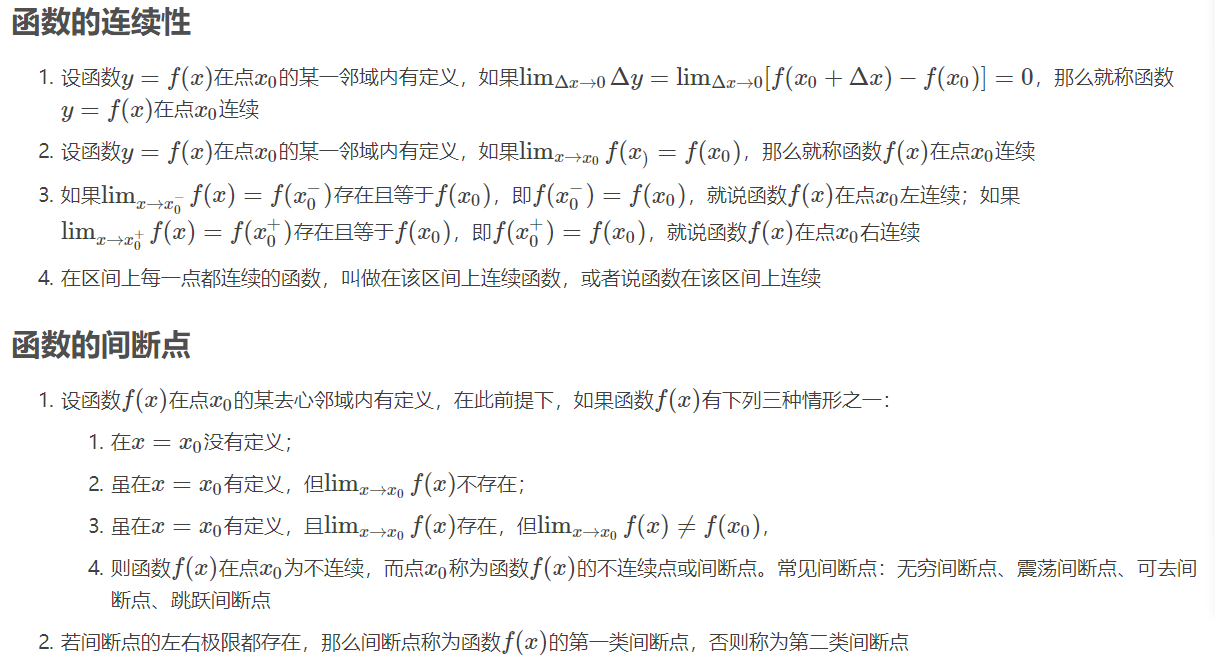
# 3.微积分

**函数、极限、连续性、偏导**

1，连续函数一定有极限，有极限不一定连续（如果函数在某点是连续的，那么它在这个点一定取一个非无穷大的值，就有极限；反之，分段函数在某点可能有极限，因为极限值不一定等于函数值）。  
2，连续不一定可导（绝对值函数），可导一定连续。  
3，可导等于可微。

**基本求导法则：**

（u+v）’=u’+v’

(uv)’ = u’v+uv’,(cu)’=cu’

(u/v)’ = (u’v-uv’ )/v^2

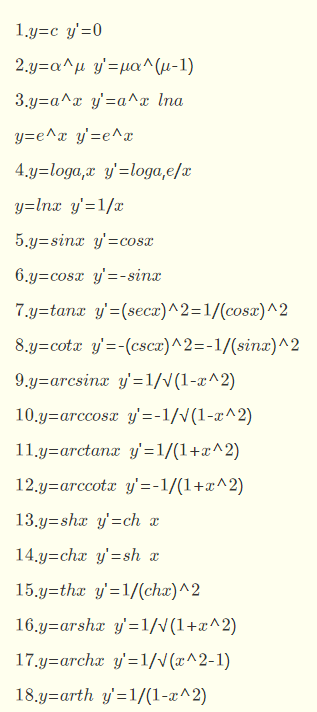
**反函数的导数：**

dy/dx =1/(dx/dy)

**复合函数导数：**

dy/dx =( dy/du) \*( du/dx)

**基本初等函数导数**



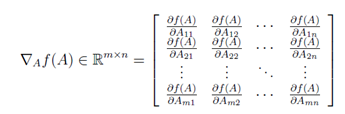


矩阵微积分

之前章节的内容，在一般线性代数的课程中都会讲到。而有些常用的内容是没有的，这就是把微积分推广到向量。事实上，我们应用的微积分都会比较繁琐，各种符号总是让问题变得更复杂。在本节中，将给出一些矩阵微积分的基本定义，并举例说明。

**4.1梯度**

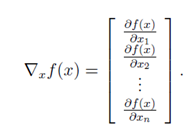
设ƒ:Rm×n→R是大小为m×n的矩阵A的函数，且返回值为实数。ƒ的梯度（关于A∈Rm×n）是一个偏导矩阵，定义如下：



即,一个m×n矩阵，其中

图片

注意∇Af(A)和A有相同的大小。所以，特别的，当A是一个向量x∈Rn时，



需要特别记住的是，函数的梯度只在函数值为实数的时候有定义。也就是说，函数一定要返回一个标量。例如，我们就不能对Ax，A∈Rn×n中的x求梯度，因为它是一个向量。

它遵循和偏导相同的性质：

图片

原则上，梯度是多变量函数偏导的延伸。然而，实际应用梯度时，会因为数学符号而变得棘手。例如，假设A∈Rm×n是一个具有固定系数的矩阵，b∈Rm是一个固定系数的向量。令ƒ ：Rm→R为由ƒ(z)=zTz，因此∇zf(z) =2z。现在，考虑表达式;

∇f(Ax)

上式该如何理解？至少有两种解释：

1. 解释一，因∇f(Ax). = 2z,所以可将∇f(Ax).理解为点Ax处的梯度，那么：

∇f(Ax) = 2(Ax) = 2Ax ∈ Rm

解释二，可以认为f(Ax)是关于变量x的函数。正式的表述为，令g(x) = f(Ax)。那么在此种解释下有：

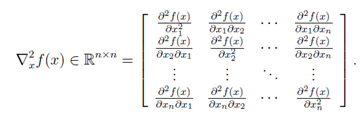
∇f(Ax) = ∇xg(x) ∈ Rn

大家可以发现，这两种解释确实不同。解释一得出的结果是m维向量，而解释二得出n维向量！怎么办？

这里的关键是确定对那个变量求微分。在第一种情况下，是让函数f对参数z求微分，然后代入参数Ax。第二种情况，是让复合函数g（x）= F（AX）与直接对x求微分。第一种情况记为∇zf（AX），第二种情况记为∇xf（AX）。你会在作业中发现，理清数学符号是非常重要的。

**4.2Hessian矩阵**

假设 ƒ ：Rn→R    是n维向量A的的函数，并返回一个实数。那么x的Hessian矩阵是偏导数的n×n矩阵，写作∇2xf（x），简记为H。



换句话说，∇2xf(x) ∈ Rn×n ，其中：

图片

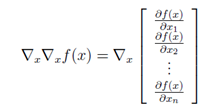
需要注意的是Hessian矩阵始终是对称的，即：

图片

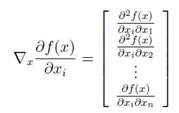
和梯度类似，Hessian矩阵只在f(x)为实数时有定义。

可以很自然联想到，偏导类似于函数的一阶导数，而Hessian类似函数的的二阶导数（我们使用的符号，也表明了这种联系）。通常这种直觉是正确的，但有些注意事项需要牢记。

首先，只有一个变量的实值函数，f : R→R，它的基本定义是二阶导数是一阶导数的导数，即：



然而，对于关于向量的函数，该函数的梯度是一个向量，我们不能取向量的梯度，即;



并且这个表达式没有定义。因此，不能说Hessian矩阵是梯度的梯度。然而，在下面的意义上比较靠谱：如果我们取第i项（∇xf（X））i =∂F（X）/∂xi，并取对x的梯度，我们得到：

图片

这是Hessian矩阵的第i列（或行）。 因此：



如果此处稍粗略一点，可以得出图片，只要将其真实的含义理解为对 (∇xf(x))的每一项求梯度，而不是对向量求梯度即可。

最后注意，虽然可求出对矩阵A∈Rn的梯度，但在本课程中，将只考虑向量x∈Rn的Hessian矩阵。这仅仅是为了方便起见（而事实上，没有计算需要求矩阵的Hessian矩阵），因为矩阵的Hessian矩阵必须表示为所有的偏导数∂2f（A）/（∂Aij∂Akℓ），而要表示为矩阵却相当麻烦。

**4.3    二次函数或线性函数的梯度和Hessian矩阵**

现在，让我们确定一些简单函数的梯度和Hessian矩阵。应当指出的是，这里给出的所有的梯度都是在CS229讲义给出的特殊情况。

当x∈Rn，对于已知向量b∈Rn，令f（X）= bT x。 得：

图片

因此

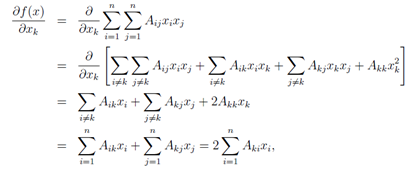
图片

由此不难看出，∇xbT x= b。这是与单变量微积分类似的情况，其中，∂/（∂x）aX =a。

现在考虑二次函数f（x）= xTAx ,A∈Sn。注意到：

图片

求其偏导数，分别考虑包含Xk和xk2因子的项：



其中最后一个等式是因为A是对称的（完全可以假设，因为它是二次型）。注意，∇xf（x）的第k项只是A的第k行和x的内积。因此，∇xxTAx=2AX。同样，与单变量微积分类似，即∂/（∂x）    ax2= 2aX。

最后，再看二次函数f（X）= xTAx的Hessian矩阵（显然，线性函数bT x的Hessian矩阵为零）。 在这种情况下，

图片

因此，应当清楚的是∇x2xTAx=2A，这完全是可证明的（并再次类似于单变量的情况∂2/(∂x2) ax2 = 2a）。

总之：

∇xbT x = b

∇xxTAx = 2Ax ( A 为对称矩阵)

∇x2xTAx = 2A ( A 为对称矩阵)

**4.4最小二乘法**

这里将用最后一节得到的公式推导最小二乘方程。假设对矩阵A∈Rm×n（为简单起见，假定A是满秩）和向量b∈Rm    ，使得b错误!未找到引用源。R（A）。在这种情况下，无法找到一个向量x∈Rn，使得Ax = b。退一步，我们找一个向量x∈Rn，使得Ax是尽可能接近b，即欧氏范数||Ax - b||22。

且知||x||22=xTx，有：

图片

取对已有x的梯度，并使用上一节推出的性质

图片

让最后一个表达式等于零，并求解X满足的标准方程

图片

这正和我们课上推导的一样。

**4.5行列式的梯度**

现在考虑一种情况，求函数对矩阵的梯度，即对A∈Rn×n，求∇A| A |。回顾之前关于行列式的讨论：

图片

因此：

图片

根据伴随矩阵的性质，可立即得出：

图片

现在，考虑函数f : Sn ++ → R, f(A) = log |A|，需要注意的是，一定要限制f的域是正定矩阵，因为这将确保| A | >0，这样log| A |是一个实数。在这种情况下，我们可以使用链式法则（很简单，只是单变量微积分的普通链式法则）得出：

图片

那么，很显然：

图片

此处，在最后一个表达式中去掉了转置符，因为A是对称的。注意当∂/(∂x) log x = 1/x时,和单值情况相似。

**4.6最优化特征值**

最后，通过直接分析特征值/特征向量，用矩阵微积分来解决一个优化问题。接下来，考虑等式约束优化问题：

图片

对于一个对称矩阵A ∈ Sn，解决等式约束优化问题的标准方法是构造拉格朗日（一个包括等式约束的目标函数）。这种情况下的拉格朗日可由下式给出：

图片

其中λ被称为与等式约束对应的拉格朗日乘子。对这问题可以找到一个x\*的最佳点，让拉格朗日的梯度在x\*上为零（这不是唯一的条件，但它是必需的）。 即：

图片

注意，这其实是线性方程组Ax =λx。这表明，假设xT x = 1，使xT Ax最大化或（或最小化）的唯一的点正是A的特征向量。

**伪逆**

**轨运算**