雷达对抗讲义

2025 年 9 月 3 日

目　录

第 1 章 基础知识

工欲善其事必先利其器, 本章节我们首先会对本课程用到的符号规则进行规定. 此外, 本课程会大量使用到矩阵方面的知识, 尤其是矩阵微积分, 因此我们也会对矩阵微积分进行简单的介绍.

1.1 符号规定

本课程中会大量使用到矩阵和向量, 相关符号书写规则如下：

1. 标量：小写字母, 如 a, b, c 或者 α, β, γ.
2. 向量：加粗小写字母, 如 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$.
3. 矩阵：加粗大写字母, 如 $\mathbf{A}, \mathbf{B}, \mathbf{C}$.

此外, 可以使用如下形式来表明向量或矩阵的维度：

1. 向量：$\boldsymbol{x} \in \mathbb{R}^{n\times 1}$ 或 $\boldsymbol{x} \in \mathbb{R}^{n}$ 表示 $n \times 1$ 的列向量.
2. 矩阵：$\mathbf{A} \in \mathbb{R}^{m\times n}$ 表示 $m \times n$ 的矩阵.

其中, $\mathbb{R}$ 表示实数域, 而复数域则表示为 $\mathbb{C}$.

需要注意的是, 对于向量和矩阵中的元素, 按照书写规则, 也应当用小写字母表示. 比如, 对于向量 $\boldsymbol{x}$, 其第 i 个元素可以表示为 x_i, 而对于矩阵 $\mathbf{A}$, 其第 i 行第 j 列的元素可以表示为 a_{ij}. 因此, $\boldsymbol{x}_i$ 或 $\mathbf{A}_i$ 这样的符号, 代表的是第 i 个向量或矩阵.

1.2 矩阵微积分

普通的求导想必大家都很熟悉, 比如对于标量函数 $f(x)$, 其关于 x 的导数可以表示为 $\frac{\partial f(x)}{\partial x}$. 而函数不仅仅可以是关于标量的函数，也可以是关于向量甚至矩阵的函数。比如，一个二元函数 $f(x_1, x_2)$ 则可以看作是关于向量 $\boldsymbol{x} = \left[\begin{smallmatrix} x_1 & x_2 \end{smallmatrix}\right]^{\mathrm{T}}$ 的一个函数 $f(\boldsymbol{x})$。自然地，我们也可以求解其关于向量 $\boldsymbol{x}$ 导数。具体而言，我们有如下定义

定义 1.1（**标量关于向量求导**） 对于向量 $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$，有映射 $f : \mathbb{R}^{n\times 1} \to \mathbb{R}$，则定义其关于向量 $\boldsymbol{x}$ 导数为

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}.$$

即计算 $f(\boldsymbol{x})$ 关于 $\boldsymbol{x}$ 的中每一个元素的偏导数，结果是一个和 $\boldsymbol{x}$ 同样维度的向量。

类似地，我们也可以定义标量对矩阵的求导，具体定义如下：

定义 1.2 （**标量关于矩阵求导**） 对于矩阵 $\mathbf{A} \in \mathbb{R}^{m\times n}$，有映射 $f: \mathbb{R}^{m\times n} \to \mathbb{R}$，则定义其关于矩阵 $\mathbf{A}$ 导数为

$$\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} = \begin{bmatrix} \frac{\partial f(\mathbf{A})}{\partial a_{11}} & \frac{\partial f(\mathbf{A})}{\partial a_{12}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial a_{1n}} \\ \frac{\partial f(\mathbf{A})}{\partial a_{21}} & \frac{\partial f(\mathbf{A})}{\partial a_{22}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial a_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{A})}{\partial a_{m1}} & \frac{\partial f(\mathbf{A})}{\partial a_{m2}} & \cdots & \frac{\partial f(\mathbf{A})}{\partial a_{mn}} \end{bmatrix}.$$

此外，函数本书也有可能不是一个标量，而是一个向量，因此我们也需要定义向量对向量的求导。具体定义如下：

定义 1.3 （**向量关于向量求导**） 对于向量 $\boldsymbol{x} \in \mathbb{R}^{m\times 1}$，有映射 $\boldsymbol{f}: \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times 1}$，则定义其关于向量 $\boldsymbol{x}$ 导数为

$$\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial \boldsymbol{x}} \\ \frac{\partial f_2(\boldsymbol{x})}{\partial \boldsymbol{x}} \\ \vdots \\ \frac{\partial f_m(\boldsymbol{x})}{\partial \boldsymbol{x}} \end{bmatrix}.$$

可以看到，对应的求导规则就是将函数的每一个元素分别对向量 $\boldsymbol{x}$ 求导，并将结果按照 $\boldsymbol{f}$ 的形式排列。比如，$m \times 1$ 列向量关于 $n \times 1$ 列向量的导数为一个 $mn \times 1$ 的列向量。如果是 $1 \times m$ 的行向量关于 $n \times 1$ 列向量的导数，那么根据规则，结果则是一个 $n \times m$ 的矩阵。

下面，我们会给出一些常见的矩阵微积分公式，这些公式将在本课程中反复用到。

例 1.1 设有向量 $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$、$\boldsymbol{y} \in \mathbb{R}^{n\times 1}$ 和矩阵 $\mathbf{A} \in \mathbb{R}^{n\times n}$，计算如下导数

1. $f = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ 关于 $\boldsymbol{x}$ 的导数
2. $f = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ 关于 $\boldsymbol{y}$ 的导数
3. $f = \boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{y}$ 关于 $\boldsymbol{x}$ 的导数
4. $f = \boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{y}$ 关于 $\mathbf{A}$ 的导数
5. $\boldsymbol{f} = \boldsymbol{x}^{\mathrm{T}}\mathbf{A}$ 关于 $\boldsymbol{x}$ 的导数
6. $\boldsymbol{f} = (\boldsymbol{x} - \boldsymbol{y})^{\mathrm{T}}$ 关于 $\boldsymbol{x}$ 的导数

解

1. 注意到 $f = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \sum_{i=1}^{n} x_i y_i$，因此有

$$\frac{\partial f}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \boldsymbol{y}.$$

2. 同理，对于 $f=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ 有

$$\frac{\partial f}{\partial \boldsymbol{y}}=\begin{bmatrix}\frac{\partial f}{\partial y_1}\\ \frac{\partial f}{\partial y_2}\\ \vdots\\ \frac{\partial f}{\partial y_n}\end{bmatrix}=\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix}=\boldsymbol{x}.$$

3. 同样地，将函数写成求和的形式 $f=\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{y}=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_iy_j$，因此有

$$\frac{\partial f}{\partial \boldsymbol{x}}=\begin{bmatrix}\frac{\partial f}{\partial x_1}\\ \frac{\partial f}{\partial x_2}\\ \vdots\\ \frac{\partial f}{\partial x_n}\end{bmatrix}=\begin{bmatrix}\sum_{j=1}^{n}a_{1j}y_j\\ \sum_{j=1}^{n}a_{2j}y_j\\ \vdots\\ \sum_{j=1}^{n}a_{nj}y_j\end{bmatrix}=\mathbf{A}\boldsymbol{y}.$$

4. 同理，对于 $f=\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{y}$ 有

$$\frac{\partial f}{\partial \mathbf{A}}=\begin{bmatrix}\frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \cdots & \frac{\partial f}{\partial a_{1n}}\\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{\partial a_{22}} & \cdots & \frac{\partial f}{\partial a_{2n}}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial f}{\partial a_{n1}} & \frac{\partial f}{\partial a_{n2}} & \cdots & \frac{\partial f}{\partial a_{nn}}\end{bmatrix}=\begin{bmatrix}x_1y_1 & x_1y_2 & \cdots & x_1y_n\\ x_2y_1 & x_2y_2 & \cdots & x_2y_n\\ \vdots & \vdots & \ddots & \vdots\\ x_ny_1 & x_ny_2 & \cdots & x_ny_n\end{bmatrix}=\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}.$$

5. 最后，对于 $\boldsymbol{f}=\boldsymbol{x}^{\mathrm{T}}\mathbf{A}$，其第 j 个元素为 $f_j=\sum_{i=1}^{n}a_{ij}x_i$，因此有

$$\frac{\partial f}{\partial \boldsymbol{x}}=\begin{bmatrix}\frac{\partial f_1}{\partial \boldsymbol{x}} & \frac{\partial f_2}{\partial \boldsymbol{x}} & \cdots & \frac{\partial f_n}{\partial \boldsymbol{x}}\end{bmatrix}=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}=\mathbf{A}.$$

6. 对于 $\boldsymbol{f}=\boldsymbol{x}-\boldsymbol{y}$，有

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}=\begin{bmatrix}\frac{\partial(x_1-y_1)}{\partial \boldsymbol{x}} & \frac{\partial(x_2-y_2)}{\partial \boldsymbol{x}} & \cdots & \frac{\partial(x_n-y_n)}{\partial \boldsymbol{x}}\end{bmatrix}=\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}=\mathbf{I}.$$

从以上的例子我们可以得到一些经验结论：

1. 对于标量关于向量的求导，如果向量位于表达式的左侧且有转置，那么导数则直接是右侧变量，比如 $\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}{\partial \boldsymbol{x}}=\boldsymbol{y}$ 和 $\frac{\partial \boldsymbol{x}^{\mathrm{T}}\mathbf{A}}{\partial \boldsymbol{x}}=\mathbf{A}$。
2. 对于标量关于向量的求导，如果向量位于表达式的右侧，那么导数则是左侧变量加转置，比如 $\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}{\partial \boldsymbol{y}}=\boldsymbol{x}$ 和 $\frac{\partial \mathbf{A}\boldsymbol{y}}{\partial \boldsymbol{y}}=\mathbf{A}^{\mathrm{T}}$。

矩阵微分有四个常用的性质：线性、乘积、商和链式法则，具体如下：

性质 1.1 **（矩阵微积分的四个性质）** 以下性质中的前三个与标量函数求导的性质类似，只有最后一个链式法则略有不同。

1. 线性

$$\frac{\partial(af(\boldsymbol{x})+bg(\boldsymbol{x}))}{\partial \boldsymbol{x}}=a\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}+b\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}. \tag{1.1}$$

2. 乘积

$$\frac{\partial f(\boldsymbol{x})g(\boldsymbol{x})}{\partial \boldsymbol{x}}=\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}g(\boldsymbol{x})+f(\boldsymbol{x})\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}. \tag{1.2}$$

3. 商

$$\frac{\partial \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}}{\partial \boldsymbol{x}}=\frac{f'(\boldsymbol{x})g(\boldsymbol{x})-f(\boldsymbol{x})g'(\boldsymbol{x})}{g^2(\boldsymbol{x})}. \tag{1.3}$$

4. 链式法则

$$\frac{\partial f(\mathbf{g}(\boldsymbol{x}))}{\partial \boldsymbol{x}}=\frac{\partial \mathbf{g}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\frac{\partial f(\mathbf{g})}{\partial \mathbf{g}}. \tag{1.4}$$

例 1.2 计算如下函数关于向量 $\boldsymbol{x}$ 的导数

$$f(\boldsymbol{x})=\|\mathbf{A}\boldsymbol{x}-\boldsymbol{y}\|^2.$$

解 方法 1：将函数展开，有

$$f(\boldsymbol{x})=\|\mathbf{A}\boldsymbol{x}-\boldsymbol{y}\|^2=(\mathbf{A}\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}(\mathbf{A}\boldsymbol{x}-\boldsymbol{y})=\boldsymbol{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\boldsymbol{x}-2\boldsymbol{y}^{\mathrm{T}}\mathbf{A}\boldsymbol{x}+\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}.$$

因此有

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}=2\mathbf{A}^{\mathrm{T}}\mathbf{A}\boldsymbol{x}-2\mathbf{A}^{\mathrm{T}}\boldsymbol{y}.$$

方法 2：记 $\mathbf{g}(\boldsymbol{x})=\mathbf{A}\boldsymbol{x}-\boldsymbol{y}$，则 $f(\boldsymbol{x})$ 可以写为如下的复合函数形式

$$f(\boldsymbol{x})=\mathbf{g}(\boldsymbol{x})^{\mathrm{T}}\mathbf{g}(\boldsymbol{x}).$$

根据链式法则，有

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}=\frac{\partial \mathbf{g}^{\mathrm{T}}(\boldsymbol{x})}{\partial \boldsymbol{x}}\frac{\partial f(\mathbf{g})}{\partial \mathbf{g}}=\frac{(\mathbf{A}\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}}{\partial \boldsymbol{x}}\frac{\partial \mathbf{g}^{\mathrm{T}}\mathbf{g}}{\partial \mathbf{g}}=2\mathbf{I}\mathbf{g}=2\mathbf{A}^{\mathrm{T}}(\mathbf{A}\boldsymbol{x}-\boldsymbol{y}).$$

1.3 复矩阵微积分

在本课程中，还有可能涉及到复数矩阵的微积分。由于在实际应用中，大部分函数都是关于复向量或者复矩阵的实值函数，因此我们只针对这种函数给出其导数的定义，具体如下：

定义 1.4 对于复数 $z=x+jy$，有映射 $g:\mathbb{C}\to\mathbb{R}$，则定义其关于复数 z 的导数为

$$\frac{\partial g(z)}{\partial z}=\frac{\partial g(z)}{\partial x}+j\frac{\partial g(z)}{\partial y}.$$

例 1.3 设有复数 $z=x+jy$，计算 $g(z)=z\overline{z}$ 关于 z 的导数。

解 将目标函数展开，有

$$g(z)=z\bar{z}=(x+jy)(x-jy)=x^2+y^2.$$

因此根据定义有

$$\frac{\partial g(z)}{\partial z}=\frac{\partial g(z)}{\partial x}+j\frac{\partial g(z)}{\partial y}=2x+j2y=2z.$$

利用上述定义，我们同样得到类似的实值函数关于复向量和复矩阵的导数。下面我们通过一些简单的例子，给出一些常用的经验公式。

例 1.4 计算如下函数关于复向量 $\boldsymbol{z}$ 的导数

$$g(\boldsymbol{z})=\boldsymbol{z}^{\mathrm{H}}\boldsymbol{z}.$$

解 记 $\boldsymbol{z}=\boldsymbol{x}+j\boldsymbol{y}$，其中 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 分别是实部和虚部向量，那么目标函数可以展开为

$$g(\boldsymbol{z})=\boldsymbol{z}^{\mathrm{H}}\boldsymbol{z}=(\boldsymbol{x}-j\boldsymbol{y})(\boldsymbol{x}+j\boldsymbol{y})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}.$$

因此根据定义有

$$\frac{\partial g(\boldsymbol{z})}{\partial \boldsymbol{z}}=\frac{\partial g(\boldsymbol{z})}{\partial \boldsymbol{x}}+j\frac{\partial g(\boldsymbol{z})}{\partial \boldsymbol{y}}=2\boldsymbol{x}+j2\boldsymbol{y}=2\boldsymbol{z}.$$

例 1.5 计算如下函数关于复向量 $\boldsymbol{z}$ 的导数

$$g(\boldsymbol{z})=\boldsymbol{z}^{\mathrm{H}}\mathbf{R}\boldsymbol{z},$$

其中 $\mathbf{R}$ 是一个共轭对称矩阵。

解 记 $\boldsymbol{z}=\boldsymbol{x}+j\boldsymbol{y}$，$\mathbf{R}=\mathbf{P}+j\mathbf{Q}$，那么目标函数可以展开为

$$\begin{aligned}g(\boldsymbol{z})&=\boldsymbol{z}^{\mathrm{H}}\mathbf{R}\boldsymbol{z}=(\boldsymbol{x}-j\boldsymbol{y})^{\mathrm{T}}(\mathbf{P}+j\mathbf{Q})(\boldsymbol{x}+j\boldsymbol{y})\\&=\boldsymbol{x}^{\mathrm{T}}\mathbf{P}\boldsymbol{x}+\boldsymbol{y}^{\mathrm{T}}\mathbf{P}\boldsymbol{y}+\boldsymbol{y}^{\mathrm{T}}\mathbf{Q}\mathbf{x}-\boldsymbol{x}^{\mathrm{T}}\mathbf{Q}\boldsymbol{y}+j(\boldsymbol{x}^{\mathrm{T}}\mathbf{Q}\boldsymbol{x}-\boldsymbol{y}^{\mathrm{T}}\mathbf{P}\boldsymbol{x}+\boldsymbol{x}^{\mathrm{T}}\mathbf{P}\boldsymbol{y}+\boldsymbol{y}^{\mathrm{T}}\mathbf{Q}\boldsymbol{y}).\end{aligned}$$

注意到，$\mathbf{R}=\mathbf{R}^{\mathrm{H}}$，即

$$\mathbf{P}+j\mathbf{Q}=\mathbf{P}^{\mathrm{T}}-j\mathbf{Q}^{\mathrm{T}},$$

因此，我们有

$$\begin{cases}\mathbf{P} &=\mathbf{P}^{\mathrm{T}}\\ \mathbf{Q} &=-\mathbf{Q}^{\mathrm{T}}\end{cases}.$$

此外，对于任意一个向量 $\boldsymbol{x}$，我们有

$$\begin{cases}\boldsymbol{x}^{\mathrm{T}}\mathbf{Q}\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}(-\mathbf{Q}^{\mathrm{T}})\boldsymbol{x}=-\boldsymbol{x}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\boldsymbol{x}\\ \boldsymbol{x}^{\mathrm{T}}\mathbf{Q}\boldsymbol{x}=(\boldsymbol{x}^{\mathrm{T}}\mathbf{Q}\boldsymbol{x})^{\mathrm{T}}=\boldsymbol{x}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\boldsymbol{x},\end{cases}$$

也就是说 $\boldsymbol{x}^{\mathrm{T}}\mathbf{Q}\boldsymbol{x}=-\boldsymbol{x}^{\mathrm{T}}\mathbf{Q}\boldsymbol{x}=0$ 。因此，$g(\boldsymbol{z})$ 可以简化为

$$g(\boldsymbol{z})=\boldsymbol{x}^{\mathrm{T}}\mathbf{P}\boldsymbol{x}+\boldsymbol{y}^{\mathrm{T}}\mathbf{P}\boldsymbol{y}+\boldsymbol{y}^{\mathrm{T}}\mathbf{Q}\mathbf{x}-\boldsymbol{x}^{\mathrm{T}}\mathbf{Q}\boldsymbol{y}.$$

根据定义，我们有

$$\frac{\partial g(\boldsymbol{z})}{\partial \boldsymbol{z}} = \frac{\partial g(\boldsymbol{z})}{\partial \boldsymbol{x}} + j\frac{\partial g(\boldsymbol{z})}{\partial \boldsymbol{y}} = 2\mathbf{P}\boldsymbol{x} - 2\mathbf{Q}\boldsymbol{y} + j(2\mathbf{P}\boldsymbol{y} + 2\mathbf{Q}\boldsymbol{x})$$
$$= 2(\mathbf{P} + j\mathbf{Q})(\boldsymbol{x} + j\boldsymbol{y}) = 2\mathbf{R}\mathbf{z}.$$

1.4 常见统计量

在概率论中，我们经常会遇到一些常见的统计量，比如均值、方差、协方差等。这些统计量在本课程中也会经常用到。不同的是，本课程中的对象都是采集到的离散数据，对应向量或矩阵而不是随机变量。因此，我们需要提前了解对于向量和矩阵，如何计算这些统计量。

设有向量 $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$ 和 $\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^{\mathrm{T}}$，两者都可以看作是某个随机变量的 n 个观测。对应的均值、方差和协方差有如下计算公式：

1. 均值：$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\boldsymbol{x}^{\mathrm{T}}\mathbf{1}$。
2. 方差：$\mathrm{Var}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}(\boldsymbol{x} - \bar{x}\mathbf{1})^{\mathrm{T}}(\boldsymbol{x} - \bar{x}\mathbf{1})$，如果均值为零，则 $\mathrm{Var}(\boldsymbol{x}) = \frac{1}{n}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$。
3. 协方差：$\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}(\boldsymbol{x} - \bar{x}\mathbf{1})^{\mathrm{T}}(\boldsymbol{y} - \bar{y}\mathbf{1})$，如果均值为零，则 $\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{n}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$。

设有矩阵 $\mathbf{X} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_m \end{bmatrix} \in \mathbb{R}^{n\times m}$，其中每一列都是一个向量，假设均值为零，那么向量两两之间的协方差可以构成一个协方差矩阵 $\boldsymbol{\Sigma}$，其第 i, j 个元素为

$$\sigma_{ij} = \mathrm{Cov}(\boldsymbol{x}_i, \boldsymbol{x}_j).$$

如果均值为零，那么协方差矩阵可以表示为

$$\boldsymbol{\Sigma} = \frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X} \in \mathbb{R}^{m\times m}.$$

例 1.6 设有一个未知的数据矩阵 $\mathbf{X} \in \mathbb{R}^{n\times m}$，但其协方差矩阵 $\boldsymbol{\Sigma}$ 是已知的，请给出任意投影方向 $\boldsymbol{v} \in \mathbb{R}^{m\times 1}$ 下数据的方差。

解　投影后的数据向量为 $\boldsymbol{x} = \mathbf{X}\boldsymbol{v}$，其方差为

$$\mathrm{Var}(\boldsymbol{x}) = \frac{1}{n}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \frac{1}{n}(\mathbf{X}\boldsymbol{v})^{\mathrm{T}}(\mathbf{X}\boldsymbol{v}) = \frac{1}{n}\boldsymbol{v}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{v} = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{v}.$$

因此，数据在任意投影方向上的方差可以由 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{v}$ 给出。

1.5 拉格朗日乘数法

对于等式约束的优化问题：

$$\begin{cases} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h(\boldsymbol{x}) = \mathbf{0}. \end{cases}$$

其取得极值的必要条件如下：

1. $h(\boldsymbol{x})=0$
2. $\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}+\lambda\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}}=0$

其中 $\lambda \geq 0$.

该如何理解上述两个条件呢？对于条件 1，很显然它必须成立，因为该条件本身就是约束。对于条件 2，则可以分为两种情况。情况 1：该目标函数取得极值时，目标函数的导数并不为 0 。此时，条件 2 表明 $\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}=-\lambda\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}}$，也就是说目标函数梯度下降的方向与等式约束对应的函数的梯度方向刚好反向。这意味着想要进一步降低目标函数值，必然会影响等式约束的成立，因此，此时函数取得了极值。情况 2：设 $\boldsymbol{x}^*$ 处目标取得极值，并且 $h(\boldsymbol{x}^*)=0$，也就是说目标函数取得极值时，等式约束刚好满足。此时 $\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}=\mathbf{0}$，显然当 $\lambda=0$ 时，条件 2 是成立的。

根据如上观察，我们可以归纳出拉格朗日乘数法，其步骤如下：

1. 构建拉格朗日函数
$$\mathcal{L}(\boldsymbol{x},\lambda)=f(\boldsymbol{x})+\lambda h(\boldsymbol{x}).$$
2. 对 $\boldsymbol{x}$ 和 λ 分别求导，并令其为 0，得到方程组
$$\begin{cases}\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}+\lambda\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}}=0\\ h(\boldsymbol{x})=0\end{cases}.$$
3. 求解上述方程组，得到 $(\boldsymbol{x}^*,\lambda^*)$。

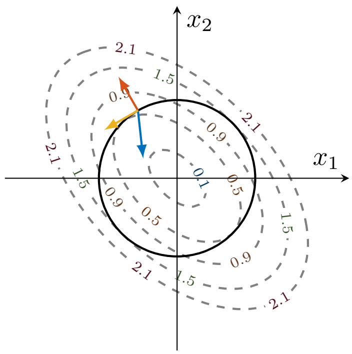
例 1.7 设有如下的优化问题
$$\begin{cases}\min \quad f(\boldsymbol{x})=x_1^2+x_1x_2+x_2^2\\ \text{s.t.} \quad h(\boldsymbol{x})=x_1^2+x_2^2-1=0\end{cases}.$$
请给出该问题的拉格朗日函数，并求解最优解。

解 令向量 $\boldsymbol{x}=\left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right]$，则目标函数和约束函数可以写成如下的形式
$$f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\mathbf{R}\boldsymbol{x},\quad h(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}-1,$$
其中 $\mathbf{R}=\begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & 1\end{bmatrix}$。对应的拉格朗日函数为
$$\mathcal{L}(\boldsymbol{x},\lambda)=\boldsymbol{x}^{\mathrm{T}}\mathbf{R}\boldsymbol{x}+\lambda(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}-1).$$
对 $\boldsymbol{x}$ 和 λ 分别求导，并令其为 0，得到方程组
$$\begin{cases}\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}+\lambda\frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}}=0\\ h(\boldsymbol{x})=0\end{cases}\Rightarrow\begin{cases}2\mathbf{R}\boldsymbol{x}+2\lambda\boldsymbol{x}=0\\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}-1=0\end{cases}.$$
不难发现 $\mathbf{R}\boldsymbol{x}=-\lambda\boldsymbol{x}$，即 $\boldsymbol{x}$ 是 $\mathbf{R}$ 的特征向量，对应的特征值为 $-\lambda$。而矩阵 $\mathbf{R}$ 有如

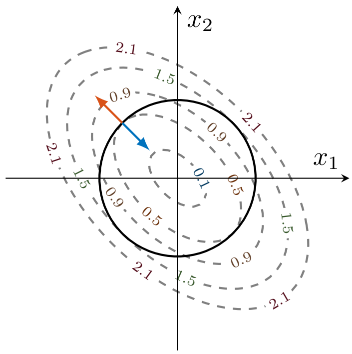
下特征分解：

$$\mathbf{R} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^{\mathrm{T}} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}.$$

显然，其最小特征值为 $\frac{1}{2}$，对应的模长为 1 的特征向量为 $\boldsymbol{x} = \pm \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$。



(a) 不满足极值条件



(b) 满足极值条件

图 1.1 等式约束的优化问题示例

从图 1.1a 可以看到，此时的点并不满足极值条件，目标函数的梯度（蓝色箭头）和约束函数的梯度（红色箭头）并不共线，这意味着沿着黄色箭头的方向，目标函数仍然是可以继续下降的，因此该点并不是极值点。而在极值点处，如图 1.1b 所示，目标函数的梯度和约束函数的梯度是共线的，这意味着在该点处，想要进一步降低目标函数值，必然会影响等式约束的成立，因此该点是极值点。