Chương 2

Đại số tuyến tính

Đại số tuyến tính là một nhánh của toán học được sử dụng rộng rãi trong khoa học và kỹ thuật. Tuy nhiên, vì đại số tuyến tính là một dạng toán học liên tục chứ không phải toán rời rạc, nên nhiều nhà khoa học máy tính có ít kinh nghiệm với nó. Hiểu biết tốt về đại số tuyến tính là rất cần thiết để hiểu và làm việc với nhiều thuật toán học máy, đặc biệt là các thuật toán học sâu. Do đó, trước khi giới thiệu về học máy, ta cần tập trung vào các kiến thức nền tảng quan trọng về đại số tuyến tính.

Nếu bạn đã quen thuộc với đại số tuyến tính, hãy thoải mái bỏ qua chương này. Nếu bạn đã có kinh nghiệm với các khái niệm này nhưng cần một tài liệu tham khảo chi tiết để ôn lại các công thức chính, bạn có thể đọc cuốn *The Matrix Cookbook* (Petersen và Pedersen, 2006, [1]). Nếu bạn chưa từng tiếp xúc với đại số tuyến tính, chương này sẽ dạy bạn đủ để đọc cuốn sách này, nhưng bạn nên tham khảo thêm các tài liệu khác chỉ tập trung vào việc giảng dạy đại số tuyến tính, chẳng hạn như *Linear Algebra* (Shilov, 1977, [2]). Chương này sẽ bỏ qua nhiều chủ đề quan trọng của đại số tuyến tính không cần thiết để hiểu học máy.

2.1 Đại lượng vô hướng, véctơ, ma trận, và tenxơ

Một số đối tượng nghiên cứu của đại số tuyến tính:

- **Đại lượng vô hướng**: Đại lượng vô hướng đơn giản chỉ là một số đơn lẻ, khác với hầu hết các đối tượng khác được nghiên cứu trong đại số tuyến tính, thường là các mảng chứa nhiều số. Ta hay viết các số vô hướng bằng chữ cái thường, in nghiêng. Khi giới thiệu, ta sẽ chỉ rõ loại số của chúng. Ví dụ, ta nói "Giả sử $s \in \mathbb{R}$ là độ dốc của đường thẳng", khi định nghĩa một số vô hướng có giá trị thực, hoặc "Giả sử $n \in \mathbb{N}$ là số lượng dữ liệu", khi định nghĩa một số vô hướng tự nhiên.

- **Véctơ**: Véctơ là một mảng gồm các số. Các số này được sắp xếp theo thứ tự. Ta có thể xác định từng số riêng lẻ bằng chỉ số của nó trong thứ tự đó. Ta hay đặt tên cho các véctơ bằng chữ cái thường, in đậm, chẳng hạn như $\boldsymbol{x}$. Các phần tử của véctơ được xác định bằng cách viết tên của nó dưới dạng chữ in nghiêng, kèm theo chỉ số dưới. Phần tử đầu tiên của $\boldsymbol{x}$ là x_1, phần tử thứ hai là x_2, và cứ tiếp tục như vậy. Ta cũng cần nói rõ loại số nào được lưu trữ trong véctơ. Nếu mỗi phần tử nằm trong $\mathbb{R}$, và véctơ có n phần tử, thì véctơ nằm trong tập hợp được tạo bởi tích Descartes của $\mathbb{R}$ với chính nó n lần, ký hiệu là $\mathbb{R}^n$. Khi cần xác định rõ ràng các phần tử của véctơ, ta viết chúng dưới dạng một cột được bao quanh bởi dấu ngoặc vuông:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. \tag{2.1}$$

Ta có thể xem véctơ như là xác định các điểm trong không gian, với mỗi phần tử cung cấp tọa độ theo một trục khác nhau. Đôi khi, ta cần truy cập một tập hợp các phần tử của véctơ. Trong trường hợp này, ta định nghĩa một tập chứa các chỉ số và viết tập hợp đó dưới dạng chỉ số dưới. Ví dụ, để truy cập x_1, x_3 và x_6, ta định nghĩa tập hợp $S = \{1, 3, 6\}$ và viết là $\boldsymbol{x}_S$. Ta sử dụng dấu "$-$" để chỉ phần bù của một tập hợp. Ví dụ, $\boldsymbol{x}_{-1}$ là véctơ chứa tất cả các phần tử của $\boldsymbol{x}$ ngoại trừ x_1, và $\boldsymbol{x}_{-S}$ là véctơ chứa tất cả các phần tử của x ngoại trừ x_1, x_3 và x_6.

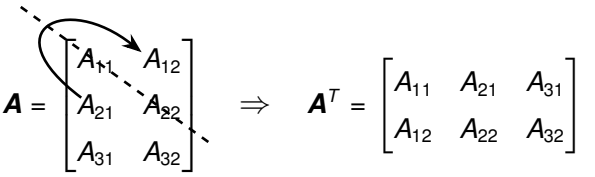
- **Ma trận**: Ma trận là một mảng 2 chiều gồm các số, vì vậy mỗi phần tử được xác định bởi hai chỉ số thay vì chỉ một. Ta thường đặt tên cho ma trận bằng các chữ cái viết hoa và in đậm, chẳng hạn như $\boldsymbol{A}$. Nếu ma trận $\boldsymbol{A}$ các phần tử giá trị thực có cỡ $m \times n$, tức là có m hàng và n cột, , thì ta nói rằng $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. Ta thường xác định các phần tử của ma trận bằng cách sử dụng tên của nó ở dạng in nghiêng nhưng không in đậm, và các chỉ số được liệt kê có dấu phẩy ngăn cách, hoặc có thể bỏ dấu phẩy nếu không gây nhầm lẫn. Ví dụ, $A_{1,1}$, hay A_{11}, là phần tử ở góc trên bên trái của $\boldsymbol{A}$, được đánh dấu hàng 1 cột 1, và $A_{m,n}$, hay A_{mn}, là phần tử ở góc dưới bên phải, ở hàng m cột n. Ta có thể xác định tất cả các số ở hàng i bằng cách viết dấu ":" thay cho chỉ số cột. Ví dụ, $\boldsymbol{A}_{i,:}$ biểu thị hàng ngang thứ i của ma trận $\boldsymbol{A}$. Tương tự, $\boldsymbol{A}_{:,j}$ là cột thứ j của $\boldsymbol{A}$. Khi cần xác định rõ ràng các phần tử của ma trận, ta viết chúng dưới dạng mảng và đặt trong dấu ngoặc vuông:

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} \tag{2.2}$$

  Đôi khi ta có thể cần đánh chỉ số các biểu thức nhận giá trị ma trận không chỉ là một chữ cái. Trong trường hợp này, ta sử dụng chỉ số dưới bên phải biểu thức, nhưng không chuyển bất kỳ thứ gì thành chữ thường. Ví dụ, $f(\boldsymbol{A})_{ij}$ cho phần tử (i, j) của ma trận được tính bằng cách áp dụng hàm số f lên $\boldsymbol{A}$.

- **Tenxơ**: Trong một số trường hợp, ta cần một mảng với nhiều hơn hai trục. Trong trường hợp tổng quát, một mảng số được sắp xếp trên một lưới đều với số lượng trục thay đổi được gọi là tenxơ. Ta ký hiệu một tenxơ có tên là "A" bằng kiểu chữ này: $\mathcal{A}$. Ta xác định phần tử của $\mathcal{A}$ tại tọa độ (i, j, k) bằng cách viết $\mathcal{A}_{ijk}$.

Một phép toán quan trọng trên ma trận là **chuyển vị**. Chuyển vị của một ma trận là hình ảnh phản chiếu của ma trận qua một đường chéo, gọi là **đường chéo chính**, chạy từ trên xuống dưới và sang phải, bắt đầu từ góc trên bên trái. Hình 2.1 mô tả trực quan phép toán này.



Hình 2.1: Chuyển vị của ma trận có thể được xem như là hình ảnh phản chiếu qua đường chéo chính.

Ký hiệu chuyển vị của ma trận $\boldsymbol{A}$ là $\boldsymbol{A}^T$, và nó được định nghĩa như sau

$$\left(\boldsymbol{A}^T\right)_{ij} = A_{ji} \tag{2.3}$$

Véctơ có thể được xem như ma trận chỉ chứa một cột. Do đó, chuyển vị của một véctơ là một ma trận chỉ có một hàng. Đôi khi, ta định nghĩa một véctơ bằng cách viết các phần tử của nó dưới dạng một ma trận hàng, sau đó sử dụng phép toán chuyển vị để biến nó thành một véctơ cột dạng chuẩn, ví dụ: $\boldsymbol{x} = [x_1, x_2, x_3]^T$, hoặc $\boldsymbol{x} = (x_1, x_2, x_3)^T$.

Một đại lượng vô hướng có thể được xem như một ma trận chỉ có một phần tử. Từ đó, ta có thể thấy rằng một đại lượng vô hướng chính là chuyển vị của chính nó: $a = a^T$.

Ta có thể cộng các ma trận với nhau, miễn là chúng có cùng cỡ, bằng cách cộng các phần tử tương ứng của chúng: $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$, trong đó $C_{ij} = A_{ij} + B_{ij}$.

Ta cũng có thể cộng một đại lượng vô hướng với ma trận hoặc nhân ma trận với đại lượng vô hướng, bằng cách thực hiện phép toán đó trên từng phần tử của ma trận: $\boldsymbol{D} = a \cdot \boldsymbol{B} + c$, trong đó $D_{ij} = a \cdot B_{ij} + c$.

Trong ngữ cảnh học sâu, ta cũng sử dụng một số ký hiệu ít tính quy ước hơn. Ta cho phép cộng ma trận với véctơ, tạo ra một ma trận khác: $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{b}$, trong đó $C_{ij} = A_{ij} + b_j$. Nói cách khác, véctơ $\boldsymbol{b}$ được cộng vào từng hàng của ma trận. Cách viết tắt này loại bỏ nhu cầu phải định nghĩa một ma trận với $\boldsymbol{b}$ được sao chép vào mỗi hàng trước khi thực hiện phép cộng. Việc sao chép ngầm $\boldsymbol{b}$ vào nhiều vị trí này được gọi là **phát sóng** (broadcasting).

2.2 Nhân các véctơ và các ma trận

Một trong những phép toán quan trọng nhất liên quan đến ma trận là phép nhân hai ma trận. Tích ma trận của các ma trận $\boldsymbol{A}$ và $\boldsymbol{B}$ là một ma trận thứ ba $\boldsymbol{C}$. Để phép nhân này được xác định, ma trận $\boldsymbol{A}$ phải có số cột bằng với số hàng của ma trận $\boldsymbol{B}$. Nếu $\boldsymbol{A}$ có cỡ $m \times n$ và $\boldsymbol{B}$ có cỡ $n \times p$, thì $\boldsymbol{C}$ sẽ có cỡ $m \times p$. Ta viết tích ma trận bằng cách đặt hai hoặc nhiều ma trận cạnh nhau, ví dụ

$$\boldsymbol{C} = \boldsymbol{AB}. \tag{2.4}$$

Phép toán nhân được định nghĩa bởi

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}. \tag{2.5}$$

Lưu ý rằng tích chuẩn của hai ma trận không chỉ đơn thuần là một ma trận chứa tích của các phần tử riêng lẻ. Phép toán như vậy cũng có và được gọi là **tích từng phần tử** hay **tích Hadamard**, và được ký hiệu là $\boldsymbol{A} \odot \boldsymbol{B}$.

Tích vô hướng giữa hai véctơ $\boldsymbol{x}$ và $\boldsymbol{y}$ có cùng số chiều là tích ma trận $\boldsymbol{x}^T\boldsymbol{y}$. Ta có thể xem tích ma trận $\boldsymbol{C} = \boldsymbol{AB}$ bằng cách tính $C_{i,j}$ là tích vô hướng giữa hàng thứ i của $\boldsymbol{A}$ và cột thứ j của $\boldsymbol{B}$.

Các phép toán nhân ma trận có nhiều tính chất hữu ích giúp cho việc phân tích toán học các ma trận trở nên thuận tiện hơn. Ví dụ, phép nhân ma trận có tính chất kết hợp:

$$(\boldsymbol{AB})\,\boldsymbol{C} = \boldsymbol{A}\,(\boldsymbol{BC})\,. \tag{2.6}$$

Nó cũng có tính chất phân phối:

$$\begin{aligned} \boldsymbol{A}\,(\boldsymbol{B} + \boldsymbol{C}) &= \boldsymbol{AB} + \boldsymbol{AC} \\ (\boldsymbol{A} + \boldsymbol{B})\,\boldsymbol{C} &= \boldsymbol{AC} + \boldsymbol{BC}. \end{aligned} \tag{2.7}$$

Tuy nhiên, phép nhân ma trận không có tính chất giao hoán (điều kiện $\boldsymbol{AB} = \boldsymbol{BA}$ không phải lúc nào cũng đúng), khác với phép nhân vô hướng. Tuy nhiên, tích vô hướng giữa hai véctơ thì có tính giao hoán:

$$\boldsymbol{x}^T\boldsymbol{y} = \boldsymbol{y}^T\boldsymbol{x} \tag{2.8}$$

Chuyển vị của tích ma trận có một tính chất đơn giản:

$$(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T. \tag{2.9}$$

Điều này cho phép ta chứng minh đẳng thức (2.8), bằng cách sử dụng kết luận rằng giá trị của phép nhân này là một vô hướng và do đó bằng với chuyển vị của chính nó:

$$\boldsymbol{x}^T \boldsymbol{y} = \left(\boldsymbol{x}^T \boldsymbol{y}\right)^T = \boldsymbol{y}^T \boldsymbol{x}$$

Vì trọng tâm của cuốn sách này không phải là đại số tuyến tính, chúng ta không cố gắng liệt kê đầy đủ các tính chất hữu ích của tích ma trận tại đây, nhưng người đọc nên biết rằng còn nhiều tính chất khác.

Giờ ta đã biết đủ ký hiệu đại số tuyến tính để viết ra một hệ phương trình tuyến tính:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{2.10}$$

trong đó $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ là ma trận đã biết, $\boldsymbol{b} \in \mathbb{R}^m$ là véctơ đã biết, và $\boldsymbol{x} \in \mathbb{R}^n$ là véctơ chứa các biến chưa biết mà ta cần giải. Mỗi phần tử x_i của $\boldsymbol{x}$ là một trong những biến chưa biết này. Mỗi hàng của $\boldsymbol{A}$ và mỗi phần tử của $\boldsymbol{b}$ cung cấp một ràng buộc. Ta có thể viết lại phương trình (2.10) thành:

$$\begin{aligned} \boldsymbol{A}_{1,:}\boldsymbol{x} &= b_1 \\ \boldsymbol{A}_{2,:}\boldsymbol{x} &= b_2 \\ &\dots \\ \boldsymbol{A}_{m,:}\boldsymbol{x} &= b_m \end{aligned} \tag{2.11}$$

hoặc chi tiết hơn là

$$\begin{array}{ccccccccc} A_{11}x_1 & + & A_{12}x_2 & + & \cdots & + & A_{1n}x_n & = & b_1 \\ A_{21}x_1 & + & A_{22}x_2 & + & \cdots & + & A_{2n}x_n & = & b_2 \\ & & \multicolumn{5}{c}{\dots\dots\dots\dots\dots\dots} & & \\ A_{m1}x_1 & + & A_{m2}x_2 & + & \cdots & + & A_{mn}x_n & = & b_m \end{array} \tag{2.12}$$

Ký hiệu tích ma trận – véctơ cung cấp một cách biểu diễn gọn gàng hơn cho các phương trình có dạng này.

2.3 Ma trận đơn vị và ma trận nghịch đảo

Đại số tuyến tính cung cấp một công cụ mạnh mẽ gọi là **phép nghịch đảo ma trận**, cho phép ta giải tích phương trình (2.10) cho nhiều giá trị của $\boldsymbol{A}$.

Để mô tả phép nghịch đảo ma trận, trước tiên ta cần định nghĩa khái niệm ma trận đơn vị. Ma trận đơn vị là ma trận không làm thay đổi bất kỳ véctơ nào khi ta

nhân véctơ đó với ma trận này. Ta nói ma trận đơn vị bảo toàn các véctơ n chiều và ký hiệu là $\boldsymbol{I}_n$. Như vậy, $I_n \in \mathbb{R}^{n\times n}$, và

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \quad \boldsymbol{I}_n \boldsymbol{x} = \boldsymbol{x}. \tag{2.13}$$

Cấu trúc của ma trận đơn vị rất đơn giản: tất cả các phần tử trên đường chéo chính đều bằng 1, trong khi tất cả các phần tử khác đều bằng 0, như hình dưới đây:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Hình 2.2: Ví dụ về **ma trận đơn vị**: Đây là $\boldsymbol{I}_3$

Với ma trận $\boldsymbol{A}$ cỡ $m \times n$, nghịch đảo của $\boldsymbol{A}$ được ký hiệu là $\boldsymbol{A}^{-1}$, và nó được định nghĩa là ma trận sao cho

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n \tag{2.14}$$

Ta có thể giải phương trình (2.10) theo các bước sau:

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} &= \boldsymbol{b} \\ \boldsymbol{A}^{-1}(\boldsymbol{A}\boldsymbol{x}) &= \boldsymbol{A}^{-1}\boldsymbol{b} \\ \left(\boldsymbol{A}^{-1}\boldsymbol{A}\right)\boldsymbol{x} &= \boldsymbol{A}^{-1}\boldsymbol{b} \\ \boldsymbol{I}_n\boldsymbol{x} &= \boldsymbol{A}^{-1}\boldsymbol{b} \\ \boldsymbol{x} &= \boldsymbol{A}^{-1}\boldsymbol{b} \end{aligned} \tag{2.15}$$

Tất nhiên, quá trình này phụ thuộc vào khả năng tìm được $\boldsymbol{A}^{-1}$. Ta sẽ thảo luận về các điều kiện để $\boldsymbol{A}^{-1}$ tồn tại trong phần sau.

Khi $\boldsymbol{A}^{-1}$ tồn tại, có một số thuật toán để tìm nó dưới dạng đóng. Về lý thuyết, cùng một ma trận nghịch đảo có thể được sử dụng để giải phương trình nhiều lần cho các giá trị khác nhau của $\boldsymbol{b}$. Tuy nhiên, $\boldsymbol{A}^{-1}$ chủ yếu hữu ích như một công cụ lý thuyết và không nên thực sự được sử dụng trong hầu hết các phần mềm. Vì $\boldsymbol{A}^{-1}$ chỉ có thể được biểu diễn với độ chính xác giới hạn trên máy tính, các thuật toán sử dụng cả giá trị của $\boldsymbol{b}$ thường có thể đưa ra ước tính chính xác hơn về $\boldsymbol{x}$, chẳng hạn **thuật toán Gauss – Seidel**.

2.4 Độc lập tuyến tính và không gian bao tuyến tính

Để $\boldsymbol{A}^{-1}$ tồn tại, phương trình (2.10) phải có đúng một nghiệm cho mọi giá trị của $\boldsymbol{b}$. Tuy nhiên, cũng có thể hệ phương trình vô nghiệm hoặc có vô số nghiệm cho một số giá trị của $\boldsymbol{b}$. Không thể có trường hợp phương trình có nhiều hơn một nghiệm nhưng ít hơn vô số nghiệm cho một giá trị cụ thể của $\boldsymbol{b}$; vì nếu cả $\boldsymbol{x}$ và $\boldsymbol{y}$ đều là nghiệm, thì

$$\boldsymbol{z} = \alpha\boldsymbol{x} + (1 - \alpha)\,\boldsymbol{y} \tag{2.16}$$

cũng là một nghiệm với số thực α bất kỳ.

Để phân tích phương trình có bao nhiêu nghiệm, ta có thể xem các cột của $\boldsymbol{A}$ chỉ các hướng khác nhau mà ta có thể di chuyển từ gốc tọa độ (điểm được xác định bởi véctơ gồm toàn số 0), và xác định có bao nhiêu cách để đạt đến $\boldsymbol{b}$. Theo cách nhìn này, mỗi phần tử của $\boldsymbol{x}$ xác định ta nên di chuyển bao xa theo từng hướng, với x_i xác định khoảng cách di chuyển theo hướng của cột thứ i:

$$\boldsymbol{Ax} = \sum_{i=1}^{n} x_i \boldsymbol{A}_{:,i} \tag{2.17}$$

Nhìn chung, phép toán kiểu này được gọi là **tổ hợp tuyến tính**. Tổng quát, một tổ hợp tuyến tính của một tập hợp các véctơ $\{\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(n)}\}$ được xác định bằng cách nhân mỗi véctơ $\boldsymbol{v}^{(i)}$ với một hệ số vô hướng tương ứng và cộng các kết quả lại:

$$\sum_{i=1}^{n} c_i \boldsymbol{v}^{(i)} \tag{2.18}$$

Bao tuyến tính sinh bởi một tập các véctơ là tập hợp tất cả các điểm có thể đạt được bằng cách lấy tổ hợp tuyến tính của các véctơ ban đầu.

Việc xác định liệu phương trình $\boldsymbol{Ax} = \boldsymbol{b}$ có nghiệm hay không tương đương với việc kiểm tra xem $\boldsymbol{b}$ có nằm trong tập sinh bởi các cột của $\boldsymbol{A}$ hay không. Tập hợp sinh này được gọi là **không gian cột** hoặc **miền giá trị** của $\boldsymbol{A}$.

Để hệ phương trình $\boldsymbol{Ax} = \boldsymbol{b}$ có nghiệm cho mọi giá trị của $\boldsymbol{b} \in \mathbb{R}^m$, ta cần không gian cột của $\boldsymbol{A}$ phải bao phủ toàn bộ $\mathbb{R}^m$. Nếu bất kỳ điểm nào trong $\mathbb{R}^m$ bị loại khỏi không gian cột, điểm đó là một giá trị $\boldsymbol{b}$ tiềm năng mà không có nghiệm. Việc yêu cầu rằng không gian cột của $\boldsymbol{A}$ phải bao phủ toàn bộ $\mathbb{R}^m$ ngay lập tức ngụ ý rằng $\boldsymbol{A}$ phải có ít nhất m cột, tức là $n \geq m$. Nếu không, số chiều của không gian cột sẽ nhỏ hơn m. Ví dụ, xét một ma trận cỡ 3×2. Mục tiêu $\boldsymbol{b}$ là 3 chiều, nhưng $\boldsymbol{x}$ chỉ có 2 chiều, vì vậy việc thay đổi giá trị của $\boldsymbol{x}$ chỉ cho phép ta truy ra một mặt

phẳng 2 chiều trong $\mathbb{R}^3$. Phương trình có nghiệm nếu và chỉ nếu $\boldsymbol{b}$ nằm trên mặt phẳng đó.

Điều kiện $n \geq m$ chỉ là điều kiện cần để mọi điểm đều có nghiệm. Nó không phải là điều kiện đủ, vì có thể một số cột là dư thừa. Ví dụ, xét một ma trận cỡ 2×2 có hai cột giống nhau. Ma trận này có không gian cột giống với không gian cột của ma trận 2×1 chỉ chứa một cột giống nhau ở trên. Nói cách khác, không gian cột vẫn chỉ là một đường thẳng và không bao phủ toàn bộ $\mathbb{R}^2$, mặc dù có hai cột.

Sự dư thừa kiểu này được gọi là **phụ thuộc tuyến tính**. Một tập hợp các véctơ được gọi là **độc lập tuyến tính** nếu không có véctơ nào trong tập là tổ hợp tuyến tính của các véctơ còn lại. Nếu ta thêm một véctơ vào một tập hợp mà véctơ đó là tổ hợp tuyến tính của các véctơ khác trong tập, thì véctơ mới này không thêm điểm nào vào bao tuyến tính của tập đó. Điều này có nghĩa là để không gian cột của ma trận bao phủ toàn bộ $\mathbb{R}^m$, ma trận phải chứa ít nhất một tập hợp gồm m cột độc lập tuyến tính. Điều kiện này vừa là điều kiện cần vừa là điều kiện đủ để phương trình (2.10) có nghiệm với mọi giá trị của $\boldsymbol{b}$. Lưu ý rằng yêu cầu là tập hợp phải có đúng m cột độc lập tuyến tính, chứ không phải ít nhất m. Không có tập hợp các véctơ m chiều nào có thể có hơn m cột độc lập tuyến tính, nhưng một ma trận có nhiều hơn m cột có thể chứa nhiều hơn một tập hợp như vậy.

Để ma trận có nghịch đảo, ta cần đảm bảo rằng phương trình (2.10) có *nhiều nhất* một nghiệm cho mỗi giá trị của $\boldsymbol{b}$. Để làm được điều đó, ta cần đảm bảo rằng ma trận có nhiều nhất m cột. Nếu không, sẽ có nhiều hơn một cách tham số hóa cho mỗi nghiệm.

Điều này có nghĩa là ma trận phải là **ma trận vuông**, tức là $m = n$ và tất cả các cột phải tuyến tính độc lập. Một ma trận vuông có các cột phụ thuộc tuyến tính được gọi là **ma trận suy biến**.

Nếu $\boldsymbol{A}$ không phải là ma trận vuông hoặc là ma trận vuông nhưng suy biến, vẫn có thể giải được phương trình. Tuy nhiên, ta không thể sử dụng phương pháp nghịch đảo ma trận để tìm nghiệm.

Đến lúc này, ta đã thảo luận về nghịch đảo ma trận khi nhân từ bên trái. Cũng có thể định nghĩa nghịch đảo được nhân từ bên phải:

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}. \tag{2.19}$$

Đối với các ma trận vuông, nghịch đảo trái và nghịch đảo phải là bằng nhau.

2.5 Chuẩn

Đôi khi ta cần đo chiều dài của một véctơ. Trong học máy, ta thường đo chiều dài của các véctơ bằng một hàm gọi là **chuẩn**. Chuẩn L^p được cho bởi

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \tag{2.20}$$

với $p \in \mathbb{R}$, $p \geq 1$.

Các chuẩn, bao gồm chuẩn L^p, là những hàm gán tương ứng mỗi véctơ với một số không âm. Về mặt trực quan, chuẩn của một véctơ $\boldsymbol{x}$ đo khoảng cách từ gốc tọa độ đến điểm $\boldsymbol{x}$. Chặt chẽ hơn, một chuẩn là bất kỳ hàm f nào thỏa mãn các tính chất sau:

- $f(\boldsymbol{x}) = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{0}$
- $f(\boldsymbol{x} + \boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$ (**bất đẳng thức tam giác**)
- $\forall \alpha \in \mathbb{R}, \quad f(\alpha \boldsymbol{x}) = |\alpha| f(\boldsymbol{x})$

Chuẩn L^2, với $p = 2$, được gọi là **chuẩn Euclid**. Nó đơn giản là khoảng cách Euclid từ gốc tọa độ đến điểm được xác định bởi $\boldsymbol{x}$. Chuẩn L^2 được sử dụng thường xuyên trong học máy đến mức thường được ký hiệu đơn giản là $\|\boldsymbol{x}\|$, với chỉ số 2 bị lược bỏ. Một đại lượng cũng phổ biến khi đánh giá chiều dài của một véctơ là chuẩn L^2 bình phương, có thể được tính đơn giản là $\boldsymbol{x}^T\boldsymbol{x}$.

Chuẩn L^2 bình phương thuận tiện hơn trong việc xử lý về mặt toán học và tính toán so với chính chuẩn L^2. Ví dụ, các đạo hàm của chuẩn L^2 bình phương theo từng phần tử của $\boldsymbol{x}$ chỉ phụ thuộc vào phần tử tương ứng của $\boldsymbol{x}$, trong khi tất cả các đạo hàm của chuẩn L^2 phụ thuộc vào toàn bộ véctơ. Trong nhiều tình huống, chuẩn L^2 bình phương có thể không được ưa chuộng vì nó tăng rất chậm gần gốc tọa độ. Trong một số ứng dụng học máy, việc phân biệt giữa các phần tử có giá trị chính xác bằng không và các phần tử nhỏ nhưng không bằng không là rất quan trọng. Trong những trường hợp này, ta sử dụng một hàm tăng với cùng tốc độ ở mọi vị trí, nhưng vẫn giữ được tính đơn giản về mặt toán học: chuẩn L^1. Chuẩn L^1 được xác định đơn giản như sau:

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|. \tag{2.21}$$

Chuẩn L^1 thường được sử dụng trong học máy khi sự khác biệt giữa các phần tử bằng 0 và khác 0 là rất quan trọng. Mỗi khi một phần tử của $\boldsymbol{x}$ dịch chuyển khỏi 0 một lượng ε, chuẩn L^1 sẽ tăng thêm ε.

Trong lĩnh vực học máy, đôi khi ta đo độ lớn của một véctơ bằng cách đếm số phần tử khác 0 của nó. Một số tác giả gọi hàm này là "chuẩn L^0", nhưng đây là thuật ngữ không chính xác. Số phần tử khác 0 trong một véctơ không phải là một chuẩn, vì việc nhân véctơ với một hệ số α không làm thay đổi số phần tử khác 0. Chuẩn L^1 thường được sử dụng như một sự thay thế cho số phần tử khác 0.

Một chuẩn khác thường xuất hiện trong học máy là chuẩn L^∞, còn được gọi là **chuẩn max**, xác định bởi giá trị tuyệt đối của phần tử có độ lớn lớn nhất trong véctơ:

$$\|\boldsymbol{x}\|_\infty = \max_i |x_i| . \tag{2.22}$$

Đôi khi ta cũng muốn đo độ lớn của một ma trận. Trong bối cảnh học sâu, cách phổ biến nhất để làm điều này là sử dụng **chuẩn Frobenius**, một chuẩn ít được biết đến:

$$\|\boldsymbol{A}\|_F = \sqrt{\sum_{i,j=1}^{n} A_{ij}^2}, \tag{2.23}$$

tương tự như chuẩn L_2 của một véctơ.

Tích vô hướng của hai véctơ có thể được viết lại theo các chuẩn. Cụ thể:

$$\boldsymbol{x}^T\boldsymbol{y} = \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 \cos\theta \tag{2.24}$$

trong đó θ là góc giữa $\boldsymbol{x}$ và $\boldsymbol{y}$.

2.6 Các loại ma trận và véctơ đặc biệt

Một số loại ma trận và véctơ đặc biệt rất hữu ích.

Ma trận đường chéo bao gồm hầu hết các phần tử bằng 0 và chỉ có các phần tử khác 0 trên đường chéo chính. Ma trận $\boldsymbol{D}$ là ma trận đường chéo khi và chỉ khi $D_{i,j} = 0$ cho mọi $i \neq j$. Ta đã thấy một ví dụ về ma trận chéo: ma trận đơn vị, trong đó tất cả các phần tử trên đường chéo chính đều bằng 1. Ta viết $\text{diag}(\boldsymbol{v})$ để biểu diễn ma trận vuông chéo mà các phần tử trên đường chéo được xác định bởi các phần tử của véctơ $\boldsymbol{v}$. Ma trận đường chéo đáng chú ý vì việc nhân với ma trận đường chéo rất hiệu quả về tính toán. Để tính $\text{diag}(\boldsymbol{v})\,\boldsymbol{x}$, ta chỉ cần nhân từng

phần tử x_i với v_i. Nói cách khác, $\text{diag}(\boldsymbol{v})\boldsymbol{x} = \boldsymbol{v} \odot \boldsymbol{x}$. Việc tìm nghịch đảo của ma trận vuông đường chéo cũng rất dễ dàng. Ma trận nghịch đảo chỉ tồn tại khi tất cả các phần tử trên đường chéo chính đều khác không, và trong trường hợp đó, $\text{diag}(\boldsymbol{v})^{-1} = \text{diag}\left(\left[\frac{1}{v_1}, \ldots, \frac{1}{v_n}\right]^T\right)$. Trong nhiều trường hợp, ta có thể xây dựng một số thuật toán học máy rất tổng quát cho các ma trận bất kỳ, từ một thuật toán tương tự nhưng đơn giản hơn bằng cách giới hạn một số ma trận là ma trận đường chéo.

Không phải tất cả các ma trận đường chéo đều phải là ma trận vuông. Có thể tạo ra một ma trận đường chéo hình chữ nhật. Ma trận đường chéo không vuông không có ma trận nghịch đảo, nhưng vẫn có thể nhân chúng với véctơ và ma trận. Đối với một ma trận chéo không vuông $\boldsymbol{D}$, tích $\boldsymbol{Dx}$ sẽ bao gồm việc co dãn từng phần tử của $\boldsymbol{x}$, và, hoặc là nối thêm một số số không vào kết quả nếu D nhiều hàng hơn cột, hoặc loại bỏ một số phần tử cuối cùng của véctơ nếu D it hàng hơn cột.

Ma trận đối xứng là ma trận bằng ma trận chuyển vị của chính nó:

$$\boldsymbol{A} = \boldsymbol{A}^T. \tag{2.25}$$

Ma trận đối xứng thường xuất hiện khi các phần tử được tạo ra bởi một hàm có hai tham số mà không phụ thuộc vào thứ tự của các tham số. Ví dụ, nếu $\boldsymbol{A}$ là một ma trận các khoảng cách, với A_{ij} biểu diễn khoảng cách từ điểm i đến điểm j, thì $A_{ij} = A_{ji}$ vì các hàm khoảng cách là đối xứng.

Véctơ đơn vị là một véctơ có **chuẩn bằng 1**:

$$\|\boldsymbol{x}\|_2 = 1. \tag{2.26}$$

Véctơ $\boldsymbol{x}$ và $\boldsymbol{y}$ gọi là **trực giao** (**vuông góc**) với nhau nếu $\boldsymbol{x}^T\boldsymbol{y} = 0$. Nếu cả hai véctơ đều có chuẩn khác không, điều này có nghĩa là chúng tạo với nhau một góc 90 độ. Trong không gian $\mathbb{R}^n$, tối đa có thể có n véctơ trực giao lẫn nhau với chuẩn khác 0. Nếu các véctơ không chỉ trực giao mà còn có chuẩn đơn vị, ta gọi chúng là các véctơ **trực chuẩn**.

Ma trận trực giao là ma trận vuông có các hàng trực chuẩn lẫn nhau và các cột cũng trực chuẩn lẫn nhau.

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{I}. \tag{2.27}$$

Điều này dẫn đến

$$\boldsymbol{A}^{-1} = \boldsymbol{A}^T, \tag{2.28}$$

vì vậy các ma trận trực giao thu hút sự quan tâm do nghịch đảo của chúng rất dễ tính toán. Cần chú ý kỹ đến định nghĩa của ma trận trực giao. Khác với tên gọi, các hàng của chúng không chỉ trực giao mà còn trực chuẩn. Không có thuật ngữ đặc biệt nào cho ma trận mà các hàng hoặc cột của nó chỉ trực giao nhưng không trực chuẩn.

2.7 Phân tích giá trị riêng

Nhiều đối tượng toán học có thể được hiểu rõ hơn bằng cách phân tách chúng thành các phần cấu thành, hoặc tìm ra một số tính chất điển hình của chúng, không phụ thuộc vào cách biểu diễn chúng.

Ví dụ, các số nguyên có thể được phân tích thành các thừa số nguyên tố. Cách ta biểu diễn số 12 sẽ thay đổi tùy vào việc ta viết nó ở hệ thập phân hay nhị phân, nhưng biểu diễn $12 = 2 \times 2 \times 3$ luôn đúng. Từ cách biểu diễn này, ta có thể suy ra các tính chất hữu ích, chẳng hạn như 12 không chia hết cho 5, hoặc bất kỳ bội số nào của 12 cũng sẽ chia hết cho 3.

Giống như việc có thể khám phá bản chất thực sự của một số nguyên bằng cách phân tích nó thành các thừa số nguyên tố, ta cũng có thể phân tích ma trận theo cách cho thấy các thông tin về thuộc tính cơ bản của chúng, là điều không hiển nhiên do cách biểu diễn ma trận như một mảng các phần tử.

Một trong những loại phân tích ma trận được sử dụng rộng rãi nhất được gọi là **phân tích giá trị riêng**, trong đó ta phân tích ma trận thành một tập hợp các véctơ riêng và giá trị riêng.

Một **véctơ riêng** của ma trận vuông $\boldsymbol{A}$ là một véctơ khác không $\boldsymbol{v}$ sao cho phép nhân với $\boldsymbol{A}$ chỉ thay đổi một tỷ lệ của $\boldsymbol{v}$:

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}. \tag{2.29}$$

Hệ số vô hướng λ được gọi là **giá trị riêng** tương ứng với véctơ riêng này. (Ta cũng có thể tìm **véctơ riêng trái** sao cho $\boldsymbol{v}^T\boldsymbol{A} = \lambda \boldsymbol{v}^T$, nhưng thường ta quan tâm đến véctơ riêng phải).

Nếu $\boldsymbol{v}$ là một véctơ riêng của ma trận $\boldsymbol{A}$, thì bất kỳ véctơ nào tỷ lệ với nó, tức là $s\boldsymbol{v}$ với $s \in \mathbb{R}$, $s \neq 0$ cũng là một véctơ riêng của $\boldsymbol{A}$. Hơn nữa, $s\boldsymbol{v}$ vẫn có cùng giá trị riêng. Vì lý do này, ta thường chỉ tìm các véctơ riêng đơn vị.

Giả sử ma trận $\boldsymbol{A}$ có n véctơ riêng độc lập tuyến tính, $\{\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(n)}\}$, với các giá trị riêng tương ứng $\{\lambda_1, \ldots, \lambda_n\}$. Ta có thể ghép tất cả các véctơ riêng lại để tạo

thành một ma trận $\boldsymbol{V}$, với mỗi véctơ riêng là một cột: $\boldsymbol{V} = \left[\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(n)}\right]$. Tương tự, ta có thể ghép các giá trị riêng lại để tạo thành một véctơ $\boldsymbol{\lambda} = \left[\lambda_1, \ldots, \lambda_n\right]^T$. Phân tích giá trị riêng của $\boldsymbol{A}$ sau đó được cho bởi

$$\boldsymbol{A} = \boldsymbol{V}\text{diag}\left(\boldsymbol{\lambda}\right)\boldsymbol{V}^{-1}. \tag{2.30}$$

Ta đã thấy rằng việc xây dựng các ma trận với các giá trị riêng và véctơ riêng cụ thể cho phép ta kéo giãn không gian theo các hướng mong muốn. Tuy nhiên, ta thường muốn phân tích ma trận thành các giá trị riêng và véctơ riêng của nó. Việc làm này có thể giúp ta phân tích một số thuộc tính của ma trận, giống như việc phân tích một số nguyên thành các thừa số nguyên tố có thể giúp ta hiểu được đặc điểm của số nguyên đó.

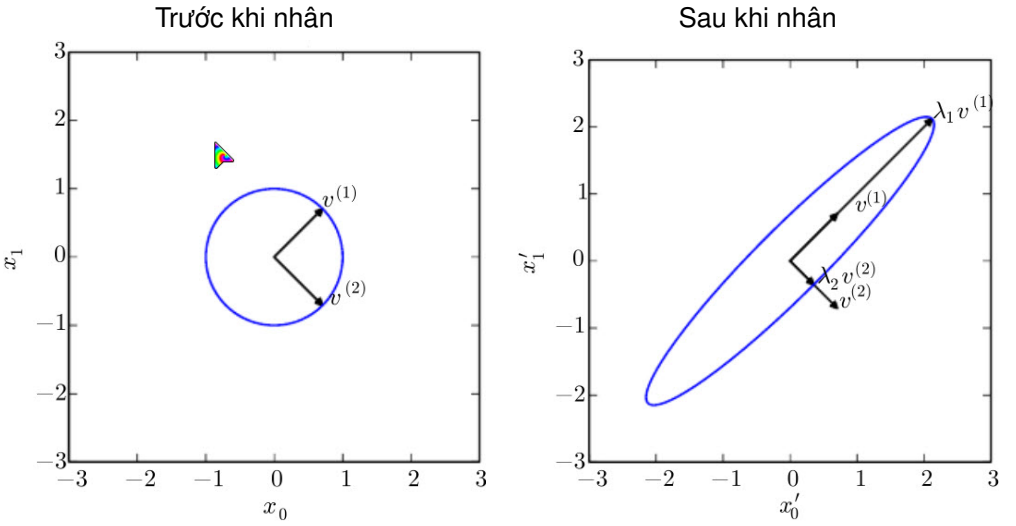
Không phải mọi ma trận đều có thể phân tích thành các giá trị riêng và véctơ riêng. Trong một số trường hợp, phép phân tích tồn tại, nhưng có thể liên quan đến số phức thay vì số thực. May mắn là, trong cuốn sách này, ta thường chỉ cần phân tích một lớp ma trận cụ thể có phân tích đơn giản. Cụ thể, mọi ma trận thực đối xứng đều có thể phân tích thành biểu thức chỉ sử dụng các véctơ riêng và giá trị riêng thực:

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T, \tag{2.31}$$

trong đó $\boldsymbol{Q}$ là ma trận trực giao bao gồm các véctơ riêng của $\boldsymbol{A}$, và $\boldsymbol{\Lambda}$ là ma trận đường chéo. Giá trị riêng Λ_{ii} tương ứng với véctơ riêng ở cột i của $\boldsymbol{Q}$, ký hiệu là $\boldsymbol{Q}_{:,i}$. Vì $\boldsymbol{Q}$ là ma trận trực giao, ta có thể xem $\boldsymbol{A}$ như việc kéo giãn không gian với mỗi hướng $\boldsymbol{v}^{(i)}$ theo tỷ lệ λ_i. Hình 2.3 là một ví dụ về mô tả trên.

Mặc dù bất kỳ ma trận thực đối xứng thực $\boldsymbol{A}$ nào cũng đều có phân tích giá trị riêng, nhưng phân tích giá trị riêng có thể không duy nhất. Nếu hai hoặc nhiều véctơ riêng có cùng giá trị riêng, thì bất kỳ tập hợp các véctơ trực giao nào nằm trong không gian con sinh bởi chúng cũng là các véctơ riêng với giá trị riêng đó, và ta có thể chọn một ma trận $\boldsymbol{Q}$ sử dụng các véctơ riêng đó. Theo quy ước, ta thường sắp xếp các phần tử của $\boldsymbol{\Lambda}$ theo thứ tự giảm dần. Theo quy ước này, phân tích giá trị riêng là duy nhất khi tất cả các giá trị riêng đều là duy nhất.

Phân tích giá trị riêng của một ma trận cho ta biết nhiều thông tin hữu ích về ma trận đó. Ma trận là suy biến nếu và chỉ nếu ma trận đó có giá trị riêng nào đó bằng 0. Phân tích giá trị riêng của một ma trận thực đối xứng cũng có thể được sử dụng để tối ưu hóa các dạng toàn phương, tức là các biểu thức bậc hai có dạng $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ với điều kiện $\left\|\boldsymbol{x}\right\|_2 = 1$. Mỗi khi $\boldsymbol{x}$ bằng một véctơ riêng của $\boldsymbol{A}$, f sẽ có giá trị bằng giá trị riêng tương ứng. Giá trị lớn nhất của f trong miền ràng buộc



Hình 2.3: Ví dụ về tác động của véctơ riêng và giá trị riêng. Ở đây, ta có ma trận $\boldsymbol{A}$ với hai véctơ riêng trực chuẩn, $\boldsymbol{v}^{(1)}$ với giá trị riêng λ_1 và $\boldsymbol{v}^{(2)}$ với giá trị riêng λ_2. (*Bên trái*) Ta vẽ tập hợp tất cả các véctơ đơn vị $\boldsymbol{u} \in \mathbb{R}^2$ dưới dạng một đường tròn đơn vị. (*Bên phải*) Ta vẽ tập hợp tất cả các điểm $\boldsymbol{Au}$. Bằng cách quan sát cách $\boldsymbol{A}$ làm biến dạng đường tròn đơn vị, ta có thể thấy rằng nó kéo giãn không gian theo hướng $\boldsymbol{v}^{(i)}$ với tỷ lệ λ_i.

là giá trị riêng lớn nhất và giá trị nhỏ nhất trong miền ràng buộc là giá trị riêng nhỏ nhất.

Ma trận có tất cả các giá trị riêng dương được gọi là ma trận xác định dương. Na trận có tất cả các giá trị riêng dương hoặc bằng 0 được gọi là ma trận nửa xác định dương. Tương tự, nếu tất cả các giá trị riêng âm, ma trận được gọi là ma trận xác định âm, và nếu tất cả các giá trị riêng âm hoặc bằng 0, ma trận là nửa xác định âm. Các ma trận nửa xác định dương rất thú vị vì chúng đảm bảo rằng $\forall \boldsymbol{x},\ \boldsymbol{x}^T\boldsymbol{Ax} \geq 0$. Các ma trận xác định dương còn đảm bảo thêm rằng $\boldsymbol{x}^T\boldsymbol{Ax} = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{0}$.

2.8 Phân tích giá trị kỳ dị

Trong Mục 2.7, ta đã thấy cách phân tích một ma trận thành các véctơ riêng và giá trị riêng. **Phân tích giá trị kỳ dị** (SVD) cung cấp một cách khác để phân tích ma trận thành các véctơ suy biến và giá trị suy biến. Phân tích giá trị kỳ dị cho phép ta khám phá một số thông tin tương tự như phép phân tích giá trị riêng. Tuy nhiên, phân tích giá trị kỳ dị có thể áp dụng tổng quát hơn. Mọi ma trận thực đều có

phân tích giá trị kỳ dị, nhưng điều này không đúng đối với phân tích giá trị riêng. Ví dụ, nếu một ma trận không phải là ma trận vuông, thì phép phân tích giá trị riêng không được định nghĩa và thay vào đó ta phải sử dụng phép phân tích giá trị kỳ dị.

Nhớ lại rằng phép phân tích giá trị riêng liên quan đến việc phân tích một ma trận $\boldsymbol{A}$ để tìm ma trận $\boldsymbol{V}$ gồm các véctơ riêng và một véctơ $\boldsymbol{\lambda}$ gồm các giá trị riêng sao cho có thể viết lại $\boldsymbol{A}$ dưới dạng

$$\boldsymbol{A} = \boldsymbol{V}\text{diag}\left(\boldsymbol{\lambda}\right)\boldsymbol{V}^{-1}. \qquad \text{(xem (2.30))}$$

Phân tích giá trị kỳ dị cũng tương tự, ngoại trừ lần này ta sẽ viết $\boldsymbol{A}$ dưới dạng tích của ba ma trận:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{T} \tag{2.32}$$

Giả sử $\boldsymbol{A}$ là ma trận cỡ $m \times m$. Khi đó $\boldsymbol{U}$ được định nghĩa là ma trận cỡ $m \times n$, $\boldsymbol{D}$ là ma trận cỡ $m \times n$, và $\boldsymbol{V}$ là ma trận cỡ $n \times n$.

Mỗi ma trận trong số này được định nghĩa có một cấu trúc đặc biệt. Các ma trận $\boldsymbol{U}$ và $\boldsymbol{V}$ đều là ma trận trực giao. Ma trận $\boldsymbol{D}$ là ma trận đường chéo. Lưu ý rằng $\boldsymbol{D}$ không nhất thiết phải là ma trận vuông. Các phần tử dọc theo đường chéo của $\boldsymbol{D}$ được gọi là các **giá trị kỳ dị** của ma trận $\boldsymbol{A}$. Các cột của $\boldsymbol{U}$ được gọi là các **véctơ kỳ dị trái**. Các cột của $\boldsymbol{V}$ được gọi là các **véctơ kỳ dị phải**.

Ta thực sự có thể diễn giải phân tích giá trị kỳ dị của ma trận $\boldsymbol{A}$ dưới dạng phân tích giá trị riêng của các ma trận liên quan đến $\boldsymbol{A}$. Các véctơ kỳ dị trái của $\boldsymbol{A}$ là các véctơ riêng của $\boldsymbol{A}\boldsymbol{A}^{T}$. Các véctơ kỳ dị phải của $\boldsymbol{A}$ là các véctơ riêng của $\boldsymbol{A}^{T}\boldsymbol{A}$. Các giá trị kỳ dị khác 0 của $\boldsymbol{A}$ là căn bậc hai của các giá trị riêng của $\boldsymbol{A}^{T}\boldsymbol{A}$. Điều này cũng đúng với $\boldsymbol{A}\boldsymbol{A}^{T}$.

Có lẽ tính năng hữu ích nhất của phân tích giá trị kỳ dị là ta có thể sử dụng nó để tổng quát hóa một phần phép nghịch đảo ma trận cho các ma trận không vuông, như sẽ thấy trong phần tiếp theo.

2.9 Ma trận giả nghịch đảo Moore – Penrose

Phép nghịch đảo ma trận không được định nghĩa cho các ma trận không vuông. Giả sử ta muốn xây dựng một ma trận nghịch đảo trái B của ma trận A, để có thể giải phương trình tuyến tính

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \tag{2.33}$$

bằng cách nhân bên trái mỗi vế để thu được

$$\boldsymbol{x} = \boldsymbol{B}\boldsymbol{y}. \tag{2.34}$$

Tùy thuộc vào cấu trúc của bài toán, có thể không thể thiết kế được một tương ứng duy nhất biến $\boldsymbol{A}$ thành $\boldsymbol{B}$.

Nếu ma trận $\boldsymbol{A}$ có nhiều hàng hơn cột, thì phương trình này có thể không có nghiệm. Nếu $\boldsymbol{A}$ ít hàng hơn cột, thì có thể có nhiều nghiệm khác nhau. **Ma trận giả nghịch đảo Moore – Penrose** cho phép ta xử lý các trường hợp này. Giả nghịch đảo của $\boldsymbol{A}$ được định nghĩa là ma trận

$$\boldsymbol{A}^{+} = \lim_{\alpha \downarrow 0} \left(\boldsymbol{A}^{T} \boldsymbol{A} + \alpha \boldsymbol{I} \right)^{-1} \boldsymbol{A}^{T}. \tag{2.35}$$

Các thuật toán thực tế để tính giả nghịch đảo không dựa trên định nghĩa này, mà dựa trên công thức

$$\boldsymbol{A}^{+} = \boldsymbol{V} \boldsymbol{D}^{+} \boldsymbol{U}^{T}, \tag{2.36}$$

trong đó $\boldsymbol{U}$, $\boldsymbol{D}$ và $\boldsymbol{V}$ là phân tích giá trị kỳ dị của $\boldsymbol{A}$, và giả nghịch đảo $\boldsymbol{D}^{+}$ của ma trận đường chéo $\boldsymbol{D}$ được tính bằng cách lấy nghịch đảo của các phần tử khác không của nó, sau đó lấy chuyển vị của ma trận kết quả.

Khi ma trận $\boldsymbol{A}$ có nhiều cột hơn hàng, việc giải một phương trình tuyến tính bằng ma trận giả nghịch đảo sẽ cung cấp một trong nhiều nghiệm có thể. Cụ thể, nó trả về nghiệm $\boldsymbol{x} = \boldsymbol{A}^{+} \boldsymbol{y}$ với chuẩn Euclid tối thiểu $\|\boldsymbol{x}\|_2$ trong số tất cả các nghiệm có thể.

Khi ma trận $\boldsymbol{A}$ có nhiều hàng hơn cột, có thể phương trình không có nghiệm. Trong trường hợp này, việc sử dụng giả nghịch đảo sẽ cho ta giá trị $\boldsymbol{x}$ sao cho $\boldsymbol{A}\boldsymbol{x}$ gần với $\boldsymbol{y}$ nhất theo chuẩn Euclid $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2$.

2.10 Vết của ma trận

Vết của ma trận là tổng tất cả các phần tử trên đường chéo chính của ma trận:

$$\mathrm{Tr}\left(\boldsymbol{A}\right) = \sum_{i} A_{ii}. \tag{2.37}$$

Toán tử vết hữu ích vì nhiều lý do. Một số phép toán khó biểu diễn mà không sử dụng ký hiệu tổng có thể được biểu diễn bằng cách sử dụng tích ma trận và toán tử vết. Ví dụ, toán tử vết cung cấp một cách viết khác cho chuẩn Frobenius của ma trận:

$$\|\boldsymbol{A}\|_F = \sqrt{\mathrm{Tr}\left(\boldsymbol{A}\boldsymbol{A}^{T}\right)}. \tag{2.38}$$

Việc viết một biểu thức dưới dạng toán tử vết mở ra cơ hội để biến đổi biểu thức bằng cách sử dụng nhiều đồng nhất thức phổ biến. Ví dụ, toán tử vết bất biến

đối với phép chuyển vị ma trận:

$$\text{Tr}\left(\boldsymbol{A}\right) = \text{Tr}\left(\boldsymbol{A}^{T}\right). \tag{2.39}$$

Vết của ma trận vuông được tạo thành từ tích nhiều ma trận cũng bất biến khi di chuyển ma trận nhân tử cuối cùng lên vị trí đầu tiên, nếu cỡ của các ma trận là tương thích, tức là phép nhân ma trận khi viết ra có thể thực hiện được.

$$\text{Tr}\left(\boldsymbol{ABC}\right) = \text{Tr}\left(\boldsymbol{C\,AB}\right) = \text{Tr}\left(\boldsymbol{B\,CA}\right) \tag{2.40}$$

hay tổng quát hơn:

$$\text{Tr}\left(\prod_{i=1}^{n} \boldsymbol{F}^{(i)}\right) = \text{Tr}\left(\boldsymbol{F}^{(n)} \prod_{i=1}^{n-1} \boldsymbol{F}^{(i)}\right). \tag{2.41}$$

Tính bất biến đối với hoán vị vòng này vẫn đúng ngay cả khi các tích ma trận có cỡ khác nhau. Ví dụ, với $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ và $\boldsymbol{B} \in \mathbb{R}^{n\times m}$, ta có

$$\text{Tr}\left(\boldsymbol{AB}\right) = \text{Tr}\left(\boldsymbol{BA}\right) \tag{2.42}$$

mặc dù $\boldsymbol{AB} \in \mathbb{R}^{m\times m}$ và $\boldsymbol{BA} \in \mathbb{R}^{n\times n}$.

Một kết quả khác cần nhớ là một số vô hướng chính là vết của nó: $a = \text{Tr}\left(a\right)$.

2.11 Định thức

Định thức của một ma trận vuông, ký hiệu là $\det\left(\boldsymbol{A}\right)$, là một hàm ánh xạ mỗi ma trận thành một số thực. Định thức bằng tích của tất cả các giá trị riêng của ma trận. Giá trị tuyệt đối của định thức có thể được coi như một thước đo cho việc phép nhân với ma trận đó mở rộng hay thu hẹp không gian bao nhiêu. Nếu định thức bằng 0, thì không gian bị thu hẹp hoàn toàn dọc theo ít nhất một chiều, khiến nó mất toàn bộ thể tích. Nếu định thức bằng 1, thì phép biến đổi bảo toàn thể tích.

2.12 Ví dụ: phân tích thành phần chính

Một thuật toán học máy đơn giản, **phân tích thành phần chính** (PCA), có thể được suy ra chỉ bằng kiến thức cơ bản về đại số tuyến tính.

Cho tập gồm m điểm $\left\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\right\}$ trong không gian $\mathbb{R}^{n}$. Giả sử ta muốn áp dụng nén mất dữ liệu cho các điểm này. Nén mất dữ liệu có nghĩa là lưu trữ các điểm theo cách yêu cầu ít bộ nhớ hơn nhưng có thể mất một phần độ chính xác. Ta muốn mất càng ít độ chính xác càng tốt.

Một cách để mã hóa các điểm này là biểu diễn chúng dưới dạng phiên bản có số chiều thấp hơn. Đối với mỗi điểm $\boldsymbol{x}^{(i)} \in \mathbb{R}^n$, ta sẽ tìm một véctơ mã hóa tương ứng $\boldsymbol{c}^{(i)} \in \mathbb{R}^\ell$. Nếu $\ell < n$, sẽ cần ít bộ nhớ hơn để lưu trữ các điểm mã hóa so với dữ liệu gốc. Ta sẽ cần tìm một hàm mã hóa tạo ra mã cho đầu vào, $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{c}$, và một hàm giải mã để tái tạo đầu vào từ mã hóa của nó, $\boldsymbol{x} \approx \boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}))$.

Phân tích thành phần chính được xác định bởi cách ta chọn hàm giải mã. Cụ thể, một giải mã rất đơn giản, sử dụng phép nhân ma trận để ánh xạ mã trở lại không gian $\mathbb{R}^n$. Giả sử $\boldsymbol{g}(\boldsymbol{c}) = \boldsymbol{D}\boldsymbol{c}$, trong đó $\boldsymbol{D} \in \mathbb{R}^{n \times \ell}$ là ma trận xác định việc giải mã.

Việc tính toán mã tối ưu cho bộ giải mã này có thể là một vấn đề phức tạp. Để giữ cho việc mã hóa dễ dàng, phân tích thành phần chính yêu cầu các cột của $\boldsymbol{D}$ phải trực giao với nhau. (Lưu ý rằng $\boldsymbol{D}$ vẫn chưa phải là "ma trận trực giao" trừ khi $\ell = n$).

Với bài toán như mô tả ở trên, có nhiều nghiệm khả thi, vì ta có thể tăng tỷ lệ của $\boldsymbol{D}_{:,i}$ nếu giảm tỷ lệ tương ứng của c_i tại tất cả các điểm. Để đảm bảo bài toán có nghiệm duy nhất, ta ràng buộc tất cả các cột của $\boldsymbol{D}$ phải có chuẩn đơn vị.

Để biến ý tưởng cơ bản này thành một thuật toán có thể triển khai, điều đầu tiên ta cần làm là tìm cách tạo ra điểm mã tối ưu $\boldsymbol{c}^*$ cho mỗi điểm đầu vào $\boldsymbol{x}$. Một cách để làm điều này là cực tiểu hóa khoảng cách giữa điểm đầu vào $\boldsymbol{x}$ và tái tạo của nó, $\boldsymbol{g}(\boldsymbol{c}^*)$. Ta có thể đo khoảng cách này bằng một chuẩn. Trong thuật toán phân tích thành phần chính, ta sử dụng chuẩn L^2:

$$\boldsymbol{c}^* = \arg\min_{\boldsymbol{c}} \left\| \boldsymbol{x} - \boldsymbol{g}(\boldsymbol{c}) \right\|_2. \tag{2.43}$$

Ta có thể chuyển sang sử dụng chuẩn L^2 bình phương thay vì chuẩn L^2, vì cả hai đều được cực tiểu hóa bởi cùng một giá trị của $\boldsymbol{c}$. Cả hai đều đạt giá trị nhỏ nhất với cùng một giá trị $\boldsymbol{c}$ vì chuẩn L^2 không âm và phép bình phương là phép toán đơn điệu tăng đối với các đối số không âm.

$$\boldsymbol{c}^* = \arg\min_{\boldsymbol{c}} \left\| \boldsymbol{x} - \boldsymbol{g}(\boldsymbol{c}) \right\|_2^2. \tag{2.44}$$

Hàm đang được cực tiểu hóa hóa được đơn giản hóa thành

$$\left[\boldsymbol{x} - \boldsymbol{g}(\boldsymbol{c})\right]^T \left[\boldsymbol{x} - \boldsymbol{g}(\boldsymbol{c})\right] \tag{2.45}$$

(theo định nghĩa của chuẩn L^2 ở công thức (2.20))

$$= \boldsymbol{x}^T\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{g}(\boldsymbol{c}) - \boldsymbol{g}(\boldsymbol{c})^T\boldsymbol{x} + \boldsymbol{g}(\boldsymbol{c})^T\boldsymbol{g}(\boldsymbol{c}) \tag{2.46}$$

(theo tính chất phân phối)

$$= \boldsymbol{x}^T\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{g}(\boldsymbol{c}) + \boldsymbol{g}(\boldsymbol{c})^T\boldsymbol{g}(\boldsymbol{c}) \tag{2.47}$$

(vì đại lượng vô hướng $\boldsymbol{x}^T\boldsymbol{g}(\boldsymbol{c})$ bằng chuyển vị của chính nó).

Bây giờ ta có thể một lần nữa thay đổi hàm đang được cực tiểu hóa, bằng cách bỏ đi hạng tử đầu tiên, vì hạng tử này không phụ thuộc vào $\boldsymbol{c}$:

$$\boldsymbol{c}^* = \arg\min_{\boldsymbol{c}} -2\boldsymbol{x}^T\boldsymbol{g}(\boldsymbol{c}) + \boldsymbol{g}(\boldsymbol{c})^T\boldsymbol{g}(\boldsymbol{c}). \tag{2.48}$$

Để tiến xa hơn, ta cần thay định nghĩa của $\boldsymbol{g}(\boldsymbol{c})$ vào:

$$\boldsymbol{c}^* = \arg\min_{\boldsymbol{c}} -2\boldsymbol{x}^T\boldsymbol{D}\boldsymbol{c} + \boldsymbol{c}^T\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{c} \tag{2.49}$$

$$= \arg\min_{\boldsymbol{c}} -2\boldsymbol{x}^T\boldsymbol{D}\boldsymbol{c} + \boldsymbol{c}^T I_\ell \boldsymbol{c} \tag{2.50}$$

(do tính trực giao và chuẩn đơn vị ràng buộc trên $\boldsymbol{D}$)

$$= \arg\min_{\boldsymbol{c}} -2\boldsymbol{x}^T\boldsymbol{D}\boldsymbol{c} + \boldsymbol{c}^T\boldsymbol{c} \tag{2.51}$$

Ta có thể giải bài toán tối ưu này bằng cách sử dụng phép tính vector (xem Mục 4.1 nếu bạn chưa biết cách thực hiện điều này).

$$\nabla_{\boldsymbol{c}}\left(-2\boldsymbol{x}^T\boldsymbol{D}\boldsymbol{c} + \boldsymbol{c}^T\boldsymbol{c}\right) = \boldsymbol{0} \tag{2.52}$$

$$-2\boldsymbol{D}^T\boldsymbol{x} + 2\boldsymbol{c} = \boldsymbol{0} \tag{2.53}$$

$$\boldsymbol{c} = \boldsymbol{D}^T\boldsymbol{x}. \tag{2.54}$$

Điều này làm cho thuật toán trở nên hiệu quả: ta có thể mã hóa x một cách tối ưu chỉ bằng một phép toán ma trận – véctơ. Để mã hóa một véc tơ, ta dùng hàm mã hóa

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{D}^T\boldsymbol{x}. \tag{2.55}$$

Sử dụng thêm phép nhân ma trận nữa, ta cũng có thể định nghĩa phép tái tạo của phương pháp phân tích thành phần chính:

$$\boldsymbol{r}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x})) = \boldsymbol{D}\boldsymbol{D}^T\boldsymbol{x}. \tag{2.56}$$

Tiếp theo, ta cần chọn ma trận mã hóa $\boldsymbol{D}$. Để làm điều này, ta quay lại ý tưởng về việc cực tiểu khoảng cách L^2 giữa các đầu vào và các điểm tái tạo. Vì ta sẽ sử dụng cùng một ma trận $\boldsymbol{D}$ để giải mã tất cả các điểm, nên không thể xem xét các

điểm một cách riêng lẻ nữa. Thay vào đó, ta phải giảm thiểu chuẩn Frobenius của ma trận sai số được tính trên tất cả các chiều và tất cả các điểm.

$$\boldsymbol{D}^* = \arg\min_{\boldsymbol{D}} \sqrt{\sum_{i=1}^{m} \left\| \boldsymbol{x}^{(i)} - \boldsymbol{r}\left(\boldsymbol{x}^{(i)}\right) \right\|_2^2} = \arg\min_{D} \sqrt{\sum_{i,j} \left(x_j^{(i)} - \boldsymbol{r}\left(\boldsymbol{x}^{(i)}\right)_j \right)^2} \quad (2.57)$$

sao cho $\boldsymbol{D}^T\boldsymbol{D} = \boldsymbol{I}_\ell$.

Để xây dựng thuật toán tìm $\boldsymbol{D}^*$, ta sẽ bắt đầu bằng cách xét trường hợp $\ell = 1$. Trong trường hợp này, $\boldsymbol{D}$ chỉ là một véctơ, ký hiệu $\boldsymbol{d}$. Thay phương trình (2.56) vào phương trình (2.57) và xét đơn giản hóa $\boldsymbol{D}$ thành $\boldsymbol{d}$, bài toán được rút gọn thành

$$\boldsymbol{d}^* = \arg\min_{\boldsymbol{d}} \sum_{i=1}^{m} \left\| \boldsymbol{x}^{(i)} - \boldsymbol{d}\boldsymbol{d}^T\boldsymbol{x}^{(i)} \right\|_2^2 \quad \text{sao cho} \quad \|\boldsymbol{d}\|_2 = 1. \quad (2.58)$$

Vì chuẩn L^2 của véctơ chính là chuẩn Frobenius, và chuẩn Frobenius của ma trận không đổi qua phép chuyển vị, nên

$$\boldsymbol{d}^* = \arg\min_{\boldsymbol{d}} \sum_{i=1}^{m} \left\| \left(\boldsymbol{x}^{(i)}\right)^T - \left(\boldsymbol{x}^{(i)}\right)^T \boldsymbol{d}\boldsymbol{d}^T \right\|_F^2 \quad \text{sao cho} \quad \|\boldsymbol{d}\|_2 = 1. \quad (2.59)$$

Bây giờ, ta viết lại bài toán dưới dạng hữu ích hơn, thông qua ma trận thiết kế của các mẫu, thay vì lấy tổng qua các véctơ mẫu riêng lẻ. Điều này sẽ cho phép ta sử dụng ký hiệu ngắn gọn hơn. Gọi $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ là ma trận được định nghĩa bằng cách xếp chồng tất cả các véctơ mô tả các điểm dữ liệu, sao cho $\boldsymbol{X}_{i,:} = \left(\boldsymbol{x}^{(i)}\right)^T$. Ta có thể viết lại bài toán dưới dạng

$$\boldsymbol{d}^* = \arg\min_{\boldsymbol{d}} \left\| \boldsymbol{X} - \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T \right\|_F^2 \quad \text{sao cho} \quad \boldsymbol{d}^T\boldsymbol{d} = 1. \quad (2.60)$$

Tạm thời bỏ qua ràng buộc, ta có thể đơn giản hóa phần chuẩn Frobenius như sau:

$$\arg\min_{\boldsymbol{d}} \left\| \boldsymbol{X} - \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T \right\|_F^2 \quad (2.61)$$

$$= \arg\min_{\boldsymbol{d}} \operatorname{Tr}\left(\left(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right)^T \left(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right) \right) \quad (2.62)$$

(do phương trình (2.38))

$$= \arg\min_{\boldsymbol{d}} \operatorname{Tr}\left(\boldsymbol{X}^T\boldsymbol{X} - \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T - \boldsymbol{d}\boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{d}\boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T \right) \quad (2.63)$$

$$= \arg\min_{\boldsymbol{d}} \operatorname{Tr}\left(\boldsymbol{X}^T\boldsymbol{X}\right) - \operatorname{Tr}\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right) - \operatorname{Tr}\left(\boldsymbol{d}\boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{X}\right) + \operatorname{Tr}\left(\boldsymbol{d}\boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right) \quad (2.64)$$

$$= \arg\min_{\boldsymbol{d}} \; -\text{Tr}\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right) - \text{Tr}\left(\boldsymbol{d}\boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{X}\right) + \text{Tr}\left(\boldsymbol{d}\boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right) \tag{2.65}$$

(vì số hạng không phụ thuộc vào d thì không ảnh hưởng tới $\arg\min$) (2.66)

$$= \arg\min_{\boldsymbol{d}} \; -2\text{Tr}\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right) + \text{Tr}\left(\boldsymbol{d}\boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right) \tag{2.67}$$

(vì theo phương trình (2.41), có thể hoán vị vòng các ma trận trong toán tử vết)

$$= \arg\min_{\boldsymbol{d}} \; -2\text{Tr}\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right) + \text{Tr}\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\boldsymbol{d}\boldsymbol{d}^T\right) \tag{2.68}$$

(dùng tiếp tính chất trên)

Bây giờ ta nhắc lại bài toán có ràng buộc:

$$\arg\min_{\boldsymbol{d}} \; -2\text{Tr}\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right) + \text{Tr}\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\boldsymbol{d}\boldsymbol{d}^T\right) \quad \text{sao cho} \quad \boldsymbol{d}^T\boldsymbol{d} = 1 \tag{2.69}$$

$$= \arg\min_{\boldsymbol{d}} \; -2\text{Tr}\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right) + \text{Tr}\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right) \quad \text{sao cho} \quad \boldsymbol{d}^T\boldsymbol{d} = 1 \tag{2.70}$$

(theo ràng buộc trên)

$$= \arg\min_{\boldsymbol{d}} \; -\text{Tr}\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right) \quad \text{sao cho} \quad \boldsymbol{d}^T\boldsymbol{d} = 1 \tag{2.71}$$

$$= \arg\max_{\boldsymbol{d}} \; \text{Tr}\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\right) \quad \text{sao cho} \quad \boldsymbol{d}^T\boldsymbol{d} = 1 \tag{2.72}$$

$$= \arg\max_{\boldsymbol{d}} \; \text{Tr}\left(\boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\right) \quad \text{sao cho} \quad \boldsymbol{d}^T\boldsymbol{d} = 1 \tag{2.73}$$

Bài toán tối ưu này có thể giải bằng phương pháp phân tích giá trị riêng. Cụ thể, véctơ tối ưu $\boldsymbol{d}$ là véctơ riêng của ma trận $\boldsymbol{X}^T\boldsymbol{X}$ ứng với giá trị riêng lớn nhất.

Lập luận này chỉ áp dụng cho trường hợp $\ell = 1$ và chỉ tìm được thành phần chính thứ nhất. Tổng quát hơn, khi ta muốn tìm một cơ sở của các thành phần chính, ma trận D được cho bởi ℓ vector riêng ứng với các giá trị riêng lớn nhất. Điều này có thể được chứng minh bằng phương pháp quy nạp. Bạn đọc có thể tự chứng minh xem như một bài tập.

2.13 Không gian Euclid

Cho không gian véctơ V trên trường số thực. Xét hàm f gán tương ứng mỗi cặp phần tử $(\boldsymbol{x}, \boldsymbol{y})$, trong đó $\boldsymbol{x}, \boldsymbol{y} \in V$, với một số thực $f(\boldsymbol{x}, \boldsymbol{y})$. Hàm f được gọi là **song tuyến tính**, nếu khi cố định thành phần này, thì nó là ánh xạ tuyến tính theo thành phần kia, tức là với mọi $\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y}, \boldsymbol{y}' \in V$ và $\alpha \in \mathbb{R}$, điều kiện sau phải thỏa mãn:

$$\begin{aligned} f\left(\boldsymbol{x} + \boldsymbol{x}', y\right) &= f(\boldsymbol{x}, \boldsymbol{y}) + f\left(\boldsymbol{x}', \boldsymbol{y}\right) \quad \text{và} \quad f(\alpha\boldsymbol{x}, \boldsymbol{y}) = \alpha f(\boldsymbol{x}, \boldsymbol{y}) \\ f\left(\boldsymbol{x}, \boldsymbol{y} + \boldsymbol{y}'\right) &= f(\boldsymbol{x}, \boldsymbol{y}) + f\left(\boldsymbol{x}, \boldsymbol{y}'\right) \quad \text{và} \quad f(\boldsymbol{x}, \alpha\boldsymbol{y}) = \alpha f(\boldsymbol{x}, \boldsymbol{y}) \end{aligned} \tag{2.74}$$

Dạng song tuyến tính f trên V được gọi là

- **đối xứng** nếu

$$f(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{y}, \boldsymbol{x}) \quad \text{với mọi } \boldsymbol{x}, \boldsymbol{y} \in V. \tag{2.75}$$

- **xác định dương** nếu

$$f(\boldsymbol{x}, \boldsymbol{x}) \geq 0 \quad \text{với mọi } \boldsymbol{x} \in V, \quad \text{và} \quad f(\boldsymbol{x}, \boldsymbol{x}) = 0 \Rightarrow \boldsymbol{x} = 0. \tag{2.76}$$

Nếu f là một dạng song tuyến tính đối xứng xác định dương trên V thì f được gọi là một **tích vô hướng** trên V, và V được gọi là **không gian Euclid**. Khi đó ta ký hiệu $f(\boldsymbol{x}, \boldsymbol{y})$ bởi $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ và đọc là "tích vô hướng $\boldsymbol{x}$ với $\boldsymbol{y}$". Ta cũng viết lại các điều kiện của tích vô hướng theo ký hiệu mới, lưu ý rằng do tính đối xứng, nên chỉ cần đảm bảo tính tuyến tính theo một biến là đủ:

$$\begin{aligned}
&\langle \boldsymbol{x} + \boldsymbol{x}', \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}', \boldsymbol{y} \rangle \quad \text{và} \quad \langle \alpha \boldsymbol{x}, \boldsymbol{y} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{y} \rangle \\
&\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle \\
&\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0 \\
&\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0 \Rightarrow \boldsymbol{x} = 0
\end{aligned}$$

Trong không gian Euclid, tích vô hướng $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ sẽ sinh ra một chuẩn:

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}. \tag{2.77}$$

Một ví dụ đơn giản, trong không gian $\mathbb{R}^n$, với

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T, \quad \boldsymbol{y} = (y_1, y_2, \ldots, y_n)^T \in \mathbb{R}^n$$

có tích vô hướng

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} x_i y_i \tag{2.78}$$

và chuẩn sinh bởi tích vô hướng

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \sqrt{\sum_{i=1}^{n} |x_i|^2} \tag{2.79}$$

chính là chuẩn L^2 trong $\mathbb{R}^n$.

... đang soạn!

Đại số tuyến tính là một trong những ngành toán học cơ bản cần thiết để hiểu về học máy. Một lĩnh vực toán học quan trọng khác thường xuyên xuất hiện trong học máy là lý thuyết xác suất, sẽ được trình bày tiếp theo.

Bài tập Chương 2

1. Cho hai hàm véctơ m thành phần với n biến $\boldsymbol{f}, \boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$ xác định bởi: với $\boldsymbol{x} = [x_1, x_2, \dots, x_n]^T$, $\boldsymbol{f}(\boldsymbol{x}) = [f_1(x), f_2(x), \dots, f_m(x)]^T$ trong đó $f_i(\boldsymbol{x}) = f_i(x_1, x_2, \dots, x_n)$ là các hàm số n biến. Ta quy ước ký hiệu $\boldsymbol{f}(\boldsymbol{x})$ bởi $\boldsymbol{f}$, và tương tự đối với các hàm khác. Ký hiệu $\boldsymbol{J}(\boldsymbol{f}) = \left(\dfrac{\partial f_i}{\partial x_j}\right)$ là **ma trận Jacobi** của $\boldsymbol{f}$ tại $\boldsymbol{x}$, $i = \overline{1, m}$ là chỉ số hàng, $j = \overline{1, n}$ là chỉ số cột. Chứng minh

a) $\boldsymbol{J}(\boldsymbol{f} + \boldsymbol{g}) = \boldsymbol{J}(\boldsymbol{f}) + \boldsymbol{J}(\boldsymbol{g})$

b) $\boldsymbol{J}(\alpha \boldsymbol{f}) = \alpha \boldsymbol{J}(\boldsymbol{f})$, trong đó $\alpha \in \mathbb{R}$

2. Cho trước ma trận $\boldsymbol{A}$ cỡ $m \times n$, véctơ $\boldsymbol{b}$ có chiều m, và véctơ $\boldsymbol{x}$ có n biến. Chứng minh

a) $\boldsymbol{J}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}$

b) $\boldsymbol{J}(\boldsymbol{b}) = \boldsymbol{0}$, từ đó suy ra $\boldsymbol{J}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{A}$

3 ($\star$)**.** Định nghĩa **gradient của hàm số** là véctơ mà mỗi thành phần lần lượt là đạo hàm riêng của hàm số đó theo mỗi biến. Cho các hàm véctơ $\boldsymbol{f}$, $\boldsymbol{g}$ xác định như trong bài 1. Khi đó $\boldsymbol{f}^T\boldsymbol{g}$ là một hàm số. Chứng minh

$$\nabla\left(\boldsymbol{f}^T\boldsymbol{g}\right) = \boldsymbol{J}(\boldsymbol{f})^T\boldsymbol{g} + \boldsymbol{J}(\boldsymbol{g})^T\boldsymbol{f}.$$

4. Cho trước ma trận $\boldsymbol{A}$, $\boldsymbol{B}$ cỡ $m \times n$, véctơ $\boldsymbol{c}$, $\boldsymbol{d}$ có chiều m, và véctơ $\boldsymbol{x}$ có n biến. Chứng minh

a) $\nabla\left[(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^T(\boldsymbol{C}\boldsymbol{x} - \boldsymbol{d})\right] = \boldsymbol{A}^T(\boldsymbol{C}\boldsymbol{x} - \boldsymbol{d}) + \boldsymbol{C}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$

b) $\nabla\left[\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2\right] = 2\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$