凸优化，一种包括最小二乘和线性规划问题的数学最优化问题

近年来一些新的发展吸引了人们对凸优化这块的兴趣。一方面，在1980s，人们用内点法去解决线性规划问题，和其他凸优化问题（如半限定规划和second-ordercone programs）,第二个发展是人们发现最优化问题在实际生活广泛存在。

将一个问题转换成凸优化问题的最主要优势在于：凸优化问题已经有非常成熟的算法（interior-point methods or other special methods for convex optimization）去解决。

通常，我们可以通过对偶将一个问题转化为凸优化问题。

最优化问题具有以下形式：

minimize f0(x)

subject to fi(x)<=bi , i=1,...,m （1.1）

最优化问题(1.1)称为一线性规划，如果目标函数f0和限制函数f1,....fm都是线性函数，即满足

fi(ax+by)=afi(x)+bfi(y) (1.2)

凸优化问题定义为目标和限制函数都是凸函数，即满足以下不等式

(ax+by)<=afi(x)+b (1.3) for all x,y and all a,b∈R,with a+b=1,a,b>=0

因为任意线性规划问题都是凸优化问题，所以我们可以认为凸优化问题是线性规划的扩展

像最小二乘和线性规划，已经有非常有效的可靠的算法

**最小二乘问题**：无限制的最优问题，an objective which is a sum of squares of terms of the form x − bi:

minimize f0(x) = ||Ax−b||22=i=1k(aiTx − bi)2 (1.4)

The solution of a least-squares problem (1.4) can be reduced to solving a set of  
linear equations,(A)x = b,so we have the analytical solution x = (ATA)−1ATb

The least-squares problem can be solved in a time approximately proportional to k,with a known constant.

在权重最小二乘中，权重最小二乘损失函数为 i=1kwi(aiTx − bi)2 (In a statistical setting, weighted least-squares arises in estimation of a vector x, given linear measurements corrupted by errors with unequal variances.)

另外一个用于最小二乘的技巧是 正则化，即损失函数添加一额外项。

In the simplest case, a positive multiple of the sum of squares of the variables is added to the cost function:  
 i=1kwi(aiTx − bi)2+ ，其中ρ>0.

**线性规划：**

the objective and all constraint functions are linear:  
 minimize x  
 subject to x ≤ bi, i = 1,... ,m (1.5)

对于线性规划问题，我们无法像最小二乘样给出解决所需时间，但我们可以大概的估计出用内点法进行求解线性规划所需的运算次数（n^2m,假设m>=n）

**using linear programming**

As a simple example, consider the Chebyshev approximation problem:  
 minimize maxi=1,…,k|aiTx−bi| (该目标函数是一个不可导函数)

此时可根据对偶改成线性规划问题：

minimize t (设t=maxi=1,…,k|aiTx−bi|，那么aiTx−bi≤t)

subjuect to x -t≤ bi, i = 1,... ,k

-aiTx -t≤- bi, i = 1,... ,k

**Solving convex optimization problems**

凸优化问题的解并没有一个解析表达式，但存在很多有效的算法求解凸优化问题，如内点法

each step requires on the order of max{n^3,n^2m,F } operations ，其中F是计算f0,f1...fm的一阶导和二阶导所需的计算量

将内点法用于非线性凸优化问题仍然是一个非常活跃的研究领域。

对于凸优化问题的几类重要问题，如二阶锥规划和集合规划问题，内点法正在逐渐成为一项成熟的技术。

**using convex optimization**

通常，凸优化问题的识别比较困难。 此外，较之线性规划问题，转换为凸优化问题的过程中存在更多的技巧。因此，判断某个问题是否属于凸优化问题或识别那些可以转换为凸优化问题的问题是具有挑战性的工作。本书的主要目的就是为读者建立这方面的知识。当我们具备了识别或表述一个凸优化问题的技巧时，我们将发现，超乎我们的想象，很多问题都可以利用凸优化求解。

**非线性优化**

对于一般的非线性规划问题(1,1),目前还没有有效的求解方法。有时看似简单的问题，变量个数可能还不到10，却非常难以求解。因此，现有的用于求解一般非线性规划问题的方法都是在放宽某些指标的条件下，采取不同的途径进行求解。

**局部优化**

在局部优化中，人们放宽对解的最优性的要求，不再搜寻使目标函数值最小的最优可行解，取而代之的是寻找局部最优解。 由于仅仅要求目标函数和约束函数可微，局部优化求解迅速，并可以处理大规模问题。

除了可能找不到全局最优解之外，局部优化方法还存在一些别的缺点。1.初始值的选取非常重要。 2，无法估计局部最优解相比全局最优解到底有多大的差距。3局部优化方法对算法那的参数值一般也较为敏感，通常需要针对某个具体的问题或某类问题进行调整。

局部优化问题相比最小二乘问题，线性规划问题以及凸优化问题需要更多的技巧。

非线性规划中的局部优化方法和凸优化进行比对：

大部分局部优化方法仅仅要求目标函数和约束函数可微，因此，将实际问题建模为非线性优化问题是相当直接的。当建模完成后，局部优化中的技巧体现在问题的求解上。 而凸优化的情形完全相反，技巧和难点体现在描述问题的环节，一旦问题呗建模为凸优化问题，求解过程相对来说就非常简单

**非凸问题中凸优化的应用**

**局部优化中利用凸优化进行初始值的选择**

对于一个非凸问题，我们先将其表述为近似凸优化问题。我们通过求解近似凸问题，得到近似问题的精确解。然后，我们用凸问题的精确解作为局部优化算法的初始值，求解原非凸问题。

**边界 bd**C = **cl**C\ **int**C

闭包 **cl**C

内部 **int**C

**凸集**

仿射组合：=1，称形式的点为x1,..xk的仿射组合（一个仿射组合就是一个点）

仿射集合定义：x1,...,xk属于C，并且=1时，属于C，则称C是一仿射集合

仿射集合C可以表示为C=V+x0={v+x0|v},即一个子空间加上一个偏移。定义仿射集合C的维数为子空间V=C-x0的维数，其中x0是C中的任意元素

线性方程组的解集是一个仿射集合。反之任意仿射集合可以表示为一个线性方程组的解集

仿射包：集合C中的点的所有仿射组合所成的集合为C的仿射包，记为affC:

affC = {=1},仿射包是包含C的最小的仿射集合（仿射包也是一仿射集合）

CaffC S是任一包含C的仿射集合。

**仿射维数与相对内部**

定义集合C的仿射维数为其仿射包的维数。

如果集合C的仿射维数小于n，那么这个集合在仿射集合affC中。

定义集合C的相对内部为affC 的内部，记为relintC,即

relintC = {xB(x,r)},其中B(x,r) = {y| ||y-x||<=r}.即半径为r,中心为x并由范数||.||定义的球； 定义集合C的相对边界为clC\relintC,此处clC表示C的闭包。

**凸集**

**集合C被称为凸集，如果C中任意两点间的线段仍然在C中，即对于任意x1,x2**和满足0<=<=1的都有。（显然凸集是一仿射集合）

**称点**为点x1,..xk的一个凸组合，其中+..+.且i=1...k

一个集合是凸集等价于集合包含其中所有点的凸组合。

**称集合C中所有点的凸组合的集合为其凸包，记为convC:，它是包含C的最小凸集**

**convC = {**i=1...k,+..+**}**

**凸与仿射定义类似，只是系数需大于0（系数大于0限定了其凸包/凸集的大小不能像仿射包/仿射包样无限大）**

单位圆环的仿射包是全空间但其凸包就是单位圆

**锥**

如果对于任意x和，都有，称集合C是锥或者非负齐次

用对称nxn矩阵的集合，即= {X}

**超平面与半空间**

超平面是具有下面形式的集合{x|=b},其中a,a=0且b

几何上，超平面{x|=b}可以解释为与给定向量a的内积为常数的点的集合

也可以看成法线方向为a的超平面，而常数b决定了这个平面从原点的偏移。

两个向量的内积<0,即钝角

范数||.||的范数锥是集合C={(x,t)| ||x||<=t}

多面体被定义为有限个线性等式和不等式的解集.

P={x|}

多面体是凸集。有界的多面体也称为多胞形。

单纯形是一类重要的多面体。

设k+1个点v0,..vk仿射独立，即v1-v0,...vk-v0线性独立， 那么，这些点决定了一个单纯形，如下 C=conv{v0,..vk}={| >=0,+..+}。这个单纯形的仿射维数为k。

单位单纯形是由零向量和单位向量0,e1,....en决定的n维单纯形。它可以表示为满足下列条件的向量的集合： C=conv{0,e1,..en} = {| >=0,+..+}

通过**保凸运算**可以从凸集构造出其他凸集。

1.交集运算（可数个凸集的交集还是凸集）

2.仿射函数

函数f:仿射的，如果它是一个线性函数和一个常数的和，即具有f(x)=Ax+b的形式，其中A,b

假设S是凸的，f:仿射函数。那么S在f下的象f(S) = {f(x)|x}是凸的

3.和运算

4.直积运算

5.透视函数

定义P:，P(z,t)=z/t为透视函数，其定义域为x。

透视函数对向量进行伸缩，或称为规范化，使得最后一维分量为1并舍弃之。

作用于透视函数的凸集得到的像还是凸集。

线性分式函数是由透视函数和仿射函数复合而成。

f=P。g给出的函数f: f(x) = (Ax+b)/() domf={x|}

称为线性分式（或投射）函数

广**义不等式**

广义不等式即上的偏序关系。这种偏序关系和R上的标准序有很多相同的性质。

正常锥K：1.K是凸的，2.K是闭的，3,K是实的，即具有非空内部4，K是尖的，即不包含直线

用正常锥K可以定义上的偏序关系如下： xy-x,这里x，也可以写成x

当K=时，偏序关系就是通常意义上R的.

**2.4.2 最小与极小元**

元素x是S中的一个最小元，当且仅当Sx+K,此处x+K表示可以与x相比并且大于或等于x的所有元素，即大于或等于x的所有元素

2.5分离与支撑超平面

2.5.1超平面分离定理

超平面分离定理：假设C和D是两个不相交的凸集，即C=，那么存在a和b，使得对于所有x有，,对于所有x. 超平面{x| }称为集合C和D的分离超平面

严格分离：

分离超平面满足更强的条件,即对于任意x有,并且任意x有

2.5.2 支撑超平面

设x0是C边界**bd**C上的一点。如果对于任意x,满足,那么称超平面{x| =}为集合C在点x0处的支撑超平面

2.6对偶锥与广义不等式

令K为一个锥。集合称为K的对偶锥

3凸函数

3.1.1定义

函数f: 是凸的，如果domf是凸集，且对于任意x，ydomf和任意0<=<=1

有 f()<= (3.1)

对于仿射函数，不等式(3.1)总成立。因此所有的仿射函数是既凸又凹的。反之，若某个函数是既凸又凹的，则其是仿射函数。

函数是凸的，当且仅当其在与其定义域相交的任何直线上都是凸的

3.1.2 扩展值延伸

如果f是凸函数，我们定义它的扩展值延伸

3.1.3 一阶条件

函数f是凸函数的充要条件是domf是凸集且对于任意x,ydomf,下式成立

f(y)>=f(x)+(tylor展开式的一阶近似。因为凸函数的二阶导恒大于0 ，所以该不等式对于凸函数恒成立)

不等式(3.2)说明从一个凸函数的局部信息(即它在某点的函数值及导数)，我们可以得到一些全局信息。

3.1.4 二阶条件

函数f是凸函数的充要条件是，其Hessian矩阵是半正定矩阵：即对于所有的xdomf，有

f(x)0

例子 ，范数均为凸函数

3.1.6 下水平集

函数f: R的下水平集定义为

={xdomf | f(x)<=}

对于任意值，凸函数的下水平集仍然是凸集。

3.1.7 上境图

函数f: R的图像定义为

{(x,f(x))|xdomf},

它是空间的一个自己。 函数f: R的上境图定义为

epi f={(x,t)| xdomf ,f(x)<=t}

一个函数是凸函数，当且仅当其上境图是凸集。

3.1.8 Jensen不等式及其扩展

f()<= (3.5)

不等式（3.5）可以扩展至无穷项和，积分以及期望。例如，如果在Sdomf上p(x)>=0,且,下式成立

f()<= f(Ex)<=E(f(x))

上述所有不等式均被称为Jensen不等式

3.1.9 不等式

很多注明的不等式都可以通过将Jensen不等式应用与合适的凸函数得到。（事实上，凸性和Jensen不等式可以构成不等式理论的基础。）

3,2 保凸运算

3.2.1 非负加权求和

f=

3.2.2 复合仿射映射

假设f: R ,A,以及b，定义g:为

g(x) = f(Ax+b)

其中dom g={x| Ax+b}。若函数f是凸函数，则g也是凸函数。

3.2.3 逐点最大和逐点上确界

函数和均为凸函数，则二者的逐点最大函数f

f(x) = max{} 其定义域为dom f=domf1domf2,仍然是凸函数。（该结论可以推广到无穷多项）

逐点上确界(它是逐点最大的一个推广)。

如果对于任意y函数f(x,y)关于x都是凸的，则函数g

g(x) = 关于x亦是凸的。

此时，函数g的定义域为 dom g={x| (x,y) }

3.2.4 复合

标量复合

首先考虑k=1的情况，即h: R->R, g:.仅考虑n=1的情况

f(x) = h(g(x)) ,dom f= {x }

因为复合函数f=h。g的二阶导数为

矢量复合

下面考虑k>=1的情况，设

f(x) = h(g(x)) = h(,

其中h: R, :

3.2.5 最小化

如果函数f关于(x,y)是凸函数，集合C是非空集合，定义函数g(x) = inf f(x,y)

则函数g(x)关于x是凸函数。

3.2.6 透视函数

假设f: R , 则f的透视函数g: 定义为 g(x,t)=tf(x/t)

其定义域为 domg ={(x,t) | x/t f,t>0}

透视运算是保凸运算

3.3 共轭函数

设函数f: R , 定义函数为

此函数称为函数f 的共轭函数。使上述上确界有限，即差值在dom f有上界的所有y构成了共轭函数的定义域。

显然，是凸函数，这是因为它是一系列y的凸函数（实质上是仿射函数）的逐点上确界。无论f是否是凸函数，

严格凸的二次函数。

考虑函数f(x) = 1/2,Q,对所有的y , x的函数1/2

x=处达到上确界，因此 1/2

3.3.2 基本性质

Fenchel 不等式

从共轭函数的定义我们可以得到，对任意x和y，如下不等式成立

f(x) + （Fenchel不等式，当f可微时，亦称为Young不等式）

共轭的共轭等于其本身

可微函数

可微函数f的共轭函数亦称为函数f的Legendre变换 ， 一般函数的共轭有时称为Fenchel共轭。

设函数f是凸函数且可微，使 - f(x)取最大的 满足y = ,反之，若满足y= - f(x)在处取最大值。因此，如果y=，我们有 = . 所以，给定任意y，我们可以求解梯度方程y=从而得到y处的共轭函数

3.4拟凸函数

3.4.1 定义及例子

函数f: R 称为拟凸函数（或者单峰函数），如果其定义域及其所有下水平集

={x},都是凸集。函数f是拟凸函数。

函数f是拟凹函数，如果-f是拟凸函数，即每个上水平集{x|f(x)>=}

若某函数既是拟凸函数又是拟凹函数，其为拟线性函数。

3.4.2 基本性质

在拟凸条件下，凸函数的很多性质仍然成立，或可以找到类似性质。

如： 函数f是拟凸函数的充要条件是，domf是凸集，且对于任意x,ydomf 及 0<=<=1,

有 f()<=max(f(x),f(y)) (3.19) （拟凸函数的Jensen不等式）

3.4.3 可微拟凸函数

如果函数f是凸函数且是函数f的全局极小点。然而，对于拟凸函数，这样的论断并不成立。

3.4.4 保拟凸运算

非负加权最大

拟凸函数的非负加权最大定义为

f=max{}

其中是拟凸函数。上述定义的函数f是拟凸函数。此性质可以扩展到一般的逐点上确界，即f(x) = 其中w(y)>=0.固定任意y,g(x,y)关于x是拟凸函数。则f(x)也是拟凸函数。

复合

如果函数g： R是拟凸函数，且函数h: R->R是非减的，则复合函数f = h。g是拟凸函数

最小化

3.4.5 通过一族凸函数进行表示

3.5 对数-凹函数和对数-凸函数

3.5.1 定义

称函数f: R 对数凹，如果对所有的x.有f(x)>0，且logf(x)是凹函数。

若logf(x)是凸函数，称f(x)是对数凸函数

3.5.2 相关性质

二次可微的对数-凸/凹函数

函数f是对数凸函数，当且仅当对任意x.，下式成立

f(x)

。。。对数凹函数，。。。。下式成立

f(x)

乘积，和，以及积分运算

对数凸性以及对数凹性对乘积以及证的伸缩运算是封闭的。

设f和g是对数凹函数，则逐点乘积函数h(x)=g(x)f(x)是对数凹函数，但对数凹函数的和一般不是对数凹函数

对数凹函数的积分

对数凹的性质在积分后仍然保留。如果函数f:x->R是对数凹函数，则

g(x)=在上是x的对数凹函数

如果函数f和g在上是对数凹的，则他们的卷积

(f\*g)(x) = 仍然是对数凹函数

3.6 关于广义不等式的凸性

3.6.1 关于广义不等式的单调性

如果下式成立

xy f(x)<=f(y).

称函数K-增

单调向量函数

函数f:R 在上非减，当且仅当对任意x,y下式成立

x1<=y1,...xn<=yn f(x)<=f(y).

矩阵单调函数。

矩阵f: 矩阵单调，如果在半正定锥内函数是单调的

3,6.2 关于广义不等式的凸性

4 凸优化问题

4.1 优化问题

minimize f(x)

subject to i=1,..,m

i=1,..,p （4.1）（优化问题的标准型）

称x为优化变量

D= 称为优化问题的定义域。

当点x满足约束时，称点x是可行的，所有可行点的集合称为可行集或约束集

问题（4.1）的最优值定义为

inf{f(x)|x}

最优点与局部最优点

如果是可行的并且f() = ,我们称最优点（最优解）。所有最优解的集合称为最优集。记为={x|x 且f(x)=}

满足f(x)<=()的可行解x称为-次优，所有-次优解的集合称为（4.1）的-次优解

称可行解x为局部最优，如果存在R>0,使得

f(x) = inf{f(z)|i=1,..,m,i=1,..,p ,||z-x||<=R}

可行性问题

find x

subject to i=1,..,m

i=1,..,p

可行性问题可以用来判断约束是否一致。

4.1.2 问题的标准表示

称（4.1）为优化问题的标准形式。

框约束(可行解是一个方框)

minimize f(x)

subject to i=1,...,n

4.1.3等价问题

如果从一个问题的解，很容易得到另一个问题的解，并且反之亦然，称这两个问题是等价的。

（如目标函数和约束函数乘以正数）

如 ：

变量变换

目标函数和约束函数的变换

松弛变量

等价于存在一个满足+.利用这个变换，我们可以得到问题

minimize f(x)

subject to >=0 , i=1,...,m

+, i=1,....m （4.7)

hi(x)=0, i=1...p

新的变量称为对应原不等式约束的松弛变量。通过引入松弛变量，可以将每个不等式约束替换为一个等式和一个非负约束（约束也变多了）

消除等式约束

消除线性等式约束

引入等式约束

优化部分变量

我们总有 (x),其中in.换言之，我们总可以通过先优化一部分变量再优化另一部分变量来达到优化一个函数的目的。

上境图问题形式

标准问题（4.1）的上境图形式为

minimize t

subject to f(x)-t<=0

(x)<=0, i=1,...m (4.11)

hi(x) = 0, i=1,..,p

其优化变量为x及 t。（x,t)是(4.11)的最优解当且仅当x是（4.1）的最优解

隐式与显式约束

采用章节3.1.2所提到的简单技巧（扩展值延伸），我们可以通过改变定义域将任何约束隐式地表达在目标函数中。

4、2 凸优化

4.2.1标准形式的凸优化问题

凸优化问题是形如

minimize f(x)

subject to (4.15)

, i =1,...,p

的问题，其中f(x),均是凸函数。和一般的标准形式相比，凸优化问题有三个附加的要求：

1.目标函数f(x)是凸的

2.不等式约束函数必须是凸的。

3.等式约束函数必须是仿射的。

凸优化问题的可行集是凸集。因为它是问题定义域 D=(m个凸的下水平集以及p个超平面的交集)。因此，在一个凸优化问题中，我们是在一个凸集上极小化一个凸的目标函数。

4.2.2局部最优解与全局最优解

凸优化问题的一个基础性质是其任意局部最优解也是全局最优解（拟凸优化不成立）

4.2.3 可微函数f0的最优性准则

f(y)>=f(x)+ 恒成立。

所以若x是f(x)的最优解，那么（4.21）()恒成立（两个向量的内积大于0，锐角）

这个最优性准则可以从几何上进行理解：如果0,那么意味着-在x处定义了可行集的一个支撑超平面

无约束问题：  
此时条件（4.21）可以简化为一个众所周知的x是最优解的充要条件是（此时只需取y=x-t）

4.2.4 等价的凸问题

4.2.5 拟凸优化

minimize f(x)

subject to (4.24)

, i =1,...,p

其中f(x)是拟凸函数

局部最优解与最优性条件

凸优化与拟凸优化问题最重要的不同在于拟凸优化问题可能有不是（全局）最优的局部最优解。

拟凸优化的最优性条件

x, (4.25)

(4.25)与（4.21）有两个重要的不同：

1.条件(4.25)仅仅是最优性的充分条件

2.条件(4.25)要求f的梯度非零

通过凸可行性问题求解拟凸优化问题

4.3 线性规划问题

当目标函数和约束函数都是仿射时，问题称作线性规划(LP).一般的线性规划具有以下形式：

minimize

subject to Gxh

Ax=b （4.27）

可行集P是多面体，目标函数是线性的

4.3.1 例子

线性规划的标准形式：

minimize

subject to Ax = b

x>=0

4.3.2 线性分式规划

在多面体上极小化仿射函数之比的问题称为线性分式规划：

4.4 二次优化问题

当凸优化问题的目标函数凸二次型并且约束函数为仿射时，该问题称为二次规划（QP）。

二次规划可以表述为

minimize (1/2)

subject to Gxh （4.34）

Ax=b

在二次规划问题中，我们在多面体上极小化一个凸二次函数

如果不仅目标函数，其不等式约束也是凸二次型，那么称该问题为二次约束二次规划（QCQP）

在QCQP中，我们在椭圆的交集构成的可行集上极小化凸二次函数

4.4.1 例子

最小二乘及回归

4.4.2 二阶锥规划

一个与二次规划紧密相关的问题是二阶锥规划（SOCP）：

minimize

subject to ||<= , i=1,...,m (4.36)

Fx = g

称这种形式的约束||<=，为二阶锥约束。

4.5 几何规划

4.5.1 单项式与正项式

函数 f: R, domf = 定义为

f(x) = c...

其中c>0, .它被称为单项式函数或简称为单项式。

单项式的和，即具有下列形式的函数称为正项式函数(具有K项),或简称为正项式

f(x) =

4.5.2 几何规划

具有下列形式的优化问题

minimize f(x)

subject to i=1,...,m （4.43）

= 1, i=1,...,p

被称为几何规划（GP），其中f(x),fi(x)为正项式。hi(x)为单项式。

4.5.3 凸形式的几何规划

几何规划一般不是凸优化问题，但通过变量代换以及目标，约束函数的转换，它们可以被转换为凸问题。（通常用定义变量）

4.7 向量优化

4.7.1 广义和凸和向量优化问题

将广义向量优化问题记为：

minimize(关于K) f(x)

subject to i=1,...,m （4.56）

= 0, i=1,...,p

K为正常锥，称向量优化问题(4.56)为凸向量优化问题。如果其目标函数f(x)是K-凸的，不等式约束函数f1,...fm是凸的并且等式约束函数h1,...,hp是仿射的。

向量优化令人困惑之处在于两个目标函数值f(x)和f(y)不一定可以比较。

4.7.2 最优解与值

4.7.3 Pareto 最优解与值

现在，考虑可达目标值集合不含最小元的情况，因此问题不含有最优解和最优值。如果f(x)是可达集合的极小元，我们称可行解x为Pareto最优（或有效的）、称f(x)为向量优化问题的一个Pareto最优值。

4.7.4 标量化

标量化是寻找向量问题Pareto最优解的标准技术。

5.1 Lagrange 对偶

5.1.1 Lagrange

Lagrange 对偶的基本思想是在目标函数中考虑问题(4.1)的约束条件，即添加约束条件的加权和，得到增广的目标函数。定义(4.1)的Lagrange函数L: x为

L(x,)=f(x)++

称为拉格朗日乘子向量，设原问题（5.1）即是（4.1）的最优解

5.1.2 Lagrange对偶函数

定义Lagrange对偶函数(或对偶函数)g：->R为Lagrange函数关于x取得的最小值：

即

g() = (确定后，取x使得L(x,)最小)

满足条件的()且g()>-,称为对偶可行

5.1.3 最优值的下界

对偶函数构成了（5.1）最优值下界：即对任意和下式成立

g()<= （5.2）（对偶函数给出原问题下界）

5.2 Lagrange 对偶问题

对于任意一组（）,其中,Lagrange对偶函数给出了优化问题(5,1)的最优值的一个下界，因此，我们可以得到和参数一个下界。一个自然的问题是：从Lagrange函数能够得到的最好下界是什么？

可以将该问题表示为优化问题

maximize g( (5.16)

subject to

上述问题成为问题(5.1)的Lagrange 对偶问题，其中定义域为dom g={()}称解()是对偶最优解，如果它是（5.16）的最优解。

Lagrange对偶问题(5.16)是一个凸优化问题，这是因为极大化的目标函数g(是凹函数(g( 有上界但不一定有下界)，且约束集是凸集。

5.2.1显式表达对偶约束

标准形式线性规划的对偶问题是只含有不等式约束的线性规划问题，反之亦然。

5.2.2 弱对偶性

Lagrange对偶问题的最优值，我们用根据定义，这是通过Lagrange函数得到的原问题最优值的最好下界。特别地，<=,即使原问题不是凸函数，上述不等式亦成立，这个性质称为弱对偶性。

5.2.3 强对偶性和Slater约束准则

如果等式成立

=

成立，即最优对偶间隙为零，那么强对偶性成立。这说明从Lagrange对偶函数得到的最好下界是紧的。

强对偶性成立的条件，称为约束准则。

一个简单的约束准则是Slater条件：存在一点xrelint D使得下式成立

,Ax=b

满足上述条件的店有时称为严格可行，这是因为不等式约束严格成立。

当Slater条件成立(且原问题是凸问题)时，强对偶性成立。

5.2.5矩阵对策的混合策略

5.3 几何解释

5.3.1 通过函数值集合理解强弱对偶性

5.5最优性条件

5.5.1 次优解认证和终止准则

设某个算法给出一系列原问题可行解以及对偶问题可行解(),k=1,2...,给定要求的绝对精度,那么终止算法的条件为  
 -g()<=

保证当算法终止的时候，

5.5.2互补松弛性

设原问题和对偶问题的最优值都可以达到且相等(即强对偶性成立)。令是原问题的最优解，

()是对偶问题的最优解，即

f() = maximize g( = g()

= (f(x)++

<= f()++(x\*是对偶问题的可行解)

<= f()

f()<= f(),所以上面两个不等式取等号。

因此有，; 其中所以后一求和表达式=0是恒成立的。 而对于，事实上，求和项的每一项都非正()，因此有=0,i=1,..m(5.48) 该条件称为互补松弛性。

可将互补松弛性条件改写成 ：

>0

=0

5.5.3 KKT最优性条件

现假设函数f0,....fm,h1,...,hp可微(因此定义域是开集)

非凸问题的KKT条件

设原问题和对偶问题的最优值都可以达到且相等(即强对偶性成立)。令是原问题的最优解，()是对偶问题的最优解。因为L(x,)关于x在因此函数在的导数必须为零，即

++=0

因此，我们有

i=1,..m

，i=1,..p

i=1,..m (5.49)

=0, i=1,..m

++=0

我们称上式为KKT条件

总之，对于目标函数和约束函数可微的任意优化问题，如果强对偶性成立，那么任何一对原问题最优解和对偶问题最优解必须满足KKT条件

凸问题的KKT条件

对目标函数和约束函数可微的任意凸优化问题，任意满足KKT条件的点分别是原，对偶最优解，对偶间隙

5.5.5 通过解对偶问题求解原问题

假设强对偶性成立，对偶最优解()已知。假设L(x,)的最小点，即下列问题的解

minimize f(x)++ （5.55）

唯一。那么如果问题(5.55)的解是原问题可行解，那么它就是原问题最优解。如果它不是原问题可行解，那么原问题不存在最优点，即原问题的最优解无法达到。

5.6 扰动与灵敏度分析

5.7 例子

5.7.1 引入新的变量以及相应的等式约束

5.7.2 变换目标函数

5.7.3 隐式约束

5.8 择一定理

5.8.1通过对偶函数建立弱择一性

正齐性：若, g()=

应用篇

6 逼近与拟合

6.1 范数逼急

6.1.1 基本的范数逼近问题

最简单的范数逼近问题是具有下列形式的无约束问题

minimize ||Ax-b|| (6.1)

逼近的解释

通过将Ax表示为

Ax=...+ 其中,...,为A的列，我们可以看出，范数逼近问题的目标是用A的列的线性组合，尽可能准确的逼近或拟合向量b

加权范数逼近问题

范数逼近问题的一个拓展是加权范数逼近问题

minimize ||W(Ax-b)||

其问题数据W称为权重矩阵，权重矩阵通常是对角矩阵，在这种情况下，它给出了对残差向量r=Ax-b，分量之间不同的相对强调程度。

6.1.2 罚函数逼近

我们将考虑-范数逼近问题的一个有用推广，其目标函数仅仅取决于残差的幅值分布。这个罚函数逼近问题具有形式

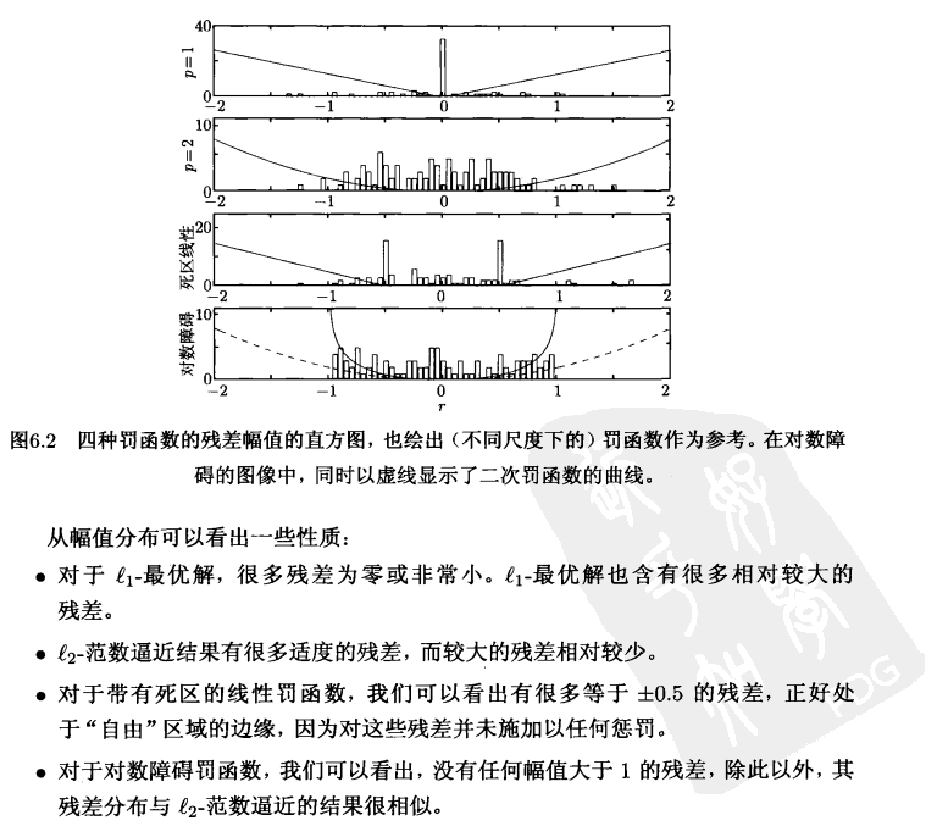
minimize + (6.2)

subject to r=Ax-b

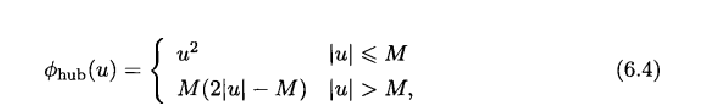
其中称为残差罚函数。在很多情况下，罚函数是对称，非负的，满足=0；

粗略地，度量了我们不喜欢残差值的程度，如果对于比较小的残差，值很小，表示我们不太关心具有这些值的残差。当r变大时，增长迅速，表示我们对于大的误差非常厌恶；例如，我们比较l1和l2范数逼近，对于|r|=1,这两个函数给与同样的罚。对于小的r,我们有

,因此,相比l2范数逼急，l1范数逼近给与大的残差的权重较小。对于大小残差的相对权重的不同，反应在相关逼近问题的解上。相对于l2范数，l1范数逼近问题最优残差的幅值分布趋向于有更多零或非常小的残差。相反，l2范数的解会趋向于有较小大的残差。(有特异点的时候，通常用l1范数会比较好) l1具有稀疏性



分段惩罚函数



当残差小于M时，这个罚函数与最小二乘的罚函数相同，而对于大的残差，它又恢复为类似l1的线性增长。

6.1.3 带有约束的逼近

变量的非负约束

变量范围

概率分布

范数球约束

6.2 最小范数问题

基本的最小范数问题具有下列形式

minimize ||x||

subject to Ax=b (6.5)

这个问题的解称为Ax=b的最小范数解，如果Ax=b有解，这样的解总是存在的。

6.3 正则化逼近

6.3.1 双准则式

在正则化逼近的基本形式中，我们的目标是寻找向量x使其较小，同时使得残差Ax-b小。自然地，这可以描述为双目标的凸向量优化问题，这两目标是||Ax-b||和||x||:

minimize(||AX-b||,||x||) （6.7）

这两个范数可能是不同的：第一个用以度量残差的规模；第二个用以度量x的规模。

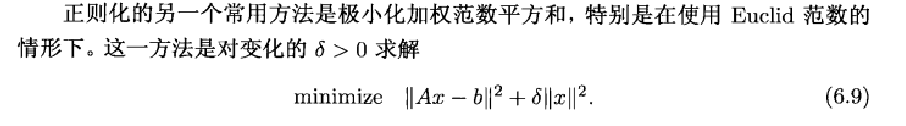
6.3.2 正则化

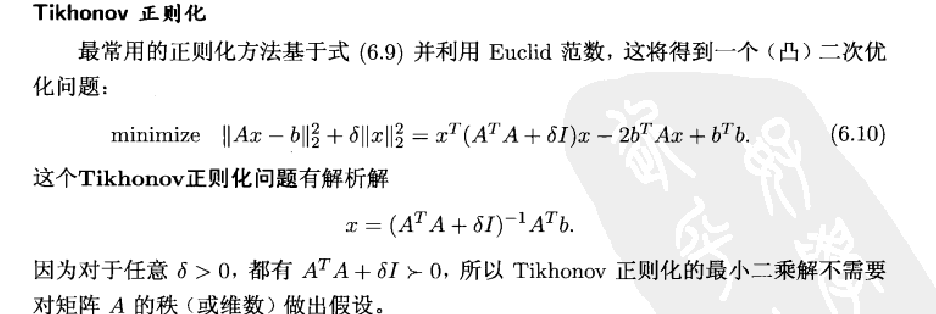
正则化是求解双准则问题(6.7)的一个常用的标量化方法。正则化的一种形式是极小化目标函数的加权和：

minimize ||Ax-b||+||x|| (6.8)

其中>0位问题参数。当在(0,)上变化时，(6.8)的解遍历了最优权衡曲线。

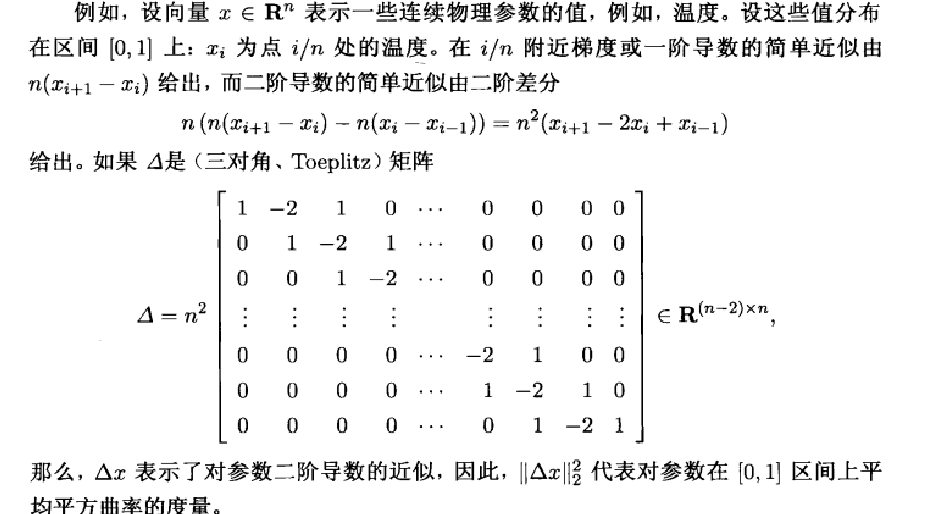
Tikhonov正则化



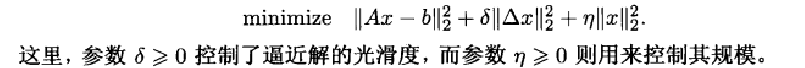
.

光滑正则化

正则化的想法，即向目标函数添加惩罚大的x的项，可以有很多扩展。其中一个有用的拓展是用||Dx||代替||x||作为正则化的项。在很多应用中，矩阵D代表近似的微分或二阶微分算子，因此||Dx||代表x的变化的度量或去光滑度。



添加多个正则化项。比如，于光滑程度和规模相关的项。



6.3.3 重构、光滑于去除噪声

信号x被加性噪声v所污染，即

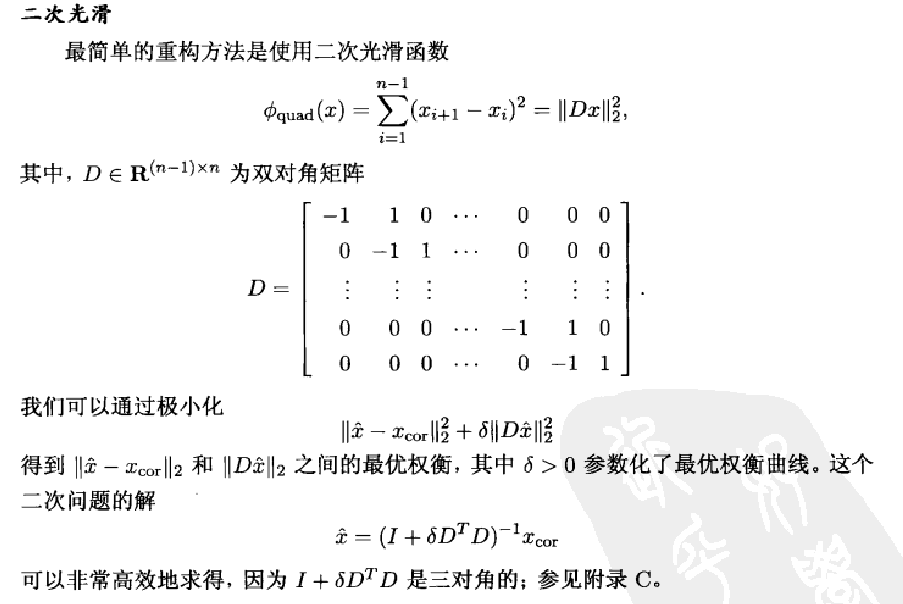
=x+v

这里，假设噪声v是未知，小值的，并且快速变化。目标是在给定受污染信号的情况下，构建对原始信号x的估计值。这一过程称为信号重构(因为我们试图从被污染的信号中构建出原信号)，或者去除噪声。

重构问题的一个简单方式是双准则问题

minimize (||||,) (6.12)

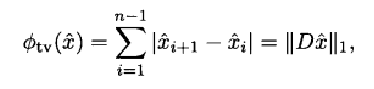
其中是变量而为问题参数。函数是凸的，称为正则化函数或光滑目标。这一目标被用来度量估计值的粗糙度，或光滑度的缺少。重构问题(6.12)寻求信号，以接近被污染信号并且光滑。



当原始信号非常光滑而噪声变化很快时，简单的二次光滑可以作为良好的重构方法。但是显然，原始信号的任意快速变化都会被二次光滑方法所减弱或者移除。

总变差重构

本节将描述一种重构方法，可以去除大部分的噪声，同时仍然保留原始信号偶尔的快速变化。这一方法基于光滑函数

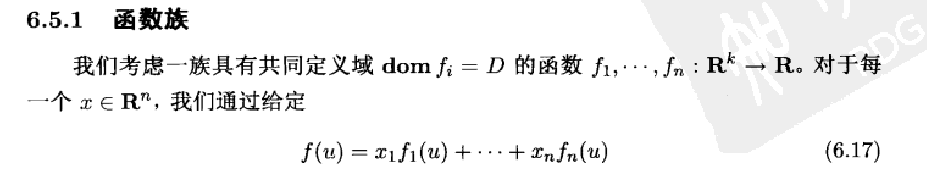


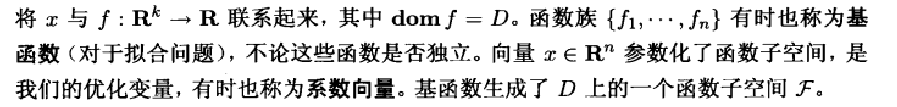
称为x的总变差。如同二次光滑性的度量总变差函数对于快速变化的|xi+1-xi|给与相对较小的惩罚。

6.4 鲁棒逼近

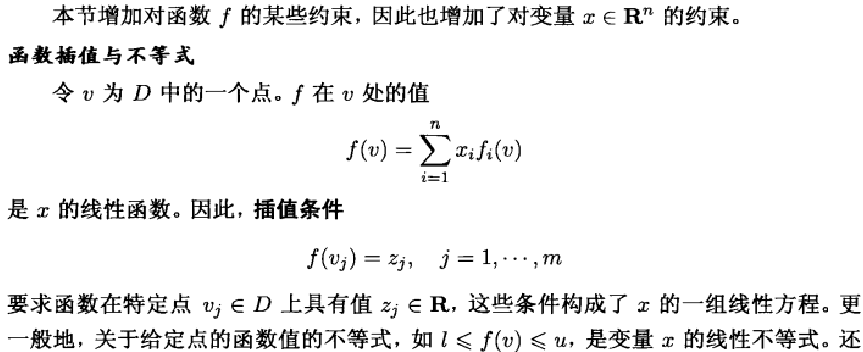
6.4.1 随机鲁棒逼近

6.5 函数拟合与插值



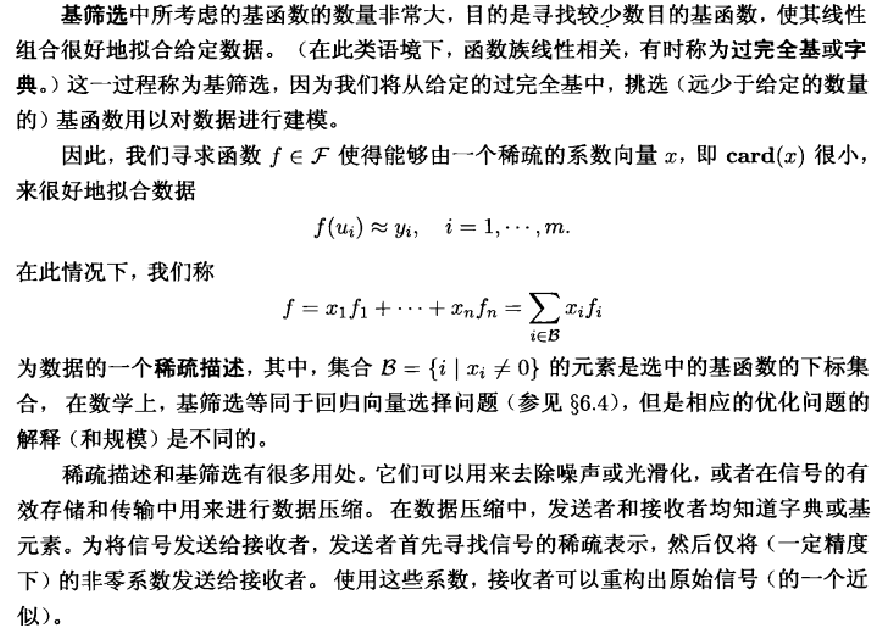


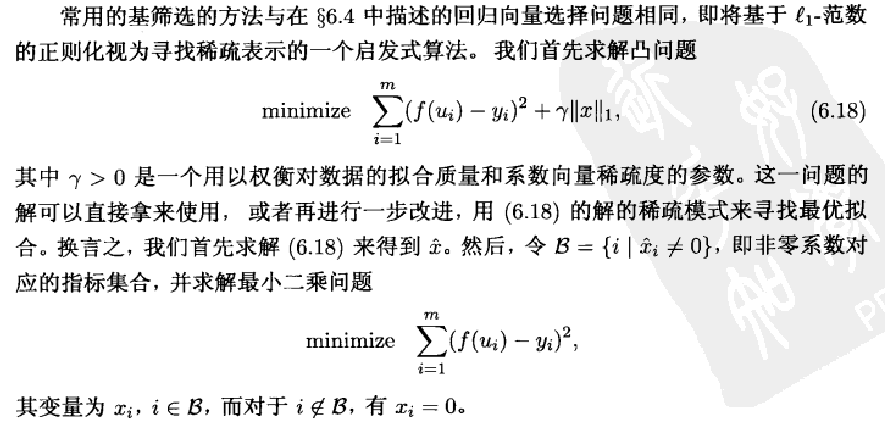
6.5.2 约束



6.5.3 拟合与插值问题

6.5.4稀疏描述与基筛选



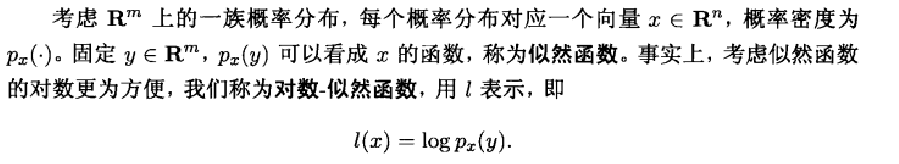


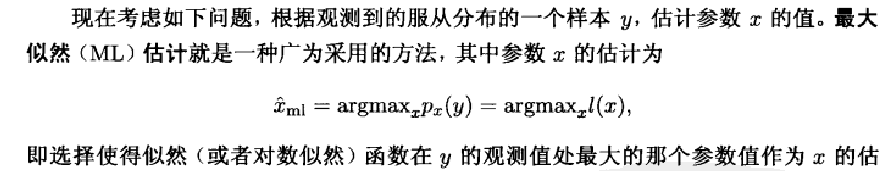
6.5.5 利用凸函数的插值

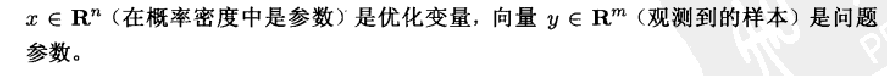
7统计估计

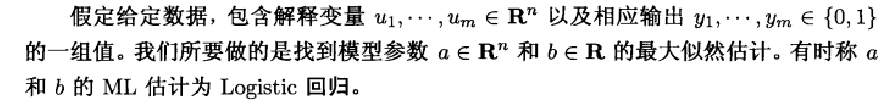
7.1 参数分布估计

7.1.1 最大似然估计









7.1.2 最大后验概率估计

最大后验概率估计问题可以看成最大似然估计的Bayes形式。此时假设未知参数x是服从某一预先设定的概率分布。假定x和y均是随机变量，而在统计估计中(如最大似然估计)，

x 是参数，而不是随机变量



