线性代数

-28 课 正定矩阵与最小值

一、知识概要

本节围绕着正定矩阵这一主题，将之前介绍的知识：主元，行列式，特征值等联系起来，并通过正定矩阵判据式 x^TAx 引出矩阵与函数之间的关系与对应特点，介绍如何找到函数最小值。并绘制对应函数图像，给出几何上的解释。

二．正定矩阵

我们从 2*2 的对称矩阵开始研究：

设 A = $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$，如何判定其是否为对称矩阵？

给出下面 4 种判定：

1）特征值判定：$\lambda_1 > 0$，$\lambda_2 > 0$。即特征值皆为正数。

2）行列式判定：a ＞ 0，ac － b^2 ＞ 0。即顺序主子式均为正值。

3）主元判定：a ＞ 0，$\frac{ac-b^2}{a} > 0$。主元均为正数。

4）*判据式：x^TAx ＞ 0。

注： 线性代数范围内，正定矩阵需要是对称阵。

例如设矩阵 A = $\begin{bmatrix} 2 & 6 \\ 6 & ? \end{bmatrix}$，则其中？处填入 18 以上的整数，该矩阵皆为正定。但是如果其中填的正好是 18，此时该矩阵行列式为 0，此时的 A 矩阵称为半正定矩阵，只有一个主元 2，且 A 为奇异矩阵。特征值为 0,20，都≥0。

好的，上面我们从普通的主元，特征值，行列式角度分析了矩阵 A，那么接下来我们就最重要的判据式：x^TAx ＞ 0 来展开讨论。

先说说 A = $\begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix}$ 这个半正定的情况，我们根据判据式，得到如下算式：

$$x^T\mathrm{A}x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underset{\substack{\uparrow \\ a}}{2}{x_1}^2 + \underset{\substack{\uparrow \\ 2b}}{12}x_1x_2 + \underset{\substack{\uparrow \\ c}}{18}{x_2}^2$$

我们看到，对应的${x_1}^2$，x_1x_2，${x_2}^2$前的系数分别为 $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ 中的 a，2b，c

这就是二次型，而所谓判定 A 是不是正定矩阵，也就是判别式x^TAx 所构造

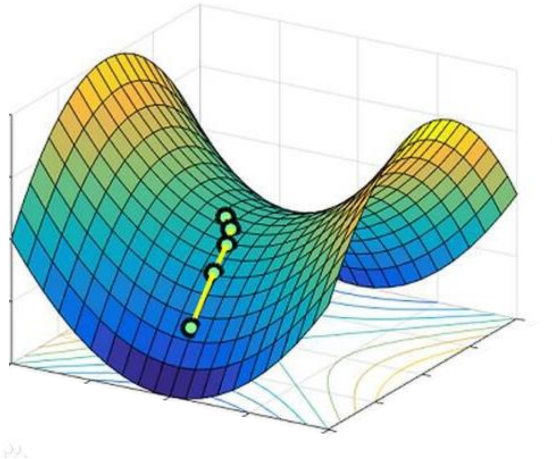
出的类似于 $2{x_1}^2 + 12x_1x_2 + 18{x_2}^2$的二次型是不是恒大于 0。对于本例中的 $2{x_1}^2 + 12x_1x_2 + 18{x_2}^2$来说，它显然不是恒正的，因为在某种取值中，其结果可以为 0。

这表明了,A 为半正定时，其对应二次型会在某种情况下得到 0，那么 A 不是正定时，其对应的x^TAx 会呈现出一种什么样的状态呢？

这次取 A = $\begin{bmatrix} 2 & 6 \\ 6 & 7 \end{bmatrix}$观察一下，此时得到的二次型为：

$$x^T\mathrm{A}x = 2{x_1}^2 + 12x_1x_2 + 7{x_2}^2$$

将其看做函数，f(x, y) = $2x^2 + 12xy + 7y^2$，绘制在坐标轴中就得到一个马鞍面，如图（1）



图（1）

其中有一个鞍点，在不同方向上观察，它会呈现出不同的性质，很容易看出来，这个鞍点是某个方向上的极大值，也是某个方向上的极小值。其中的最佳观测方向是沿特征向量的方向。OK,到这里我们就不在深入说了，很显然此时的 A 并不是一个正定矩阵。

说完了半正定与非正定矩阵，接下来我们举一个是正定矩阵的例子，并展开说一说它的特点以及二次型与我们的正定矩阵之间的联系。

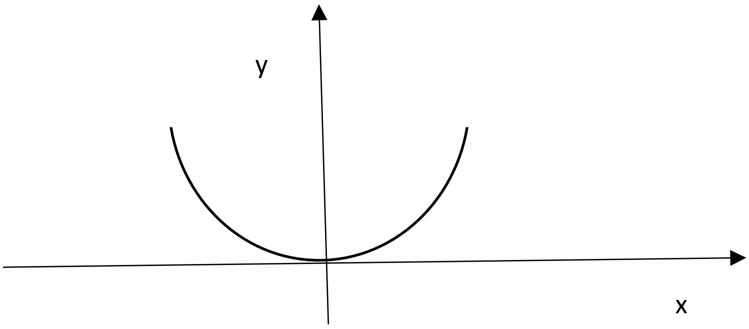
取矩阵 A = $\begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix}$，根据上面给出的（1）（2）（3）判据，很容易得到：这是一个正定矩阵。我们重点在判据式x^TAx，经计算，它的二次型为：

$$x^T\mathrm{A}x = 2{x_1}^2 + 12x_1x_2 + 20{x_2}^2$$

写成函数形式就是：

$$\mathrm{f}(x, y) = 2x^2 + 12xy + 20y^2$$

其图像切面为类似二次



图（二）

很明显，$f(x,y) = 2x^2 + 12xy + 20y^2$的极小值点在原点位置，根据微积分中学习的知识，原点一阶偏导数为 0，二阶偏导数大于 0。故可以判断其为最小值点。就是说在微积分中，我们判断是否有最小值可以通过求导判断是否大于 0，而在线性代数中，我们判断是否有最小值是通过判断二阶导数矩阵是否为正定，这是正定矩阵的用法之一，将数字转化为矩阵。

另外，我们将 $f(x,y)$配方，得到：

$$f(x,y) = 2x^2 + 12xy + 20y^2 = 2(x+3y)^2 + 2y^2$$

很容易想到，如果 A 不是正定的，那么y^2前的系数势必为负数。而如果我们用 f = 1 这个平面来截这个曲线，则上面的 $f(x,y) = 2x^2 + 12xy + 20y^2$截得的就是椭圆。

另外，这里还要介绍一点，配方法反映在线代中就是消元，$A = \begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix}$。这上的每一个元素都表示着 $f(x,y)$中对应项前的系数，经过消元，得到消元后的矩阵为：$\begin{bmatrix} 2 & 6 \\ 0 & 2 \end{bmatrix}$，对应看来：

$$2(x+3y)^2 + 2y^2$$

$$\begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} * \begin{bmatrix} 2 & 6 \\ 0 & 2 \end{bmatrix}$$

关系显而易见。

括号外的系数是主元，这样一切就联系起来了，

正定→主元为正→二次型平方项外系数为正→图像朝上→原点为最小值点。

此理论可推广到 n 维。同样，在微积分中我们学习二阶偏导数极值时的f_{xx}，

f_{yy}与$(f_{xy})^2$之间的关系也可以反映到矩阵的正定判断上，即矩阵$\begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$, 这就是二阶导数矩阵。直接将原来的$\begin{bmatrix} 2x^2 & 6xy \\ 6xy & 20y^2 \end{bmatrix}$对应元素求导。判定是否有极小值的条件即为它是否为正定矩阵。同样，这个判断也可以推广到 n*n 的矩阵中。我们接下来举一个 3 维椭球体与矩阵联系得到的特殊性质：

【例】 矩阵 A = $\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$，判定这个矩阵反映的性质。

很容易判断这个矩阵是一个正定矩阵，写出其二次型：

$$x^T\mathrm{A}\mathrm{x} = 2{x_1}^2 + 2{x_2}^2 + 2{x_3}^2 - 2x_1x_2 - 2x_2x_3$$

沿用上面的方法，截出一个椭球体 $2{x_1}^2 + 2{x_2}^2 + 2{x_3}^2 - 2x_1x_2 - 2x_2x_3 = 1$ 这时此椭球体的三个轴的方向即为 A 矩阵的特征向量方向，三个轴的长度即为特征值的大小。我们可以通过分解 A 矩阵： $\mathrm{A} = \mathrm{Q} \wedge Q^T$ 得到其对应值。这个特殊的性质我们称为：主轴定理。

三．学习感悟

本节中我们从学习正定矩阵的判定入手，主要研究了判定式$x^T\mathrm{Ax}$的二次型以及其二次型反映到具体函数上所表现出的具体性质。这一节将微积分中学习的二阶导判定以及几何角度上的图像走势与正定矩阵相联系起来，使我们学会使用正定矩阵解决更多的实际问题。