờ, khi (1) đúng với $n \in \mathbb{N}$, dùng công thức tích phân từng phần với

$$\begin{cases} u = f^{(n)}(t), \\ dv = \dfrac{(y-t)^{n-1}}{(n-1)!} dt \end{cases}$$

ta được

$$\begin{cases} du = f^{(n+1)}(t)dt, \\ v = -\frac{(y-t)^n}{(n-1)! \cdot n} dt \end{cases},$$

và do đó

$$\int_x^y \frac{(y-t)^{n-1}}{(n-1)!} f^{(n)}(t)dt = -\frac{(y-t)^n}{n!} f^{(n)}(t)\Big|_x^y + \int_x^y \frac{(y-t)^n}{n!} f^{(n+1)}(t)dt$$

$$= \frac{(y-x)^n}{n!} f^{(n)}(x) + \int_x^y \frac{(y-t)^n}{n!} f^{(n+1)}(t)dt,$$

nghĩa là (1) cũng đúng cho trường hợp $n+1$.

♦ **Công thức Taylor với phần dư vi phân**. Với điều kiện của công thức Taylor, tồn tại $z \in [x, y]$ sao cho

$$f(y) = f(x) + f'(x)(y-x) + f''(x)\frac{(y-x)^2}{2!} + \cdots + f^{(n-1)}(x)\frac{(y-x)^{n-1}}{(n-1)!} + f^{(n)}(z)\frac{(y-x)^n}{n!}.$$

Chứng minh. Với $m = \min_{z\in[x,y]} f^{(n)}(z)$ và $M = \max_{z\in[x,y]} f^{(n)}(z)$, ta có

$$m\int_x^y \frac{(y-t)^{n-1}}{(n-1)!}dt \le \int_x^y \frac{(y-t)^{n-1}}{(n-1)!} f^{(n)}(t)dt \le M\int_x^y \frac{(y-t)^{n-1}}{(n-1)!}dt.$$

Do

$$\int_x^y \frac{(y-t)^{n-1}}{(n-1)!}dt = -\frac{(y-t)^n}{n!}\Big|_x^y = \frac{(y-x)^n}{n!}$$

ta suy ra

$$m \le \frac{n!}{(y-x)^n}\int_x^y \frac{(y-t)^{n-1}}{(n-1)!} f^{(n)}(t)dt \le M$$

và do định lý giá trị trung gian, tồn tại $z \in [x, y]$ sao cho

$$\frac{n!}{(y-x)^n}\int_x^y \frac{(y-t)^{n-1}}{(n-1)!} f^{(n)}(t)dt = f^{(n)}(z),$$

và do đó

$$\int_x^y \frac{(y-t)^{n-1}}{(n-1)!} f^{(n)}(t)dt = f^{(n)}(z)\frac{(y-x)^n}{n!}. \qquad \blacksquare$$

C. XÁC SUẤT THỐNG KÊ

♦ Xét một hiện tượng ngẫu nhiên có **không gian mẫu** (sample space) $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$, tập hợp tất cả các **kết quả** (outcome) có thể xảy ra của hiện tượng ngẫu nhiên. Phép gán $A \subset \Omega \mapsto P(A)$ được gọi là **xác suất** (probability) trên Ω nếu

(i) Với mọi A, $P(A) \geq 0$,

(ii) $P(\Omega) = 1$,

(iii) Với mọi dãy các biến cố $A_1, A_2, A_3, \cdots$ xung khắc từng đôi (mutually exclusive), i.e., $A_i \cap A_j = \emptyset$ khi $i \neq j$, ta có

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i).$$

♦ **Tính chất**. (i) $P(\emptyset) = 0$.

(ii) Với mọi dãy hữu hạn các biến cố $A_1, A_2, \cdots, A_n$ xung khắc từng đôi (mutually exclusive), i.e., $A_i \cap A_j = \emptyset$ khi $i \neq j$, ta có

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

(iii) $P(\bar{A}) = 1 - P(A)$, trong đó $\bar{A} = \Omega \backslash A$ là biến cố đối của biến cố A.

(iv) Công thức cộng: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

♦ Xác suất để biến cố A xảy ra, khi biết biến cố B đã xảy ra

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ với } P(B) > 0.$$

♦ **Tính chất**. Cho P là một xác suất xác định trên một không gian mẫu Ω. Xét biến cố B, với $P(B) > 0$. Phép gán $A \subset \Omega \mapsto P(A|B)$ cũng là một xác suất trên Ω, i.e.,

(i) Với mọi A, $P(A|B) \geq 0$,

(ii) $P(\Omega|B) = 1$,

(iii) Với mọi dãy các biến cố $A_1, A_2, A_3, \cdots$ xung khắc từng đôi (mutually exclusive), i.e., $A_i \cap A_j = \emptyset$ khi $i \neq j$, ta có

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots |B) = \sum_{i=1}^{\infty} P(A_i|B).$$

♦ **Tính chất**. (i) $P(\emptyset|B) = 0$.

(ii) Với mọi dãy hữu hạn các biến cố $A_1, A_2, \cdots, A_n$ xung khắc từng đôi (mutually exclusive), i.e., $A_i \cap A_j = \emptyset$ khi $i \neq j$, ta có

$$P(\cup_{i=1}^{n} A_i|B) = \sum_{i=1}^{n} P(A_i|B).$$

(iii) $P(\bar{A}|B) = 1 - P(A|B)$, trong đó $\bar{A} = \Omega \backslash A$ là biến cố đối của biến cố A.

(iv) Công thức cộng: $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$.

(v) Công thức nhân: $P(A \cap B) = P(A|B) \cdot P(B)$. Tổng quát, với mọi dãy hữu hạn các biến cố $A_1, A_2, \cdots, A_n$, ta có

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1|A_2 \cdots A_n) P(A_2|A_3 \cap \cdots \cap A_n) \cdots P(A_{n-1}|A_n) P(A_n).$$

♦ **Công thức xác suất toàn phần**

Cho $A_1, \cdots, A_k$ là một họ **đầy đủ** các biến cố, i.e., i.e., $A_i \cap A_j = \emptyset$ khi $i \neq j$, và $A_1 \cup A_2 \cup \cdots \cup A_k = \Omega$. Với mọi biến cố B,

$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

♦ **Công thức Bayes**

Cho $A_1, \cdots, A_k$ là một họ **đầy đủ** các biến cố, i.e., i.e., $A_i \cap A_j = \emptyset$ khi $i \neq j$, và $A_1 \cup A_2 \cup \cdots \cup A_k = \Omega$, với các **xác suất tiên nghiệm** (prior probabilities). Với mọi biến cố B, với $P(B) > 0$, **xác suất hậu nghiệm** (posterior probability) của A_j khi biết B xảy ra cho bởi

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}, \quad j = 1, \cdots, k.$$

TÀI LIỆU THAM KHẢO

[1] Jeffrey Holt, *Linear Algebra with Applications*, W.H. Freeman and Company, 2013.

[2] Stephen Boyd, Lieven Vandenberghe, *Introduction to Applied Linear Algebra, Vectors, Matrices, and Least Squares*, Cambridge University Press, 2018.

[3] Howars Anton, Chris Rorres, *Elementary Linear Algebra – Applications Version*, 11th edition, Wiley, 2014.

[4] Stephen Boyd, Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[5] James Ramsay, Giles Hooker, *Dynamic Data Analysis, Modeling Data with Differential Equations*, Springer Series in Statistics, Springer Science+Business Media LLC 2017.

[6] R. Kent Nagle, Edward B. Saff, Arthur David Snider, *Fundamentals of Differential Equations and Boundary Value Problems*, Sixth Edition, Pearson Education, Inc., 2012.

[7] Douglas C. Mongemery, George C. Runger, *Applied Statistics and Probability for Engineers*, Third Edition, John Wiley & Sons, Inc, 2013.

[8] Đặng Đức Trọng, Đinh Ngọc Thanh, *Lý thuyết thống kê*, NXB ĐHQG TP. HCM, 2016.

[9] Đinh Ngọc Thanh, Đặng Đức Trọng, Nguyễn Văn Thìn, Phan Phúc Doãn, Nguyễn Thị Hồng Nhung, Nguyễn Thị Hiên, Nguyễn Thị Nhàn, *Bài tập & Thực hành Lý thuyết thống Kê*, NXB ĐHQG TP. HCM, 2016.

[10] Đặng Đức Trọng, Đinh Ngọc Thanh, *Lý thuyết Độ đo và xác suất*, NXB ĐHQG TP. HCM, 2016.

CƠ SỞ TOÁN CHO KHOA HỌC DỮ LIỆU

Nguyễn Thanh Bình (chủ biên)
Đinh Ngọc Thanh, Nguyễn Đình Thúc, Nguyễn Đăng Minh
Trường Đại học Khoa học Tự nhiên,
Đại học Quốc gia Thành Phố Hồ Chí Minh

NHÀ XUẤT BẢN ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH

Phòng 501, Nhà Điều hành ĐHQG-HCM, P. Linh Trung, TP Thủ Đức, TP.HCM.
ĐT: 028 62726361
E-mail: vnuhp@vnuhcm.edu.vn
Website: vnuhcmpress.edu.vn

Chịu trách nhiệm xuất bản và nội dung
PGS.TS NGUYỄN MINH TÂM

Biên tập
TRẦN THỊ ĐỨC LINH

Sửa bản in
ÁI NHẬT

Trình bày bìa
CÔNG TY TNHH MTV IN KINH TẾ

Đối tác liên kết
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN (ĐHQG-HCM)

Tái bản lần thứ 1. Số lượng in: 1.000 cuốn, khổ 14.5 x 20.5cm. Số XNĐKXB: 1456-2024/CXBIPH/50-13/ĐHQGTPHCM. QĐXB số: 10/QĐ-NXB-TB cấp ngày 08/5/2024. In tại: Công ty TNHH MTV In Kinh Tế. Địa chỉ: 279 Nguyễn Tri Phương, phường 5, quận 10, TP.HCM. Nộp lưu chiểu: Năm 2024. ISBN: **978-604-479-567-6.**