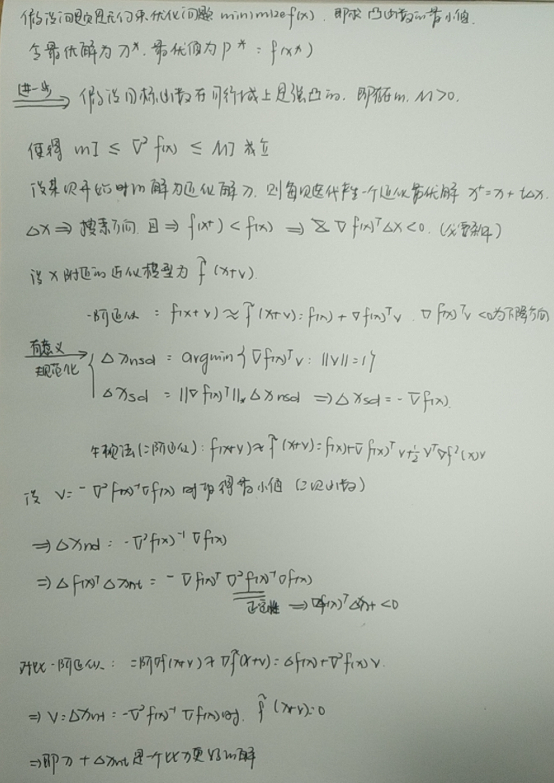
1. （必填）自己提出的问题的理解（罗列全部）：
2. **提出的问题1：**

从数学上来说，为什么牛顿法和拟牛顿法能够更快的找到极小值？

讨论后的理解：



在梯度下降法中，f(x)将以几何级数的方式收敛到p\*，即本次迭代的初始解x和优化后的解x+满足



有精确直线搜索性能超过回溯直线搜索的情况，但性能提升也不大。和对迭代次数的影响是noticeable的，但并不是dramatically，真正麻烦的是条件数。

牛顿法迭代分成两个过程，第一个过程是阻尼牛顿阶段，第二个才是二次收敛阶段。假设f(x)满足利普希茨条件

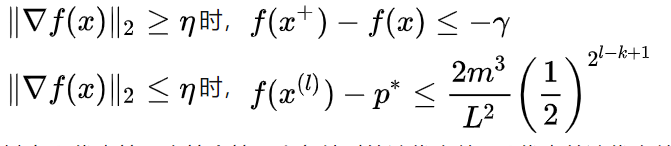


其中L为常数

可以证明，存在满足



使得



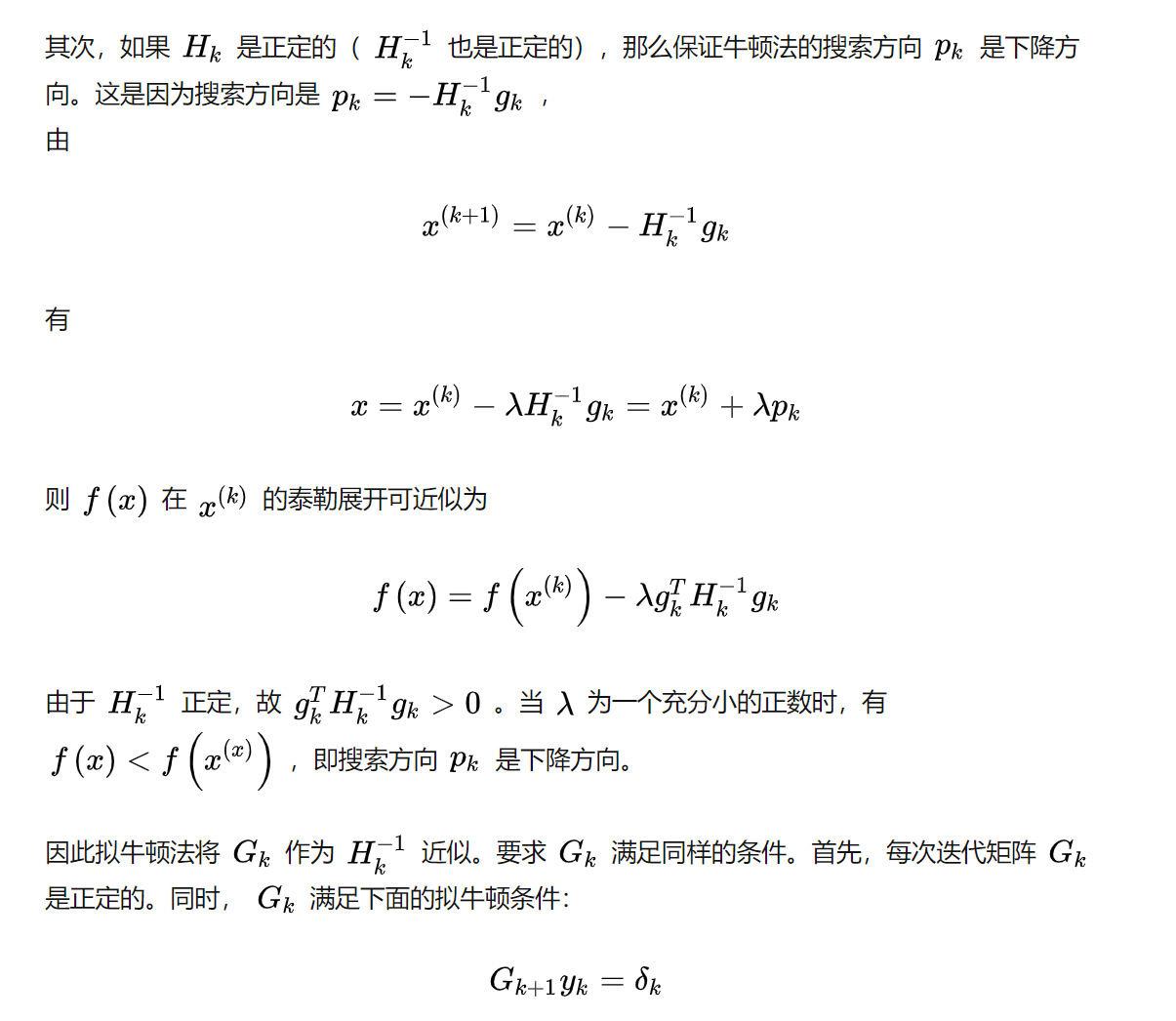
其中k代表第一次符合第二个条件时的迭代次数，l代表总迭代次数第一个阶段回溯直线搜索时步长t<1，因此称为阻尼牛顿阶段。在第二个阶段已经很靠近了，此时始终有t=1，牛顿法的收敛会极为迅速，称为二次收敛阶段，也称为纯牛顿阶段。在绝大多数的情况下，第二个阶段的迭代次数不会超过五六次，这也是牛顿法相较于梯度下降法最大的优势，而且牛顿法对高维计算的应对也非常好

1. **提出的问题2：**

P443为什么Hk是正定的可以保证牛顿法搜索方向pk是下降的方向？

讨论后的理解：

因为Hk是正定矩阵，因此Hk的逆矩阵也是正定矩阵，可以得Hkgk非负，根据公式可以推出pk恒小于0，即可以保证牛顿法的搜索方向为下降方向。

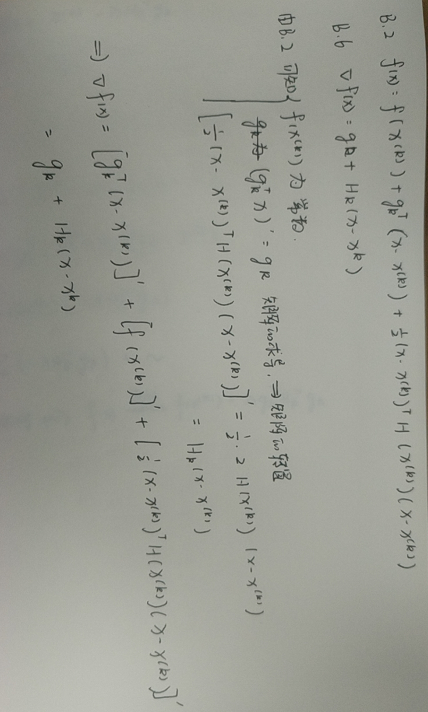


1. （必填）别人提出的问题的理解（选择几个问题罗列，并给出理解）：

1. **问题3**：

442，式B.2是怎么推式B.6

自己的理解：



2. **问题4：**

441页，为什么“特别是当H是正定矩阵时，函数f的极值为极小值”

自己的理解：

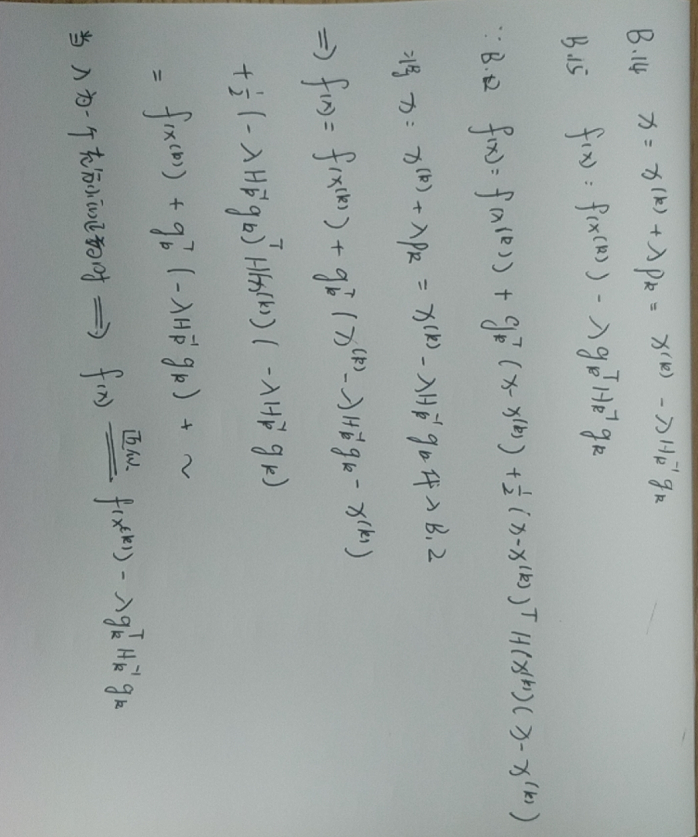
首先函数的一阶导数=0，可以确定极值点，根据二阶导数与极值点的关系可以知道，二阶导数大于0的极值点为最小值，二阶导数小于0的极值点为最大值。当Hk为正定矩阵时，Hk对应为f(x)在xk处的二阶导数。因此H为正定矩阵时，对应二阶导数大于0，即此处对应的极值为极小值。

根据定义

1. **问题5：**

如何从B.14推到B.15的？

讨论后的理解：



三、（必填）读书计划

1、本周完成的内容章节：统计学习方法（第六章：附录ABC）

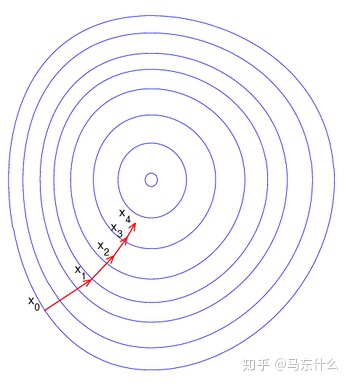
2、下周计划：统计学习方法（第七章：支持向量机）

四、（选做）读书摘要及理解或伪代码的具体实现（读书摘要、伪代码的具体实现代码等可以写到这个部分）

1、读书摘要及理解（选做）

**梯度下降算法：**

梯度下降法是最早最简单，也是最为常用的最优化方法。梯度下降法实现简单，当目标函数是凸函数时，梯度下降法的解是全局最优解。一般情况下，其解不保证是全局最优解，梯度下降法的速度也未必是最快的。梯度下降法的优化思想是用当前位置负梯度方向作为搜索方向，因为该方向为当前位置的最快下降方向，所以也被称为是”最速下降法“。最速下降法越接近目标值，步长越小，前进越慢。梯度下降法的搜索迭代示意图如下图所示：



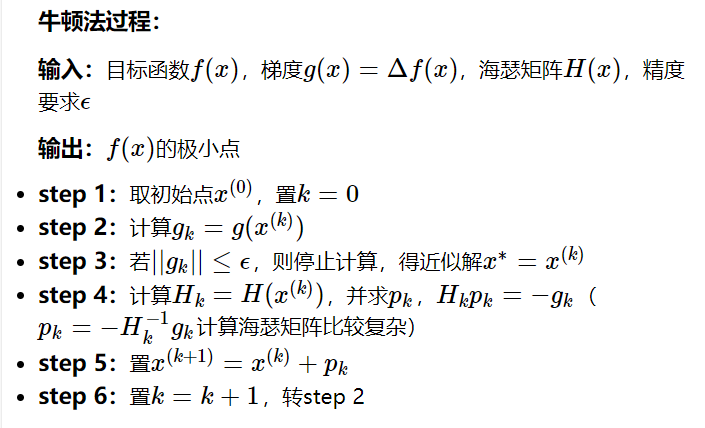
梯度下降法的缺点：

　　（1）靠近极小值时收敛速度减慢，如下图所示（这是因为越靠近极小值，参数的权重更新量越小，自己实现一个线性回归就知道了）；

　　（2）直线搜索时可能会产生一些问题（比较容易在极小值点附近震荡）；

（3）可能会“之字形”地下降。（特征没做归一化，不同量纲导致的梯度更新量差别大，迭代过程中容易偏来偏去的）

**牛顿法和拟牛顿法：**



优点：

二阶收敛，收敛速度快；

如果G∗正定，且初始点合适，算法二阶收敛、对正定二次函数，迭代一次就可以得到极小点

缺点：

牛顿法是一种迭代算法，每一步都需要求解目标函数的Hessian矩阵的逆矩阵，计算比较复杂。

牛顿法需要Hessian矩阵正定，如果非正定，会陷入鞍点

当初始点远离极小点时，牛顿法可能不受理，原因可能是因为牛顿方向不一定是下降方向，经迭代，目标函数值可能上式，此外，即使目标函数值下降，得到的点x(k+1)也不一定是沿牛顿方向的最好点或极小点。

**拉格朗日对偶性：**

对于拉格朗日乘子法求解，当原问题是凸优化的时候，考虑满足KKT的条件，对于拉格朗日函数求导等于0即可得到解。

对于对偶问题求解，转换为对偶函数求解，为了得到原始问题的解，要求强对偶，而在强对偶（不一定满足原问题凸优化）的情况下，考虑满足KKT的条件，对于拉格朗日函数求导等于0即可得到解。

当原问题是凸优化的时候，强对偶和KKT条件是互为充要条件的。

2、伪代码的具体实现(选做)